A Comprehensive Survey on Recent Trends and Techniques for Robust Face Detection and Recognition

[1]Seshaiah M,[2]Dr. Shrishail Math
[1]Research Scholar,[2]Professor
[1]Visvesvaraya Technological University - Regional Center
Belagavi, India.
[2]Dept. of CSE, Sri Krishna Institute of Technology
Bengaluru, India.

Abstract : Since last several years, image analysis and recognition have many successful applications; out of them the face recognition has become very popular. This popularity and usefulness of face recognition process became possible mainly because of two causes: the accessibility of reasonable technologies and an extensive variety of profitable and security related applications. The success of existing machine recognition systems success is controlled by the real time circumstances. The process of face recognition in an open-air atmosphere is influenced by the posture and proportion of light. Though the face recognition technology is very advanced but it is still less capable than the observation power of human system. We have projected the latest and significant survey of video and still based face identification and recognition study. In this article, we discuss about the recent methods for face detection and recognition using computer vision applications.

***IndexTerms* - Face Detection, Face Recognition, Image Processing, Visual Computing, Literature Survey.**

I. INTRODUCTION

Since last ten years there is a surge of interest in automated video surveillance from business and research [1]. The first prototypes and commercial products were typically embodied by monolithic systems, where video sensors, processing units and output devices were tightly coupled; thanks to the significant advances of the last ten years in algorithms, hardware platforms and networks, the paradigm of video surveillance is increasingly drifted toward the mobility and distribution of its modules (i.e. source, processing, monitoring, storage, etc.) [2].

The automatic analysis of surveillance videos is a key area of investigation for computer vision research and industry. This holds especially for techniques that interpret the behaviour in the monitored scene and is mainly due to the enormous variety of situations that occur in practice [3]. When unanticipated events take place, it becomes essential to perceive and inform circumstances of exceptional interest. Some solutions exist that work in well-constrained surveillance settings and show good results if they are tuned to a specific pre-defined application. However, for many application scenarios, an ideal automated visual surveillance system would autonomously interpret the scene and automatically recognize abnormal events. It would then notify operators or users accordingly, ideally including some semantic information with respect to the detected event [4].

With the replacement of conventional analogue technologies by the digital IP-based technologies, there is a huge increase in profits (billions $US) in the universal bazaar for video surveillance (VS) tools. There are many cameras present in the VS networks, and the network is capable of transmitting and buffering the enormous amounts of data which is helpful in reliable decision support systems. The selection and situation analysis can be improved by utilizing the automatic recognition and tracing ability of VS network in an extensive range of operating circumstances [5]. In active visuals, the visual surveillance makes an effort to detect, recognize and trace specific objects from image series or videos. Moreover, it also tries to realize and define the nature of the object. The main objective is to swap the conventional inactive video observation by advanced visual surveillance as the traditional videos are becoming ineffective with the increase in the number of cameras which surpasses the competence of individual operators to observe them. The visual surveillance is not only efficient to keep cameras as a replacement human vision, but also it tries to automate the complete surveillance procedure.

There is an extensive range of applications in active visual surveillance like deployment of security services in different types of organizations, traffic flow in metropolises and towns and discovery of army camps, etc. The human oriented surveillance applications are as follows:

(a) Entrance limitation in exceptional zones: The army camps and vital governmental organization need special mechanism to allow entry of peoples. A person without unique identity is not allowed to enter. Through biometric techniques a database of biometric features of authorized person has been built up. If a human wants to enter in a secure zone, the security system will measure the features of that person like facet emotions, height, and movement gesture and compare them with the features of authorized persons stored in the database. If a match is found, take decision if the sightseer can be allowed to access in the building [6].

(b) Identification of a Particular Human in certain situation: The police can use a smart surveillance system to catch suspects at a distance through their personal identification [7]. After building a database of biometric features of suspects, the police can install the visual surveillance mechanism at important sites where the suspects may emerge, like, airport, railway platforms, liquor shops etc. The identification process matches the suspects and decides if the person is actually a criminal or not. In case a suspect is found, an alarm has been raised instantly. However, the credibility of these systems is below the expectation of police requirements as these systems are generally installed in public real time scenarios.

(c) Analysis of traffic jam statistics: The visual surveillance structures may spontaneously calculate the flow of humans at significant civic zones like markets and streets and afterwards bestow the jamming investigation which helps in managing the humans in those areas. Likewise, the visual surveillance structures may observe and analyse the traffic flow in other types of road network such as superhighways and road intersections, and it can be of huge benefit for crux flow management [8].

(d) Irregularity recognition and warning: Sometimes, it is compulsory to analyse and determine the activities (usual or unusual) of human and automobiles [9]. The visual surveillance structures examine the unusual activities of criminals in public areas like parking zone and markets etc. If there is some unusual activity of a person is detected, then an alarm can be raised either by spontaneously playing a recorded public statement or by informing the police mechanically.

(e) Multiple cameras based co-operating surveillance: To safeguard the community of an area the interactive surveillance uses multiple cameras. It can trace suspects in a large zone through the collaboration of various cameras [10]. Interactive surveillance uses many cameras for traffic management, to aid the traffic police to identify and trace a rule breaker, and clutch the automobiles which are engaged in traffic offences [11].

1.1 Visual Surveillance Systems

Now a day, the main focus of the video surveillance applications is to examine human conducts and to recognize matters of deadlock risk examination and fortitude. The current survey covers the up-to-date developments and competences of visual surveillance structures. It also measures the possibility and contests of surveillance systems in detecting the unusual conduct of a person, in detecting the aggressive tentation of a person, and in identifying human substance.

Visual (or video) surveillance machines are being used to collect data and to observe humans, occasions and actions. In the visual surveillance marketplace, the video surveillance methodologies, optical / thermal cameras and night vision apparatuses, are the most popular devices. Since some last years, the highly vigorous research matter in computer vision and artificial intelligence is video surveillance of active sites, particularly for peoples. The video surveillance has a number of applications for community protection like access control, traffic management and jamming investigation, analysis of human conduct in various atmospheres, etc.

The processing stages of the structure of a computerized video surveillance structure is as follows: detection of object movement[13], object classification [14], object tracing [15], human identification, action and conduct investigation [9], and data synthesis and camera management[12]. Table 1 demonstrates the probable appliances of face recognition.

Table.1 Appliances of face recognition

Application Area	Specific application
Entertainment	Video game, virtual reality, training programs Human-robot-interaction, human-computer-interaction
Information security	TV Parental control, personal device logon, desktop logon Information security Application security, database security, file encryption Intranet security, internet access, medical records Secure trading terminals
Smart cards	Human-robot-interaction, human-computer-interaction Drivers' licenses, entitlement programs Smart cards Immigration, national ID, passports, voter registration Welfare fraud
Law enforcement and surveillance	Advanced video surveillance, CCTV control Portal control, post-event analysis Shoplifting, suspect tracking and investigation

The working of nearly each video surveillance structure begins with object and movement detection. To detect motions in an image, the image is divided into various segments regions and movements are matched against these segments. Object tracing, performance analysis and identification processes are largely dependent on the segmentation process. Background environment modelling and motion segmentation are two main parts of the object motion detection process. During the detection process, these two parts intersect each other. In an image series, the motion segmentation tries to detect areas which have moving objects like automobiles and peoples. Regions including movement are used in next methods like tracing and conduct study because merely these zones are required to be studied for next examination.

When the task of movement and entity detection is finished, moving objects are traced from one frame to another in the surveillance structures. The motion detection method involves the tracing methods which uses pictures main points like spot, curve and line to identify movement from one frame to another.

Movement patterns examination, recognition and the yielding of accurate depiction of movements and communications among entities are the main steps of Behaviour understanding. In certain situations, it is better to analyse the conduct of humans and it is decided if they are behaving in a usual or unusual manner. It is important to ensure that a person with abnormal behaviour does not enter a security area. In video surveillance structures, the person' face and gesture are considered as the chief biometric characteristics which may be utilized for an individual identification.

