Face recognition technologies for evidential evaluation of video traces

Xingjie Wei and Chang-Tsun Li

Abstract Human recognition from video traces is an important task in forensic investigations and evidence evaluations. Compared with other biometric traits, face is one of the most popularly used modalities for human recognition due to the facts that its collection is non-intrusive and requires less cooperation from the subjects. Moreover, face images taken at a long distance can still provide reasonable resolution, while most biometric modalities, such as iris and fingerprint, do not have this merit. In this chapter, we discuss automatic face recognition technologies for evidential evaluations of video traces. We first introduce the general concepts in both forensic and automatic face recognition, then analyse the difficulties in face recognition from videos. We summarise and categorise the approaches for handling different uncontrollable factors in difficult recognition conditions. Finally we discuss some challenges and trends in face recognition research in both forensics and biometrics. Given its merits tested in many deployed systems and great potential in other emerging applications, considerable research and development efforts are expected to be devoted in face recognition in the near future.

1 Introduction

The UK currently has the most widely deployed CCTV coverage in the world. In 2013, the British Security Industry Authority (BSIA) estimated that there are up to 5.9 million CCTV in the UK equating to 1 camera for every 11 people. With increasing emphasis on national and global security, there is a growing and acute need for

Xingjie Wei
School of Computing Science, Newcastle University, Newcastle upon Tyne, NE1 7RU, U.K. e-mail: xingjie.wei@ncl.ac.uk

Chang-Tsun Li
Department of Computer Science, University of Warwick, Coventry, CV4 7AL, U.K. e-mail: C-T.Li@warwick.ac.uk

human recognition (e.g., identifying or searching victims/witnesses/suspects) from videos.

Biometrics is the science of identifying an individual based on the physiological and behavioural characteristics. The physiological characteristics are related to the shape of the body including face, iris, retina, fingerprint, palmprint, palm vein, hand geometry, DNA, earlobe, etc. The behavioural characteristics are related to the pattern of behaviour of a person such as gait, signature, keystroke dynamics, voice, etc. Among these biometric traits, face is the most commonly seen and used for human recognition due to the fact that its collection is non-intrusive and requires less cooperation from the subjects. Everyone has a face and it is widely accepted as a means of recognition.

Forensic science is the scientific knowledge and technical methods in gathering and examining traces which might be useful for the investigations of crime. The biometric traces are of great interest for both biometrics and forensics research [40]. A large number of biometric applications such as fingerprint identification systems and computerised DNA databases, were implemented to serve forensic purposes. Among them, the face recognition system is becoming increasingly important due to the abundance of image and video data provided by surveillance cameras, mobile devices, Internet social networks, etc. The videos collected during an investigation needs to be evaluated by investigators throughout the forensics process. The evaluation is to explore whether additional line of questions can be identified and make sure that the current investigated actions have been completed.

There are broadly two types of evaluation during a forensics process [1]: *investigative evaluation* and *evidential evaluation*. The investigative evaluation concerns 1) *what is known* and 2) *what is unknown*, 3) *what are the consistencies* and 4) *conflicts* in the cases. On the other hand, the evidential evaluation focuses on the *relevance*, *reliability* and *admissibility* of the materials considering issues such as *the overall strength of the case* and *whether sufficient evidence exists against the offender to proceed to charge*.

In a typical forensic face recognition scenario, a forensic expert is given face images of a suspect and a person in question, who may or may not be the suspect. The forensic expert gives a value which represents the degree to which these images appear to be of the same person. When a large amount of images and videos have been gathered, identifying the possible suspects manually is extremely time consuming. Face recognition technologies can be used as a tool in the forensic scenarios for helping with queries against a large enrolled database. The query can be a *one-to-many* search, or a *one-to-one* check, as shown in Figure 1. The West Yorkshire Police in UK has tested a face recognition system for matching CCTV images against a database of mugshots and they have reported that although the technology is not fully mature, it has proven to be a useful investigation tool [46]. We will introduce the face recognition technologies for forensics evaluation of video traces in the rest of this chapter.

---

[1] National Police Library, http://www.college.police.uk/

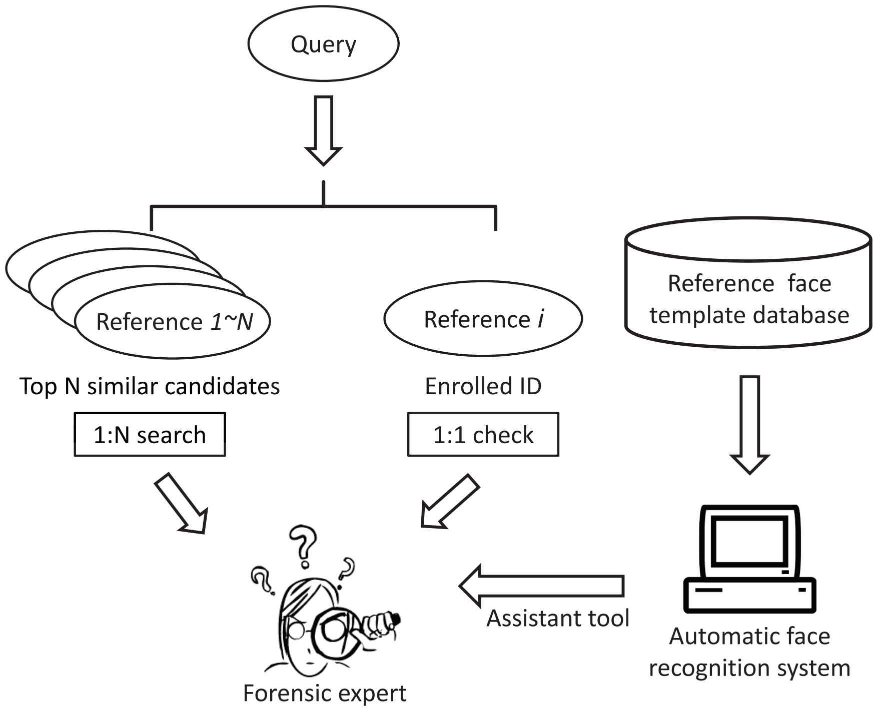

Fig. 1: The process of the forensic face recognition augmented with automatic face recognition system on queries against a large enrolled database. The query can be a one-to-many search or a one-to-one check.

2 Automatic face recognition

A general automatic face recognition system usually consists of the following modules: a face detector, a feature extractor and a matcher, as shown in Figure 2. The face detector first determines which image area contains a face. With the detected face pre-processed, facial features are extracted by the feature extractor and are fed to the matcher as input for comparison against the features of the enrolled images. A similarity score between the query face and each enrolled face is generated by the matcher and used for recognition in either identification mode or verification mode.

2.1 Face detection

The face detector locates and tracks the face area from the background of an image or video frames. Facial components such as eyes, nose, and mouth are located based on the facial landmarks. As shown in Figure 2, this *first-detetion/tracking-then-*

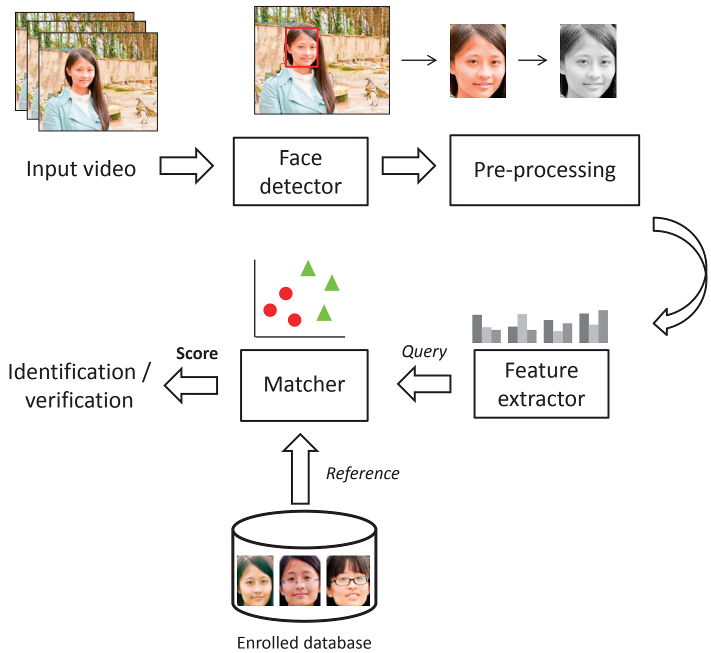

Fig. 2: Basic framework of the automatic face recognition. It usually consists of a face detector, a feature extractor and a matcher. Recognition typically works in identification mode or verification mode.