One camera based video surveillance structures can be used for movement finding, tracing, actions comprehension, and personal identification from a particular distance. On the other hand, manifold camera based video surveillance structures may be enormously useful if the area to be observed from different view point or if the view is obstructed by different objects. If the scene contains depth or obstruction, then it is difficult to track by a single camera. However this problem can be resolved by installing multiple cameras. Nonetheless, some problems are also associated with multiple camera based video surveillance systems like camera deployment, fixing of minimum number of cameras to cover the entire scene, camera adjustment, entity matching, automatic exchange of camera and data merging.

In this field of computer vision based applications, the face recognition is also plays important role to improve the system security.

Hence, recently developed visual surveillance systems are also focusing on the face recognition from CCTV footage or other types of videos for surveillance.

1.2 Face Detection and Recognition from Video Data

Since last one era, there has been an intensified research on face detection from an active scene (video). Generally the videos are recorded in real environment scenarios which produce a low quality picture; therefore the main focus face recognition is to find accurate face from a video without degrading the quality [16]. In conventional facial picture procurement surroundings, like police bases or passport organizations; face appearance are measured in the form of variables stretching from head pose to skin colour. Since the visual surveillance structures are not conspicuous, the actions of the documented persons and the influences of the atmosphere on them may differ considerably. After several experiments, it has been concluded that the face identification procedures which work fine in a restricted atmosphere have a tendency to lower down the quality in surveillance circumstances [17]. These problems lead to the expansion of face identification procedures which are developed from the pool of knowledge given by videotapes and these procedures has provided good result in unrestricted environment situations. Human with known faces can be easily identified with the help of temporal dynamics. Fig.1 shows a basic architecture of face recognition model.

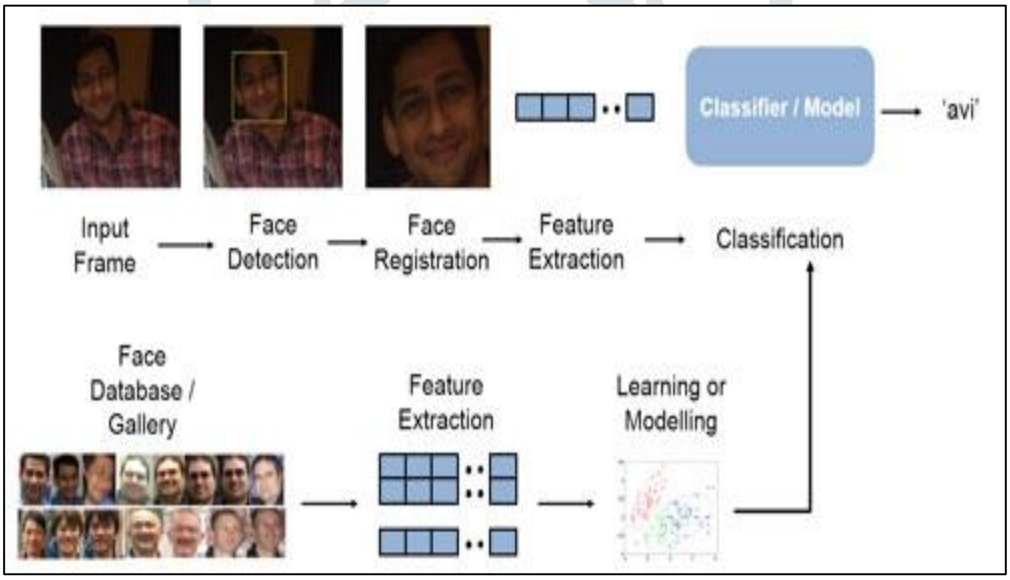

Fig.1. Face Recognition model

1.2.1 Challenges in face recognition

Unrestricted face recognition appliances may have following problematic issues:

- Posture disparity – By considering the equivalences among pixel positions and spots through different angles, the unrestrained cameras can film a non-perfect face snaps and it can vary from one picture to another.
- Lighting disparity - The texture of a face looks dissimilar at different times as a person moves through various types of light shades and through a wide range of changing positions and sharpness during the course of the video recording.
- Appearance disparity – Along with facial expression the appearance of the face also varies.
- Measure disparity – If the face moves towards the camera the video frame will become large, however if the face move away from the camera, the video frame will become small. The spatial resolution of the video frame can be so low that face becomes so distorted that it cannot be recognized finely. The quality of the picture may also be influenced by depth of field of lens of the camera.
- Movement distortion - if the camera uncovering period is kept very lengthy or the face moves quickly then there can be substantial distortion in the image.

- Occlusion – When an image is captured, there can be several objects present in the atmosphere which can obstruct few parts of the face. It can make hard to recognize the face and differentiating it from the context.

These issues are so influencing that these can instigate the variances in looks of same person in two different shots. However in the absence of these factors difference between two human under similar conditions can be very less. In the uncontrolled atmosphere, not only posture and lighting but other factors are also equally responsible to distort an image, Phillips et al. [17].

On the basis of the computer vision method, the complete practice of visual surveillance system for face identification can be categorized into two important groups such as: (1) face detection and (2) face recognition.

1.3 Face Detection Techniques

In this unit recent techniques for face detection from natural scene video sequences have been discussed. Face detection using still images has been studied widely. Similarly, face detection-based schemes are widely adopted for video database. We have also analysed the recent techniques of face detection in video scenes and still images.

In the actual world, the face detection faces huge visual disparity, like posture problem, appearance, illumination variations, as a result it needs an improved and distinguish framework to precisely identify faces from a scene, Li et al. [18]. Thus, computation excessiveness is required for problem solving models. In order to resolve the several disruptive issues, the authors have suggested cascade design which is based on convolutional neural networks (CNNs). The CNN method has a strong distinguish ability with super functioning. The CNN method performs with manifold purposes. It rapidly discards the contextual zones obtained in the previous low resolution phases. Additionally it estimates a short range of participating contenders from the previous excessive resolution phase. The number of contenders is decreased in later phases to enhance the localization efficiency. The authors have employed a calibration phase after every detection phase of the cascade. The detection window position is adjusted according to the outcome of every calibration phase and taken as input for the succeeding phase.

A fully automatic FAST-DT method capable of identifying and tracing an entity was proposed by Comaschi et al. [21]. Devoid of trusting on some scene precise cues, a number of faces exist online. It incorporates a generic face detector and a flexible organized SVM tracker output. Further it exploits the detector’s constant buoyancy to resolve the object formation and elimination issue.

Owing to the significance of safety of humans in various organization, it is important to observe the actions of people entering in the zone as well as identifying suspicious humans by surveillance cameras, Rasti et al. [26]. Usually, the face recognition methods which involve the use of camera for surveillance face problem due to low resolution of camera. The authors has used deep learning convolutional network to resolve the problem of low resolution.

The capability of face detection can be enhanced through automatic skin colour differentiator, Jairath et al. [27]. But it is very challenging work to achieve skin colour steadiness through different lightening conditions, changed camera functions, and varied background. Fixed skin colour prototypes which depend on image pre-processing are not able to achieve full constancy. This work suggests a flexible skin colour prototype to decrease the incorrect consent instances of Adaboost face indicator. Through the earlier Adaboost reactions, the face colour distribution framework was frequently updated; therefore this system proved to be added operative for physical world atmosphere conditions.

In some applications, Fast face detection high precision and less energy depletion is required Yin et al. [28]. The authors have projected an Ada-Boost related face detection method because the Ada-Boost face identification shows the features of fact calculation in both ways along with disparity in data. Moreover the Ada-Boost performs the parallel structure mechanism with optimal calculation. The framework has two level mutual memory arrangements as well as parallel structure arrays. It advances parallel programming proficiency and facilitates the sub-window flexible cascade based categorization for facts disparity via double route based fundamental image processing that increases the recognition competence in dissimilar types of faces.

Elrefaei et al. [29] have suggested a model which can identify a criminal and help police to match the facet of a suspect or offender. It provides the real-time surveillance through video based client-server face identification. The model uses Android based cell phone gadgets at the client part and video based face identification at the server part for face identification and criminal tracing. To avoid the impact of illumination an error free Viola-Jones methods has been used for the different phases of face detection. The tracking phase follows the Optical Flow process. In the projected model, the optical flow is employed by some attribute extraction procedures like Fast Corner attributes and Regular attributes. They have implemented the face identification and tracking by Android studio and OpenCV archive, and verified it by Sony Xperia Z2 Android 5.1 cell phone.