recognition framework is applied in most image based face recognition systems. For face recognition from videos, approaches based on this scheme first selects key frames (with good image quality, ideal face size, pose, etc.) and then perform detection/tracking and recognition sequentially. On the other hand, in order to deal with the uncertainties in tracking as well as in recognition, some methods [32, 57] perform simultaneous tracking and recognition by using the temporal information in videos. Such *tracking-and-recognition* provide more flexibility for face recognition from videos. It is applicable to the *image-to-video* and *video-to-video* recognition scenarios.

2.2 Feature extraction

Usually pre-processing such as face alignment and normalisation by the facial landmarks are performed before feature extraction. The face area is cropped from an image and normalised according to its geometrical properties (e.g., size and pose) using geometrical transforms or morphing. Usually a detected face area is further normalised with respect to its photometrical properties (e.g., illumination). After that, the feature extractor extracts discriminative information from the face image for distinguishing different individuals. The features can be holistic representations of the face [47, 7], or local patterns around certain salient points of the face [38, 3], depending on the methodology of the matching algorithm. Formulating an effective feature set is a non-trivial task. An effective feature set ideally should be able to characterise the discriminating features in a compact manner. Redundant features not only add little value, but also reduce the power of other useful features. An overly expressive feature set may also lead to the so-called *Curse of Dimensionality*, which requires dimension reduction in order to facilitate efficient matching.

2.3 Matching

The matcher compares two faces according to the extracted features producing a similarity score for the ensuing face recognition to be based on. There are two modes in face recognition: *identification* and *verification*, corresponding to the forensic tasks in Figure 1. In the identification mode, the face recognition system matches the features of the query face to the features of each enrolled reference face template in the database. The templates and query sample can be still images or videos. An ordered list of the *top n* most similar candidates are returned as the possible identities of the query according to the similarity scores. The performance of the system in the identification mode is measured in terms of *rank-n* identification rate which is the rate at which the true association is included in the top *n* matches. The identification rate usually refers to *rank-1* identification rate where the system returns a single match (the best match), as the most probable association with the query face. This rate is also called the *recognition rate*.

On the other hand, verification is the task where the recognition system attempts to confirm an individual's claimed identity by comparing the query sample to the individual's previously enrolled reference templates. Verification is based on a decision threshold which is set by computing the similarity scores of all samples pairs in the database. The threshold is chosen to separate the genuine (i.e., match) similarity scores distribution from the impostor (i.e., non-match) similarity scores distribution and gives the best performance based on one of the following metrics. Here an impostor is a person who submits a sample attempting to claim the identity of another person.

- *False acceptance rate* (FAR) is the rate at which the comparison between two different individuals' samples is erroneously accepted by the system as the true match. In other words, FAR is the percentage of the impostor scores which are higher than the decision threshold.
- *False rejection rate* (FRR) is the percentage of times when an individual is not matched to his/her own existing reference templates. In other words, FRR is the percentage of the genuine scores which are lower than the decision threshold.

- *Equal error rate* (EER) is the rate at which both acceptance and rejection errors are equal (i.e., FAR=FRR). Generally, the lower the EER value, the higher the accuracy of the system.

The first fully automatic face recognition system was presented by Kanade [28] in 1973, which marked a milestone at that time. Over the past four decades, there has been significant progress in automatic face recognition. However, albeit the fact that some unconstrained factors have been considered in the still-image based face recognition, recognising face from videos remains a relatively less explored area due to various factors, such as low resolution, occlusions, lack of subject's cooperation, constantly changing imaging conditions, etc. In the next section we will briefly introduce the challenges in face recognition from videos in the real-world environment.

3 Face recognition from videos traces

During the *London Riots* in 2011, the London Metropolitan Police used automatic face recognition software attempting to find the suspects from video traces, but in many cases it failed [16]. The poor quality of most CCTV footage makes it very difficult to trust standard automatic facial recognition techniques. Changes in illumination, image quality, background and orientation can easily fool the face recognition systems.

In 2013, Klontz and Jain [30] conducted a case study that used the photographs from the CCTV footage of the two suspects in the *Boston Marathon bombings* to match against a background set of mugshots. The suspects' images released by the FBI were captured in the uncontrolled environment and their faces were partially occluded by sunglasses and hats. The study showed that current commercial automatic face recognition systems have the notable potential to assist law enforcement. But the matching accuracy was not high enough, suggesting that further technical progress needs to be made in order to be able to recognise faces from CCTV footage taken from unconstrained environments.

Compared with still images, videos contain more information: e.g., multiple views of the face or different illumination conditions and temporal information. However, face recognition from CCTV footages is an even more challenging problem than recognising faces from still images taken from a short distance. First of all, there are two main difficulties for face tracking [59]: 1) real-world videos (e.g., surveillance videos) are usually of low resolution, which contains less useful information of faces, and 2) illumination changes have a strong impacts on the accuracy of tracking.

In addition, face recognition from videos in real scenarios usually involves the remote face recognition (i.e., *face recognition at a distance* (FRAD)), which is also one of the most challenging problems in face recognition. *Remote face* usually refers to the face data captured at 10m or further away from the cameras. FRAD is often related to the security and defense applications. Compared with the traditional face recognition (e.g., near-distance face recognition), FRAD is faced with two difficulties: 1) less or even no user-cooperation is available, and 2) the environment condition is more difficult to control. The FRAD algorithms need to consider more serious unconstrained variations in face data in the real-world scenarios. These variations are mainly due to two main factors: the user and the environment, as shown in Table 1.

Table 1: Unconstrained variations in face recognition from video traces

Caused by users	Pose Facial expression Occlusion
Caused by environment	Illumination Low resolution Blur (motion blur & focus blur)

The approaches of recognising face from videos can be divided into three categories [6, 9]: 1) *set-based approaches*, *sequence-based approaches*, and *dictionary-based approaches*. The set-based approaches [49] regard a video as a set of unordered images and the recognition robustness is achieved by exploiting the multiple viewing conditions contained in the image set. Traditional still image based methods can be extended to videos image sets. Recognition is performed by fusing the matching results from multiple frame-pairs at different levels (i.e., feature level, score level and decision level), or by measuring the similarity between the query and reference subspaces/manifolds which are learned from image sets. On the other hand, sequence-based approaches explicitly utilise the temporal information between video frames during recognition. This category of approaches exploits the appearance as well as the motion information of faces to obtain better recognition accuracy. To utilize the temporal information, time series state models such as sequential importance sampling (SIS) [56], and Hidden Markov Models [37] are used to model the video sequences. In recent years, matching face images based on sparse representation via a dictionary becomes popular [13]. *Dictionary* here is a large collection of face images where every individual has multiple images containing a wide range of variations such as poses, expressions and illumination changes. Face frames from a given query video are projected onto the sub-space spanned by the atoms in the dictionary and the representation residuals are computed for matching.

4 Handling uncontrollable factors present in videos

Videos provide abundant information in the form of multiple frames and temporal information compared to still images. These information can be exploited for improving the performance of face recognition and provide robustness to large variations in facial poses, expressions and illumination conditions, occlusions and low image quality factors. In the following sub-sections, we introduce face recognition approaches according to the uncontrollable factors they are dealing with. Although some approaches are developed for still images, they do shed light on the new development for video-based systems.

4.1 Approaches for handling pose variations

Forensics experts, or even ordinary people are able to recognise a person by the face from different views. However, due to the complex 3D structures of faces, the view generalisation is non-trivial and often ill-posed task for automatic face recognition systems. The main challenge of pose problem is that appearance variations caused by variable poses of the same person (i.e., intra-class variations) can be larger than those caused by identity differences (i.e., inter-class variations). The similarity between two faces from different persons in the same pose can be larger than that of the same person in different poses.