Haghighat et al. [30] have suggested a Discriminant Correlation Analysis (DCA) based low resolution face detection method. DCA estimates the relationship of the attributes in high quality resolution as well as low quality resolution pictures and then discover the predictions which raise the couple wise relationships among the two feature groups. It also separates the classes inside every set with time. The matching can be employed to predict the attributes mined from high quality and low quality resolution pictures into a shared zone. The suggested process is efficient enough for complex real life applications like identification of different faces in a congested frame from a surveillance videotape.

Dadi et al. [31] has presented a Gaussian mixture model based method for face identification and human tracing. The prototype of GMM has been distributed into various zones for the tracing of a particular person. The zones are positioned over one another and traced concurrently. To identify a person, the maps of positioned slopes attributes of the face area are provided to the support vector device categorizer. Different consecutive frames are considered while performing experiments of training faces.

Yu et al. [32] have trained the categorizer by CAS-PEAL-R1 facial database which differ in posture, illumination, decorations and appearance. In pursuance to resolve the complication of face recognition in investigation video apply the categorizer on the database. Noise from a frame of single frame has been removed through the median and average filtering, subsequently, the skin colour dissection of the pre-processed picture was accomplished through the uncomplicated skin colour prototype created in YCbCr area. They have followed symmetrical guidelines to eliminate some sections of face so that face detection process can take place at fast pace and it does not have to waste time on redundant face parts.

According to Mutneja et al. [33], the computation proficiency of training course of Viola–Jones face recognition procedure should be enhanced to improve the working of framework. The authors have speed up detection process by parallelizing the preparation of rectangular Haar characteristics on GPU. They have picked up attributes via Ada-Boost techniques used in choosing modest categorizer. The employment of parallelization methods has been done to improve Viola–Jones face recognition procedure in amalgamation with skin colour straining to decrease the search area. With the help of skin color filtering, they have reduced the search space up to a great extent and also reduce time/cost to detect a face.

To overawed issues in video related face detection, Ding et al. [34] have planned an inclusive structure which is related to Convolutional Neural Networks (CNN). In pursuance to explain the shortage in real life video training facts, initially they have learnt distorted and error free face illustrations and then synthetically blur training facet that are organized of perfect static pictures. The CNN is motivated to understand blur oblivious attributes spontaneously through training facts that are constituted of static pictures and artificially distorted data. Next, they have suggested a Trunk-Branch Ensemble CNN model (TBE-CNN) that improves sturdiness of CNN characteristics for posture disparities and obstruction. It also mines corresponding evidence from general face pictures and spots picked from facet parts. Through division of low and middle level convolutional layers among the trunk and branch networks. They has also employed an upgraded triplet loss function to endorse the distinguish features of the presentations acquired by the TBE-CNN. They have verified their method through orderly simulations.

Liu et al. [35] have encountered several problem which are associated with face detection and encryption. In face recognition, skin-color related methods uses the fuzzy clustering techniques to detect facial applicants roughly. After that faces are upgrade through SVM categorizer. For face enciphering, a changeable and fused encryption (decryption) scheme is proposed which is based on space and value jumbling prototypes.

Although, we humans can efficiently identify a face, but we cannot handle the different types of face simultaneously, Singh et al. [36]. On the other hand, the computers are equipped with large memory, processing power and high speed computation. The authors have worked to identify a face contained in a frame of video. They performed it by mining the face and then by computing the Eigen face value by normalizing the face picture which can be compared with Eigen faces available in the database. To detect a face, authors have utilized the Viola-jones method and to match a face they have used Eigen face method. In visual surveillance, the face comparison operation should be quick enough. They have used their method for detection of a suspect in a video.

Table.2. Comparative analysis for face detection schemes.

Reference	Database (Image/Video)	Pre/Post-processing	Segmentation	Feature	Classifier
Li et al. [18]	Annotated Faces in the Wild (AFW) [19] and Face Detection Data Set and Benchmark (FDDB) [20]	NA	x	CNN bounding box calibration	CNN
Comaschi et al. [21]	TA2 dataset(Video)	Bounding box Post-Processing	x	Haar	Adaboost Classifier
Liao et al. [23]	FDDB Database [20], GENKI Database [24], CMU-MIT Database [25]	Post processing to merge the nearby detection.	x	Normalized Pixel Difference	Ada-Boost classifier
Rasti et al. [26]	FERET, Head-Pose, and Essex University databases and iCV Face Recognition database (iCV-F)	x	x	Hidden markov Model and Singular Value Decomposition	CNN (Convolutional Neural Network)
Yin et al. [28]	Low power reconfigurable architecture for hardware application by using FPGA.				
Elrefaei et al. [29]	Real-time mobile application	Image cropping and unwanted bounding box removal	x	Corner point extraction and Optical flow	Cascade Classifier
Haghighat et al. [30]	SCface database [72], FRGC database [73]	Image up-sampling and down-	x	Low and High Resolution features	Minimum Distance Classifier

		sampling			
Dadi et al. [31]	AITAM1(simple) AITAM2 (moderate) and AITAM3 (complex),	x	Gaussian Mixture model	Histogram oriented gradient (HoG) and Block normalization	Support Vector Machine Classifier
Yu et al. [32]	CASPEAL-R1	Noise removal using median and mean filtering	Skin Colour Segmentation	Haar-Like Features	Ada-Boost
Mutneja et al. [33]	CALTECH Griffin et al. [75], WIDER Yang et. al. [76]	x	X (Segmentation can improve the performance as future work)	Haar Features	Viola–Jones algorithm
Ding et al. [34]	PaSC [77], COX Face [78], and YouTube Faces [79]	x	x	Blur intensive features	Trunk-Branch Ensemble Convolutional Neural Networks
Sharma et al. [39]	FDDB [20], YALE [80]	x	x	Modified Affine Transformation	Haar-Cascade Classifier
Owusu et al. [40]	FDDB, CMU-MIT [25]	Image down-sampling	x	Haar features	Neural Network back propagation

Face detection of a person in varying illumination and colour context is a complex process, Chawla et al [37]. If a person face is not towards the camera or his face is not according to the angle of the camera, the system may fail in identify a face. In their work, several vital examples have been studied and these examples are mainly related to safety surveillance which is a must in our lives.

Movement and skin colour based segmentation may be utilized to mine the area of concern for face detection, Mutneja et al. [38]. Real time face recognition can be further accelerated through GPU. Initially, they have provided flexibility for selection of highest and lowest standards of scaling features to enable the multi scale face recognition which is related to the study of zones separated by movement and skin colour pixels. To enable multi scale face recognition, they have explored a pre- prepared Haar classifiers and picture scaling instead of Viola–Jones method which relies on detector scaling.

An innovative and real time based face detection method has been proposed to identify the tilted and occluded faces under various lightings conditions and in difficult postures, Sharma et al. [39]. Basically it is a desktop application consists of a user friendly interface which gathers the pictures from a web camera and recognises the faces from an image through a Haar-cascaded classifier with improved Census Alter characteristics. To overcome the problems (like slanted or obstructed faces with diverse illuminations) associated with cascaded classifier, they have planned Viola Jones based Improved Affine Transformation structure.