For pose-invariant face recognition, a natural solution is to collect multi-view images/videos which increase the chances of the face being captured in a favourable frontal pose. This is straight-forward and the frontal face recognition algorithms can be directly extended to solve the non-frontal pose problem. Usually pose estimation, which is the process of inferring the orientation of a head, is needed when processing the multi-view data. Then view selection is performed to select images/frames with ideal view conditions based on the pose information. Li *et al.* [35] estimate the head pose using the SVM regression. The pose information is used to choose the appropriate face detector for multi-view face detection which provides improved performance in terms of accuracy and speed. Only the frontal faces are retained for recognition.

Another direction is to model the pose variations. One of the popular methods is the face manifold. An image of a given size can be viewed as a high-dimensional vector in a Euclidean image space where the dimensionality equals to the number of pixels. The surface and texture of a face is mostly smooth confining it to an embedded face manifold. The basic idea of face manifold based methods is to cluster the images of similar pose and train a linear sub-space to represent each pose cluster. Face manifold is approximated by the linear sub-space model. In this way, two face manifolds of similar poses can be compared under a small variation in pose parameters. Arandjelović and Cipolla [5] propose a method, which first decomposes an appearance manifold to Gaussian pose clusters then fuses the fixed-pose comparison results using a neural network for recognition. Wang *et al.* [49] define the manifold-manifold distance as the distance between the gallery manifold learned from the training gallery image set and the probe manifold learned from the probe image set.

The above methods deal with the pose variations by performing explicit pose estimation or model registration from the multi-view image set or videos. In a surveillance environment, these processes are still very challenging due to the fact that the image quality is poor and the calibration of cameras is not accurate. Du *et al.* propose a pose-insensitive feature [14] which does not require explicitly estimate the pose of the face. The proposed feature is developed using the spherical harmonic representation of the face texture-mapped onto a sphere. The texture map itself is generated by back-projecting the multi-view video data. One limitation of the method is that the pose insensitive feature relies on the assumption that the spherical function remains unchanged other than a rotation. The assumption works in normal illumination conditions but does not hold in extreme lightings.

Another limitation for pure 2D image based methods as described above is that they assume the pose transformation is continuous within the 2D space. On the other hand, approaches with assistance of 3D models [27, 11] which are estimated from videos achieve better performance when addressing pose variations. Compared with 2D image based methods, 3D model based methods usually incur a higher computational cost.

4.2 Approaches for handling occlusion variations

In the real-world environments, faces are easily occluded by facial accessories (e.g., sunglasses, scarf, hat, veil) or objects in front of the face (e.g., hand, food, mobile phone). These occlusions can largely change the appearance of the face, which makes the face detection more difficult. In forensic investigations, although occluded faces can be manually detected and cropped, recognising partially occluded faces images is still challenging for automatic systems.

One intuitive idea is to first determine whether a face image is occluded or not [44], and then reject the occluded images in applications. This rejection mechanism is not always suitable for face recognition especially in forensic scenarios where no alternative image can be obtained due to the lack of subject cooperation. Some approaches first segment the occluded regions from face images and then perform recognition based on the remaining parts [41, 25, 26]. These models require a skin colour based *occlusion mask* to detect the occluded areas in faces. The occlusion detectors are usually trained on specific types of occlusions (i.e., the training is data-dependent) and hence generalise poorly to various types of occlusions in real-world environments.

So performing recognition with the presence of occlusions is very challenging. There are three main categories of approaches for face recognition without detecting occlusion in advance: 1) *reconstruction based approaches*, 2) *local matching based approaches*, and 3) *occlusion-insensitive feature based approaches*.

The first category, reconstruction based approaches treats occluded face recognition as a reconstruction problem [53, 55, 51]. A clean image is reconstructed from an occluded query image by a linear combination of reference images. Then the occluded image is assigned to the class with the minimal reconstruction error. A common drawback of reconstruction based methods is that they usually require a large number of samples per subject to represent a query image. Most of them assume that the reference/training images are captured in well controlled conditions. However, this assumption does not usually hold in real-world scenarios. Another drawback is this category of approaches usually incur a high computational cost [53].