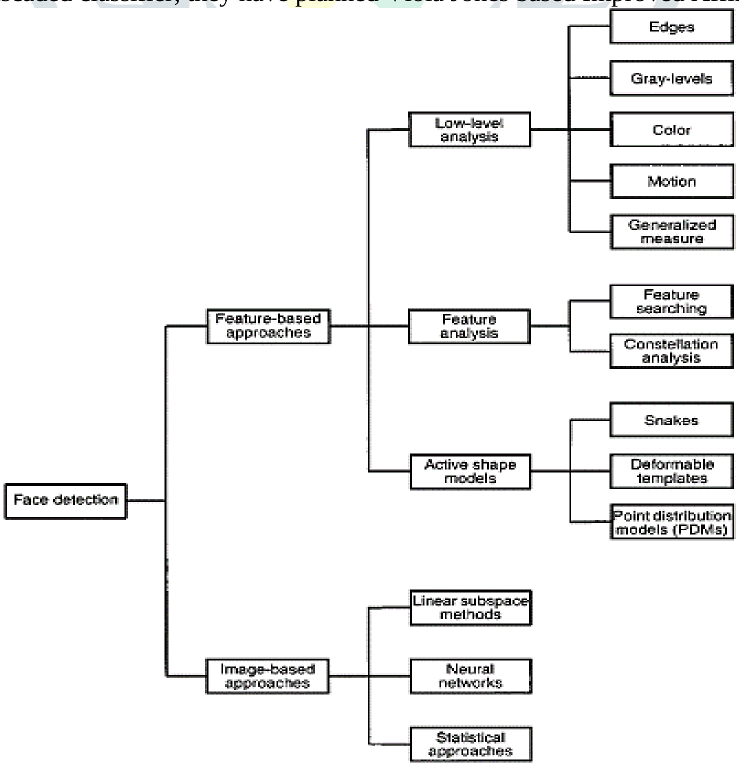

Fig.2. Face detection techniques

The face and facial appearance recognition system needs quick and precise discovery of facial data, Owusu et al. [40]. The types of these structures are internet protocol based visual observation structures; crime site snapshots structures, and database of criminals. The authors have mainly focused on preciseness and speed of the system. The Haar methodologies are used to mine the prominent facial features. Bessel down sampling method is used to decrease the dimension of the picture. Moreover, this process conserves the quality and particulars of the innovative image. Afterwards, anisotropic smoothing is used for image normalization. To classify the images, they used Multilayer feed-forward neural network with a back-propagation method.

The classifier with low computation price may be used initially for contraction of the contextual cascade in the face detection process, Yang et al. [41]. The authors have proposed a new cascaded CNN process comprising of two stages. The first phase consists of a low-pixel element window which is employed as an input for the shallow CNN and rapidly mines the candidate window. The next phase consist of the window achieved from the previous phase, this window is altered and work as an input to the analogous network layer. The cooperative online training is organised for rigid samples and the lenient non maximum dominance procedure is employed to examine the dataset.

This section shows various recent techniques in this field of face recognition. Dissimilar type of datasets is used to implement these techniques and the characteristics of these datasets are shown in table 3.

Table.3. Face detection dataset, their performance and properties

Dataset Name	No. of subjects	Images /Videos	Properties	Performance	Sample Images
Annotated Faces in the Wild (AFW) [19]	468	205	Cluttered background with aging, sunglasses, make-ups, skin colour, expression, etc. occlusions.	Precision - 97.9[18], Fast-HyperFace = 97.6% [65]	
Face Detection Data Set and Benchmark (FDDB) [20]	5171	2845	Occlusions, difficult poses, and low resolution and out-of-focus faces	Precision [65]=90.1%	
GENKI Database [24]	141	3500	wide range of backgrounds, illumination conditions, geographical locations, personal identity, and ethnicity	ROC=96.33 [82]	
CMU-MIT Database [25]	130	507	Different angles and illumination	Detection rate [40]=98.5 %	
FERET [69]	1199	14126	Various types of expression, age groups etc.	Accuracy= 80.73 [83]	
SCface database [72]	130	4160	Resolution, pose, illumination, expression	ROC=56.44 [84]	
FRGC database [73]	466	50000	poor quality, such as large illumination variations, low resolution, and blurring	Accuracy [81] =89.74%	

AITAM1[31]	5	Video frames (25 fps and 68s video length)	Occlusions	Accuracy=86.21 [31]	

1.4 Face recognition

In this section we present face recognition-based studies for visual surveillance system based on the computer vision system techniques. One of the principal applications of the FR is to enhance security at sensitive areas. Many times, it is desired to recognize faces from videos. Three major categories in video FR are: (1) still image FR, (2) multimodal FR, and (3) spatiotemporal algorithms.

1.4.1 Still Image Face Recognition

In still image-based schemes, FR is achieved in two steps. (1) The applied algorithms first automatically detect and subdivide the face from the input video and (2) employs some suitable static image technique to conclude the face identity. As illustrated in Fig. 3, still image FR has several subtasks. A fully automated FR system must perform these subtasks. From research point of view, an insight to subtasks is important. These algorithms are useful in many decisive situations and also required for the improvement of object subtasks,as shown in Fig. 3. In Ref. [50], the authors reported 100% recognition accuracy for rank-1 on FG-Net database by using Walsh–Hadamard transform encoded LBP. More recent advances in still image FR can be found in [51,52].

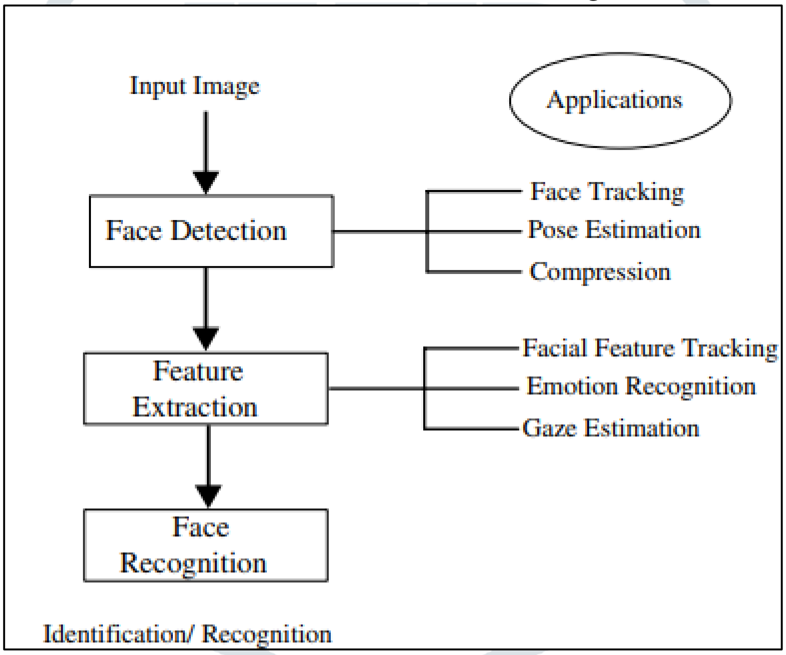

Fig.3. Still image face recognition systems.

1.4.2 Multimodal FR algorithms

This category lies in video- and audio-based FR. Such schemes are very important in non-cooperative environment, such as robbery or criminal footages. Recently, researchers in [53] experimented using Local LBP with the WPCA on FERET database and reported 91.5–99.6% accuracy. Whereas in [54] researchers calculated principal facial curvatures and achieved classification using the ICP and reported an accuracy of 66.7–79.5% on FRGCV2.0 database. Tang et al. presented a FR algorithm by introducing harmonization of spatial frames, analysis of multilevel discriminant subspace, and through Multi Classifier Integration (MCI) [55]. Developed scheme established the face resemblance through information of audio and video. Later, MCI was deployed on synchronized sequence. Proposed technique preserved time and space based knowledge preserved in a video series. Simulations conducted on the XM2VTS database revealed 99.3% accuracy.

1.4.3 Spatiotemporal algorithms

This category coherently exploits both space (all frames in a video) as well as time based information (like facial attributes and curves of a face). In [56], the authors proposed an interesting approach based on facial level curves. This approach calculated pair and section based distances among the levels of the curves. These level comprised of the space and time based attributes for recognition of expressions. Experiments conducted on BU-4DFE database reported 92.22% accuracy. A precision of 92.23% was informed for database developed by the authors. In [57], the researchers presented a sparse compensation based technique on a set of spots of face. The FR was achieved in three distinct steps. (1) A universal face picture was built through optimum figures of the incorporated training pictures. (2) An excessive patch dictionary was used to track the high quality residual picture through the sparse demonstration. (3) An imagined face picture was achieved by joining the first two stages. The proposed algorithm acquired HR image in the presence of small training image pairs. Published results showed a high PSNR on FERET database. In [58], the authors used monogenic signals, phase-quadrant encoding, multiple kernel learning, and the LBP for efficient recognition. Experiments conducted on Extended Cohn-Kanade and Oulu-CASIA face databases reported 92% and 80%, respectively. In [59], recognition was examined in unconstraint environment with multiple cameras using dynamic Bayesian network. The proposed scheme was confirmed in an open surveillance video dataset with a three-camera arrangement. Results were compared with different benchmark classifiers. Published results had an accuracy of 73.6%. In [60], the authors introduced Spatio-Temporal Texture Map (STTM) for spontaneous facial recognition. Experiments conducted on CASME II database reported up to 98.43% accuracy. Specifically, for disgust, happiness, repression, sadness, surprise, and without eyeglasses an average accuracy of 100%, 100%, 96.15%, 100%, 96%, and 91.71% was reported. Authors in [61] presented a novel video-based algorithm to recognize facial expressions. The approach consisted of face detection and face registration in video frames, while the FR was performed using linear SVM. The proposed algorithm performed exceptionally well on unseen data. For anger, disgust, and surprise, algorithm yielded 82%, 88%, and 97% accuracy, respectively on FERA database. Kamgar and Lawson developed a NN-based approach, imitating human perceptual ability of recognizing faces [62]. Proposed approach was based on identification of facial regions that belonged to target subject in face space. Recognition was accomplished by producing two groups of marginal pictures, envisioning them inside and outside of the well-defined conclusion section. A devoted classifier was trained for each person on the watch-list. Investigations were conducted on alive system with humans in actual surroundings. Reported results had accuracy up to 98.5%. More recent advances in FR dealing with video sequences can be seen in Refs. [63] and [64]. Table 4 summarizes comparison of various techniques used for videos FR.

Table. 4. Summary of face recognition and detection schemes using video or image database.

Ref	Issues addressed					Dataset Used	Accuracy
	Pose	Illumination	Resolution	Frontal	Occlusion		
[50]	No	No	No	Yes	No	2d and 3D face database	98-100%
[51]	No	No	No	Yes	No	Yale B database and the CMU-PIE database	85.8-91.56
[52]	Yes	No	No	No	No	CMU-PIE, FERET and one eastern database CAS-PEAL.	90.6-100
[53]	No	No	No	Yes	No	FERET data set	91.5-99.6
[54]	Yes	No	No	No	No	West Virginia University(video dataset)	66.7-79.5
[55]	Yes	No	No	Yes	No	XM2VTS database	99.30
[56]	Yes	No	No	No	No	BU-4DFE database	92.22
[57]	No	No	No	Yes	No	CAS-PEAL,FERET and CMU	32.14(PSNR)
[58]	Yes	Yes	No	No	No	Extended Cohn-Kanade and Oulu-CASIA	80-92
[59]	No	Yes	Yes	Yes	No	ChokePoint datase	73.60
[60]	Yes	No	No	No	No	CASME II database	91.71–98.43
[61]	Yes	No	No	Yes	No	FERA2011/GEMEP-FERA	82-97
[62]	Yes	No	No	Yes	No	Yale, Rice, PIE-CMU, and FERET	97–98.50
[63]	Yes	No	No	No	Yes	ORL, FERET and AR	96.50
[64]	No	No	Yes	No	No	USB recorded video clips	94.59

Below given figure shows a classification of face recognition techniques.

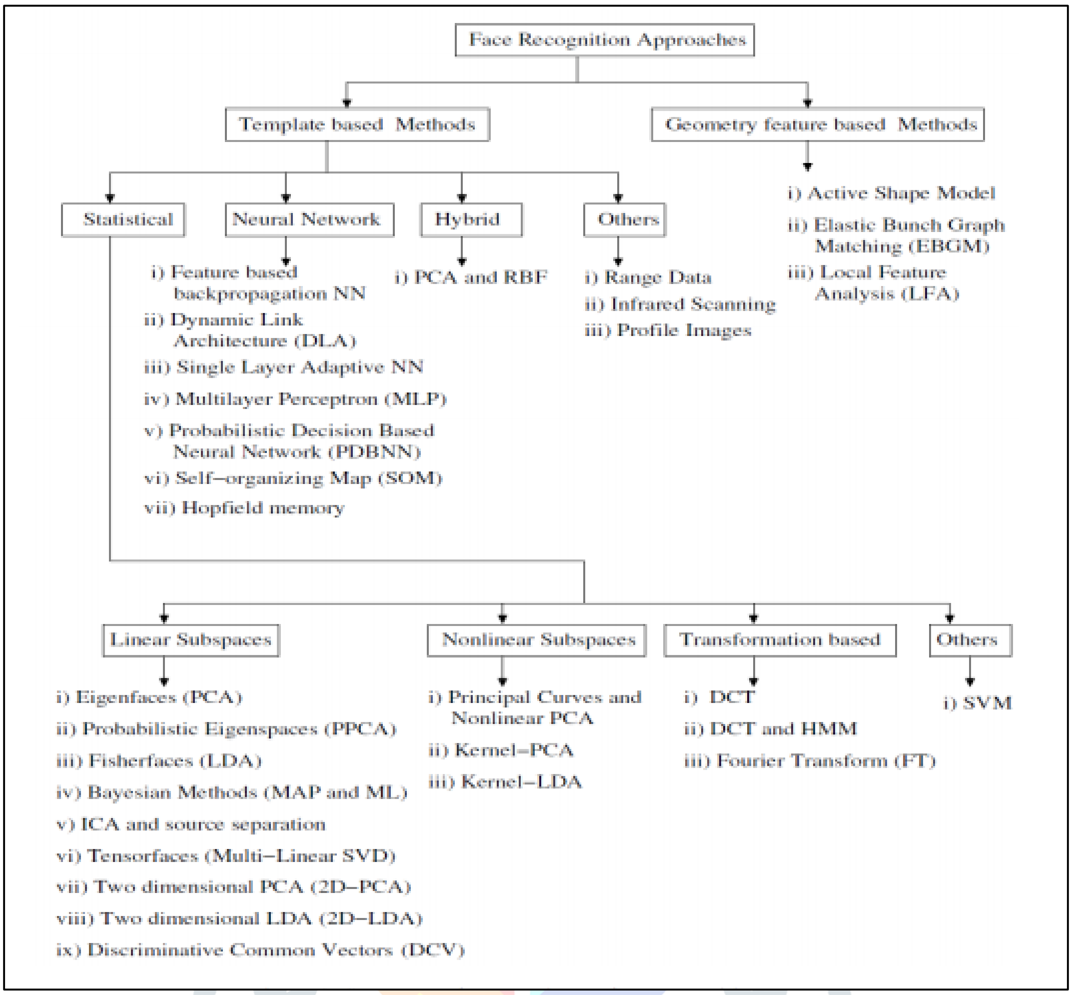

Fig.4. Face recognition techniques

Further, we discuss about the recent techniques of face recognition in video database which has more complexity when compared with the face recognition from still images. Parkhi et al. [42] performed face recognition on a single image taken from a group of faces traced in a video. The accessibility of extremely huge scale training datasets and the end to end learning through CNN lead to improvement the area of face recognition. Initially the authors have demonstrated the assembly of extremely big scale dataset (2.6M pictures, of around 2.6K humans) through a automatic amalgamation of people in the iteration. They have discussed about the negotiation among data clarity and time. Next they have navigated across the difficulties of profound network preparation and face recognition to represent procedures that can attain analogous condition of the consequences for the regular LFW and YTF face standards.

To solve the issue of video related face recognition Huang et al. [43] have established an original investigation procedure using the multiple Grassmann. They have used the Fisher LDA type model to study the projection metric through mapping of data from the initial Grassmann multiple to a fresh distinguish data. The proposed method also works to reduce the dimension of the object and uses it as a learning process. They have conducted experiments to demonstrate the efficiency of their method for various difficult datasets on video based face identification.

By collaboration of 2-class classifiers, De-la-Torre et al. [44] have suggested a flexible MCS for video-to-video FR which model the face of every target entity. The proposed model incorporates knowledge from a face chaser and person specific collections for vigorous Spatio-temporal acknowledgment as well as well-organized self- renew of facet prototype. The tracker describes a face curve for all objects that acts in a video. Spatio-temporal FR takes place when the figures of non-negative projections gather beside a curve outstrip the recognition level for precise object collaboration. An advanced updated threshold permits the model to conclude if the curve integrates ample certainty for self-regulate of face prototypes. To renew a facial prototype, all target sections mined from the curve are united with non-target sections chosen from the associate and worldwide prototypes. Facial prototypes are apprised by a study and mix approach to evade information degeneracy which may surface through self-regulate via a learning classifier. A memory organization approach the Kullback–Leibler deviation is exploited to mark and chose the most appropriate object and non-object orientation ROI samples for authentication.

Table 5. Face recognition performance

Reference	Database (Image/Video)	Pre/Post-processing	Segmentation	Feature	Classifier
[51]	Yale B and the	Image cropping	NA	Haar-Like	Ada-Boost

	CMU-PIE databases.	and angle compensation		Features	
[52]	CMU-PIE, FERET, CAS-PEAL	Not applied	Image rendering	Image normalization, Sparse modeling	Gabor PCA +LDA
[54]	Face Recognition Grand Challenge version 2.0	NA	NA	2D-Gabor and ridges	Similarity measurement
[56]	BU-4DFE database	Image resampling and noise removal	NA	Spatio-temporal shape features	Hidden Markov Model
[59]	Chokepoint database	NA	NA	LBP,LPQ and HoG	Dep Belief network
[60]	CASME II	NA	NA	Volume Local Binary Pattern (VLBP) and Local Binary Pattern from Three Orthogonal Planes	SVM
[62]	Yale, Rice, PIE-CMU, and FERET	NA	NA	Texture and Active Shape Models	Neural Net Based
[63]	ORL, FERET, and AR face databases	NA	NA	PCA model	Nearest neighbour classifier
[43]	YouTube Celebrities (YTC), YouTube Face (YTF) and Point-and-Shoot Face Recognition Challenge (PaSC)	Histogram equalization	NA	Gray feature, intensity feature	PLSbased Covariance Discriminant Learning (CDL) and Localized Multi-Kernel Metric Learning (LMKML)
[44]	Carnegie Mellon University Face in Action (FIA) database	NA	Yes	PCA based feature	Adaptive multiple classifier systems
[46]	Honda/UCSD, CMU Mobo and YouTube Celebrities	Histogram Equalization	NA	LBP and intensity features	Dictionary leaning and low-rank approximation

On the basis of position based face static pictures or mug-shots, Dewan et al. [45] have proposed a system for still-to-video FR which pursues to identify the occurrence of an object. Owing to disparity in seizure circumstances (posture, scale, brightness, fuzziness, appearance and camera setting), the model confronts numerous experiments in visual surveillance appliance. Other than these complexities, some position based stills are accessible in enrolment to plan illustrative facial prototypes of target objects. The still-to-video FR based systems should depend on modification, manifold face demonstration, or artificial production of position stills to improve the intra-class inconsistency of facial prototypes. Some FR models merely compare extreme quality faces caught in a video and it also lessens the likelihood of perceiving target objects. They have used AAMT to progressively study a track facial prototype for all objects showing in the sight. The Successive Karhunen–Loeve procedure has been utilized for online knowledge of the trace facial prototypes in an element filter related face tracer. These prototypes are coordinated across consecutive frames compared to the position static pictures of all target objects registered in the system. After this the complementary scores are gathered across numerous frames for vigorous space and time based identification. A target object is identified if scores gathered for a trace facial prototype across a static time exceed few conclusion threshold.

Xu et al. [46] have planned an organized vocabulary learning structure for video related based face identification. The authors have determined the changing structural data through numerous videos of all objects. To maintain the different configuration of face pictures in videos, vocabulary learning and low-rank approximation has been employed. The knowledgeable vocabulary is both discriminative as well as renovate. Therefore, they have reduced the up gradation inaccuracy for entire set of the face images and also motivated a sub- vocabulary to characterize the analogous substance from dissimilar videos. Furthermore, with the introduction of low-rank estimate, the projected technique is capable of discovering changing organized facts from dissimilar videos of the similar substance.

Hamedani et al. [47] have worked on a system which can recognize a person even if he is revolving his head in short span of time. The neural network (NN) based prototypes were used to mine high-dimensional video space and video frames were extracted from them. It has enhanced the detection rate in contrast to a straightforward NN design. These prototypes were stimulated by various explanations of the mind's pictorial observation. The posture and human multiples were disjointed through the neurons expertise in the concealed or blockage layer of the system.

The video related human re-identification have been suggested by McLaughlin et al. [48] through a repeated neural network frame. The features of a person are mined from every frame of a provided video series through a CNN that integrates a recurring concluding layer. The layer permits knowledge flow among time-steps. Temporal pooling is used to combine features from all time-steps and it provides a complete presence for the entire structure.

The static-to-video FR is useful in watch-list broadcasting applications. Bashbaghi et al. [49] have proposed a vigorous MCS method for FR based applications. To prototype a sole allusion still of object characters, they have developed a person specific Ee-SVMs. When numerous arbitrary subspaces are produced for dissimilar face descriptors mined from facet spots, an innovative collaborative associated learning is explored to provide cooperative assortment. Contrary to the traditional RSM which pick up all attribute subspaces arbitrarily from an entire ROI, the semi-arbitrary subspaces are utilized to count the dissemination of facet descriptors as well as to establish the confined spatial relation amongst all spots. Additionally, through numerous training patterns an unsupervised DA process train an e-SVM classifiers of the ED. To transmit information from the ODs, the video ROIs of non-object persons are utilized instead of a sole still ROI. The authors have also validated the set of non-object faces mined from static and video routes of unidentified persons in the OD. They also investigated the effect of exhausting numerous training patterns for DA. Therefore, these systems are able to integrate information of the ODs and these can increase the sturdiness in counter to numerous problematic issues recurrently detected in visual surveillance operative surroundings.

Chen et al. [66] have suggested an operative prototype to resolve face recognition related issues. They have followed both approaches, the visual sense (face images) and audio sense (speech). They have proposed HOG-TOP which can mine active surfaces from video series to illustrate the changes in facial appearance. To pick up the changes in facial configuration, they have developed a new operative geometric attribute from the distortion conversion of facial markers. They have employed the manifold attribute synthesis to solve the visual related face appearance identification issues in lab restrained atmosphere and also in the open air.

A face and body association (FBA) method for video-based face recognition was represented by Kim et al. [67]. In multiple continuous shots, several similar subjects appear, therefore the FBA method track and relate the resemblance of these subjects across consecutive frames. The concluding retrieved track is considered to be a representative of face in the image.

A context-aware local binary feature learning (CA-LBFL) process for face identification was proposed by Duan et al. [68]. Contrasting to the prevailing discriminant face descriptor (DFD) and compact binary face descriptor (CBFD), the CA-LBFL utilizes the background info of contiguous bits via restraining the figure of moves from dissimilar binary bits. As a result more vigorous info may be developed for face depiction. They have taken a face picture and then mined pixel difference vectors (PDV) in regional patches. Afterwards they studied a distinguish mapping in an unconfirmed way to plan every PDV into a background attentive binary vector. Finally, they accomplished grouping for the studied binary codes to build a codebook. Then they have mined a histogram attribute for all face shots by the studied codebook by means of the concluding representation. They have developed CA-LBMFL to utilize the local info from various scales.

On the basis of sparse demonstration Su et al. [74] have projected error free video face recognition through posture disparity in video (RVPose). The face are aligned according to the sparse representation and then identified. The RVPose lines up a series of faces with posture disparity through many posture faces of the same entity keeping the identical surface and 3-D figure. This alignment is decreased to a 3-D figure controlled video alignment challenge. Conclusively, associated video series is accepted on the basis of sparse outline.

2. Conclusion

In the current research graft we have anticipated a wide survey of human face recognition through machine and a concise analysis of associated psychological reports. Two types of face recognition tasks have been taken into account: one from video and the other from static pictures. We have classified all kinds of procedures employed and also study their advantages, disadvantages and features. We have also summarised the existing methods improvements and challenges face by them. Two significant problems have been recognized in real time face recognition structures: the posture and illumination issues. Hereby we represent a brief work of our survey, along with conclusions:

- Mechanical recognition of faces is an emerging and popular research field like pattern identification, image handling, computer vision, and neural networks. It is also used in several commercial appliances like face biometric supported ATM and access mechanism, criminal capture and traffic management applications etc. Although there are many available powerful biometric methods like fingerprint investigation and iris tests but owing to its simplicity and ease of use, face recognition also proved an influential method of personal identification.
- There are different image intensities based face recognition methods are available in the literature. Although these methods are very useful but they also have pros and cons. However the selection of the method is done according to the particular requests of a given application.
- Generally videos are recorded in uncontrolled atmosphere; therefore the videos are of low quality, it becomes difficult to detect a face from a series of video. Sometimes the humans are shy or not much interested in recording, it makes it difficult to identify the face and to obtain good quality pictures. Manifold cues based face recognition structures using have emerged fine outcomes in comparatively managed atmosphere.
- In spite many available advanced face recognition procedures accurate face recognition is not much feasible. Basically the key challenges are as follows: lighting, posture, and identification in open air environment. There are several methods available to solve these issues in face recognition procedures. However, few fundamental issues still exist and need to be resolved; like, posture judgement is not tough but it is hard to estimate precise pose of a person. Along with the mentioned issues, there are some other problems also, like identification of a face from a picture captured before many years.

References

1. Rajpoot, Q. M., & Jensen, C. D. (2014, June). Security and privacy in video surveillance: Requirements and challenges. In IFIP International Information Security Conference (pp. 169-184). Springer, Berlin, Heidelberg.
2. Chatzopoulos, D., Bermejo, C., Huang, Z., & Hui, P. (2017). Mobile augmented reality survey: From where we are to where we go. IEEE Access, 5, 6917-6950.
3. Adams, A. A., & Ferryman, J. M. (2015). The future of video analytics for surveillance and its ethical implications. Security Journal, 28(3), 272-289.
4. Mohd Ali, A., & Angelov, P. (2018). Anomalous behaviour detection using heterogeneous data (Doctoral dissertation, Lancaster University).
5. Pagano, C., Granger, E., Sabourin, R., Marcialis, G. L., & Roli, F. (2014). Adaptive ensembles for face recognition in changing video surveillance environments. Information Sciences, 286, 75-101.
6. Hu, W., Tan, T., Wang, L., & Maybank, S. (2004). A survey on visual surveillance of object motion and behaviours. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34(3), 334-352.
7. Sivic, J., Everingham, M., & Zisserman, A. (2009, June). “Who are you?”-Learning person specific classifiers from video. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 1145-1152). IEEE.
8. Li, T., Chang, H., Wang, M., Ni, B., Hong, R., & Yan, S. (2015). Crowded scene analysis: A survey. IEEE transactions on circuits and systems for video technology, 25(3), 367-386.
9. Wang, Y. K., Fan, C. T., Cheng, K. Y., & Deng, P. S. (2011, July). Real-time camera anomaly detection for real-world video surveillance. In Machine Learning and Cybernetics (ICMLC), 2011 International Conference on (Vol. 4, pp. 1520-1525). IEEE.
10. Niu, W., Long, J., Han, D., & Wang, Y. F. (2004, June). Human activity detection and recognition for video surveillance. In *ICME* (Vol. 1, pp. 719-722).
11. Ozkurt, C., & Camci, F. (2009). Automatic traffic density estimation and vehicle classification for traffic surveillance systems using neural networks. Mathematical and Computational Applications, 14(3), 187-196.
12. Drayer, B., & Brox, T. (2016). Object detection, tracking, and motion segmentation for object-level video segmentation. arXiv preprint arXiv:1608.03066.
13. Fragkiadaki, K., Arbelaez, P., Felsen, P., & Malik, J. (2015). Learning to segment moving objects in videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4083-4090).
14. Zhang, L., Li, S. Z., Yuan, X., & Xiang, S. (2007, June). Real-time object classification in video surveillance based on appearance learning. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on (pp. 1-8). IEEE.
15. Babenko, B., Yang, M. H., & Belongie, S. (2011). Robust object tracking with online multiple instance learning. IEEE transactions on pattern analysis and machine intelligence, 33(8), 1619-1632.
16. J. Phillips, “Video Challenge Problem Multiple Biometric Grand Challenge: Preliminary Results of Version 1,” 2008.
17. P. Phillips, J. Beveridge, B. Draper, G. Givens, A. O’Toole, D. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer, “An Introduction to the Good, the Bad, the Ugly Face Recognition Challenge Problem,” Proc. 2011 IEEE Conference on Automatic Face and Gesture Recognition, pp. 346–353, 2011.
18. Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015). A convolutional neural network cascade for face detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5325-5334).
19. X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2012. 2, 6, 8
20. V. Jain and E. Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Technical Report UMCS-2010-009, University of Massachusetts, Amherst, 2010. 2, 6, 7
21. Comaschi, F., Stuijk, S., Basten, T., & Corporaal, H. (2015, August). Online multi-face detection and tracking using detector confidence and structured SVMs. In Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on (pp. 1-6). IEEE.
22. S. Duffner, P. Motlicek, and D. Korchagin. The ta2 database: A multi-modal database from home entertainment. In ICSAP, 2011.
23. Liao, S., Jain, A. K., & Li, S. Z. (2016). A fast and accurate unconstrained face detector. IEEE transactions on pattern analysis and machine intelligence, 38(2), 211-223.
24. http://mplab.ucsd.edu , “The MPLab GENKI Database, GENKI-SZSL Subset.”
25. H. Rowley, S. Baluja, and T. Kanade, “Rotation invariant neural network-based face detection,” in IEEE International Conference on Computer Vision and Pattern Recognition, 1998
26. Rasti, P., Uiboupin, T., Escalera, S., & Anbarjafari, G. (2016, July). Convolutional neural network super resolution for face recognition in surveillance monitoring. In International conference on articulated motion and deformable objects (pp. 175-184). Springer, Cham.
27. Jairath, S., Bharadwaj, S., Vatsa, M., & Singh, R. (2016). Adaptive skin color model to improve video face detection. In Machine Intelligence and Signal Processing (pp. 131-142). Springer, New Delhi.
28. Yin, S., Ouyang, P., Dai, X., Liu, L., & Wei, S. (2017). An adaboost-based face detection system using parallel configurable architecture with optimized computation. IEEE Systems Journal, 11(1), 260-271.
29. Elrefaei, L. A., Alharthi, A., Alamoudi, H., Almutairi, S., & Al-rammah, F. (2017, March). Real-time face detection and tracking on mobile phones for criminal detection. In Anti-Cyber Crimes (ICACC), 2017 2nd International Conference on (pp. 75-80). IEEE.

30. Haghighat, M., & Abdel-Mottaleb, M. (2017, May). Low resolution face recognition in surveillance systems using discriminant correlation analysis. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on* (pp. 912-917). IEEE.
31. Dadi, H. S., Pillutla, G. K. M., & Makkena, M. L. (2017). Face Recognition and Human Tracking Using GMM, HOG and SVM in Surveillance Videos. Annals of Data Science, 1-23.
32. Yu, M., Yun, L., Chen, Z., & Cheng, F. (2017, June). Research on video face detection based on AdaBoost algorithm training classifier. In Electronics Instrumentation & Information Systems (EIIS), 2017 First International Conference on (pp. 1-6). IEEE.
33. Mutneja, V., & Singh, S. (2017). Modified Viola–Jones algorithm with GPU accelerated training and parallelized skin color filtering-based face detection. Journal of Real-Time Image Processing, 1-21.
34. Ding, C., & Tao, D. (2018). Trunk-branch ensemble convolutional neural networks for video-based face recognition. IEEE transactions on pattern analysis and machine intelligence, 40(4), 1002-1014.
35. Liu, S., Kong, L., & Wang, H. (2018, November). Face Detection and Encryption for Privacy Preserving in Surveillance Video. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV) (pp. 162-172). Springer, Cham.
36. Singh, M., & Sahran, R. (2018). A New Approach for Suspect Detection in Video Surveillance. In Information and Communication Technology for Sustainable Development (pp. 433-441). Springer, Singapore.
37. Chawla, D., & Trivedi, M. C. (2018). Difference in Lights and Color Background Differentiates the Color Skin Model in Face Detection for Security Surveillance. In Networking Communication and Data Knowledge Engineering (pp. 127-135). Springer, Singapore.
38. Mutneja, V., & Singh, S. (2018). GPU accelerated face detection from low resolution surveillance videos using motion and skin color segmentation. Optik-International Journal for Light and Electron Optics, 157, 1155-1165.
39. Sharma, R., Ashwin, T. S., & Guddeti, R. M. R. (2019). A Novel Real-Time Face Detection System Using Modified Affine Transformation and Haar Cascades. In Recent Findings in Intelligent Computing Techniques (pp. 193-204). Springer, Singapore.
40. Owusu, E., Abdulai, J. D., & Zhan, Y. (2019). Face detection based on multilayer feed-forward neural network and Haar features. Software: Practice and Experience.
41. Yang, W., Zhou, L., Li, T., & Wang, H. (2019). A Face Detection Method Based on Cascade Convolutional Neural Network. Multimedia Tools and Applications, 1-18.
42. Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015, September). Deep face recognition. In BMVC (Vol. 1, No. 3, p. 6).
43. Huang, Z., Wang, R., Shan, S., & Chen, X. (2015). Projection metric learning on Grassmann manifold with application to video based face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 140-149).
44. De-la-Torre, M., Granger, E., Radtke, P. V., Sabourin, R., & Gorodnichy, D. O. (2015). Partially-supervised learning from facial trajectories for face recognition in video surveillance. Information Fusion, 24, 31-53.
45. Dewan, M. A. A., Granger, E., Marcialis, G. L., Sabourin, R., & Roli, F. (2016). Adaptive appearance model tracking for still-to-video face recognition. Pattern recognition, 49, 129-151.
46. Xu, H., Zheng, J., Alavi, A., & Chellappa, R. (2016, March). Learning a structured dictionary for video-based face recognition. In Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on (pp. 1-9). IEEE.
47. Hamedani, K., Seyyedsalehi, S. A., & Ahamdi, R. (2016). Video-based face recognition and image synthesis from head frames using nonlinear manifold learning by neural networks. Neural Computing and Applications, 27(6), 1761-1769.
48. McLaughlin, N., Martinez del Rincon, J., & Miller, P. (2016). Recurrent convolutional network for video-based person re-identification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1325-1334).
49. Bashbaghi, S., Granger, E., Sabourin, R., & Bilodeau, G. A. (2017). Dynamic ensembles of exemplar-svms for still-to-video face recognition. Pattern recognition, 69, 61-81.
50. Z. Hailing, M. Ajmal and W. Lei, Recent advances on singlemodal and multimodal face recognition: A survey, IEEE Trans. Human Mach. Syst. 44(6) (2014) 701–716.
51. D. J. Kim and S. H. Lee, Face recognition in unconstrained environments, Int. Conf. Consumer Electronics (2012), pp. 143–144.
52. L. Ding and X. Ding, Continuous pose normalization for pose-robust face recognition, IEEE Signal Process. Lett. 19(11) (2012) 721–724.
53. Thanh et al., Face recognition using multi-modal binary patterns, 21st Int. Conf. Pattern Recognition (2012), pp. 2343–2346.
54. H. Mohammad and M. Mohamed, A multimodal approach for face modeling and recognition, IEEE Trans. Inform. Forensics Security 3(3) (2008) 431– 440
55. X. Tang and L. Zhifeng, Audio-guided video-based face recognition, IEEE Trans. Circuits Syst. Video Technol. 19(7) (2009) 955–964.
56. Vuong et al., Expression recognition from 3D dynamic faces using robust spatiotemporal shape features, Int. Conf. Automatic Face and Gesture Recognition and Workshops (2011), pp. 414–421.
57. M. Mohamed, Effective two-step method for face hallucination based on sparse compensation on over-complete patches, IET Image Process. 7(6) (2013) 624–632
58. Xiaohua et al., Spatiotemporal local monogenic binary patterns for facial expression recognition, IEEE Signal Process. Lett. 19(5) (2012) 243– 246.

59. L. An, M. Kafai and B. Bhanu, Dynamic Bayesian network for unconstrained face recognition in surveillance camera networks, IEEE J. Emerg. Sel. Topics Circuits Syst. 3(2) (2013) 155–164.
60. Siti et al., Spatio-temporal texture-based feature extraction for spontaneous facial expression recognition, 14th IAPR International Conference on Machine Vision Applications (MVA) (2015), pp. 467–470.
61. S. Yang and B. Bhanu, Understanding discrete facial expressions in video using an emotion avatar image, IEEE Trans. Syst. Cybern. 42(4) (2012) 980–992
62. B. Kamgar and W. Lawson, Toward development of a face recognition system for watchlist surveillance, IEEE Trans. Pattern Anal. Mach. Intell. 33(10) (2011) 1925–1937.
63. X. Yong and D. Zhang, A two-phase test sample sparse representation method for use with face recognition, IEEE Trans. Circuits Syst. Video Technol. 21(9) (2011) 1255–1262.
64. Y. Kim, J. H. Yoo and K. Choi, A motion and similarity-based fake detection method for biometric face recognition systems, IEEE Trans. Consumer Electron. 57(2) (2011) 756–762.
65. Ranjan, R., Patel, V. M., & Chellappa, R. (2019). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(1), 121-135.
66. Chen, J., Chen, Z., Chi, Z., & Fu, H. (2018). Facial expression recognition in video with multiple feature fusion. IEEE Transactions on Affective Computing, 9(1), 38-50.
67. Kim, K., Yang, Z., Masi, I., Nevatia, R., & Medioni, G. (2018, March). Face and Body Association for Video-Based Face Recognition. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 39-48). IEEE.
68. Duan, Y., Lu, J., Feng, J., & Zhou, J. (2018). Context-aware local binary feature learning for face recognition. IEEE transactions on pattern analysis and machine intelligence, 40(5), 1139-1153.
69. Phillips, P.J., Moon, H., Rizvi, S., Rauss, P.J., et al.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. 22(10), 1090–1104 (2000).
70. Head pose image database. http://www-prima.inrialpes.fr/perso/Gourier/Faces/HPDatabase.html
71. Collection of facial images Faces94. http://cswww.essex.ac.uk/mv/allfaces/ faces94.html
72. M. Grgic, K. Delac, and S. Grgic, “SCface–surveillance cameras face database,” Multimedia tools and applications, vol. 51, no. 3, pp. 863– 879, 2011
73. P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, “Overview of the face recognition grand challenge,” in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, 2005, pp. 947–954.
74. Su, Y. (2018). Robust video face recognition under pose variation. Neural Processing Letters, 47(1), 277-291.
75. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology (2007)
76. Yang, S., Luo, P., Loy, C.C., Tang, X.: Wider face: a face detection benchmark. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
77. J. R. Beveridge, J. Phillips, D. S. Bolme, B. Draper, G. H. Givens, Y. M. ui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer et al., “The challenge of face recognition from digital point-and-shoot cameras,” in Proc. IEEE Int. Conf. Biometrics, Theory, Appl. Syst., 2013, pp. 1–8
78. Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kuerban, and X. Chen, “A benchmark and comparative study of video-based face recognition on cox face database,” IEEE Trans. Image Process., vol. 24, no. 12, pp. 5967–5981, 2015
79. L. Wolf, T. Hassner, and I. Maoz, “Face recognition in unconstrained videos with matched background similarity,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2011, pp. 529–534
80. http://vision.ucsd.edu/content/yale-face-database
81. Lu, Z., Jiang, X., & Kot, A. (2018). Color Space Construction by Optimizing Luminance and Chrominance Components for Face Recognition. Pattern Recognition.
82. Jang, Y., Gunes, H., & Patras, I. (2017). SmileNet: registration-free smiling face detection in the wild. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1581-1589).
83. Kishore Kumar, K., & Pavani, M. (2018). Periocular Region-Based Age-Invariant Face Recognition Using Local Binary Pattern. Microelectronics, Electromagnetics and Telecommunications, 713–720.
84. Fydanaki, A., & Geradts, Z. (2018). Evaluating OpenFace: an open-source automatic facial comparison algorithm for forensics. Forensic Sciences Research, 1–8.