The second category is the local matching based approaches [39, 45, 52, 50]. Facial features are extracted from local areas of a face, for example, overlapping or non-overlapping patches of an image, so the affected and unaffected parts of the face can be analysed in isolation. In order to minimize matching errors due to occluded parts, different strategies such as local space learning [45, 39], multi-task sparse representation learning [36] or voting [50] are performed. The common intuition behind the local matching based approaches is based on that the facial appearance changes caused by occlusions are local in nature. Only part of a face is distorted by occlusions while others are less affected and reliable for recognition. So compared with the reconstruction based methods, local matching based methods are less likely to be able to handle the situation in which more than half of the face is occluded.

In addition to the above approaches, which focus on improving the robustness to occlusions during the matching stage, researchers also pay attention to image representation. The third category is occlusion-insensitive feature based approaches [48, 58] which attempts to extract occlusion-insensitive features from face images. Tzimiropoulos *et al.* [48] pointed out that PCA learning in the gradient orientation domain with a cosine-based distance measure helps reduce the effects due to occlusions in face images. The distribution of image gradient orientation (IGO) differences and the cosine kernel provide a powerful way to measure image correlation/similarity when image data are corrupted by occlusions. The face representations learned from the image gradient orientations are relatively invariant to the occlusion effects. Inspired by their work, Zhu *et al.* [58] further proposed a Gabor phase difference representation for occluded face recognition. They find that the Gabor phase (GP) difference is more stable and robust than gradient orientation to occlusions.

4.3 Approaches for handling illumination variations

Due to the 3D structure and various surface reflectance of faces, light sources can cast shading and shadows, creating non-uniform illumination on faces, which accentuates or diminishes certain facial features. The differences induced by this impact in the facial appearance can be greater than that between individuals.

There are two categories of approaches for addressing the illumination problem - *active* and *passive* ways [60]. The active approaches attempt to obtain face images which are invariant to the illumination changes. Usually specific devices such as 3D scanners, thermal cameras, infrared cameras, etc. other than the visible light cameras are required. A good survey on 3D face recognition can be found in [1]. Thermal images and near-infrared images are more insensitive to large illumination changes as compared to visible light images. Introductions to illumination invariant face recognition using thermal images and near-infrared images are presented in [23] and [34], respectively.

On the other hand, the passive approaches attempt to directly deal with images which have already been affected by illuminations. There are usually three classes of approaches. 1) Illumination normalisation [15, 12], which seeks to suppress the illumination variations either by image transformations or by synthesising an unaffected image from affected ones, 2) illumination invariant representation [17, 3], which attempts to extract features invariant to illumination changes, and 3) illumination variation modelling. Illumination core [19, 8] and sparse representation [53] are based on the theoretical principle that the set of images of a convex Lambertian object [31] obtained in a wide variety of illumination conditions can be approximated by a low-dimensional linear sub-space in which the recognition can be performed. These methods require well-registered face images and the availability of face images with different illumination conditions. Such requirements limit their applicability in practice, especially for videos captured in unconstrained environments. Some models for pose variations are also applicable to solving of the illumination problem. Using the manifold model introduced in Section 4.1, the illumination variation for each of the pose clusters can be modelled using a linear, pose-specific illumination sub-space. Given a reference template and a novel cluster with the same pose, the illumination changes can be normalised by adding a vector from the pose illumination sub-space to the frame so that its distance from the reference cluster is minimized [5].

4.4 Approaches for handling low image quality variations

In security-related face recognition applications such as surveillance, usually the face images/videos captured are degraded by low resolution and blur effects. When sufficient videos are available, one simple idea is to select a set of frames which yield the best recognition accuracy by a classifier. This can help to remove or give lower weight to the poor quality frames during recognition.

For the low resolution problem, another intuitive solution is the super-resolution (SR) based method [4]. SR is a technique for synthesising high-resolution images from low resolution images for visual enhancement. After applying SR, a higher resolution image can be obtained and then used for recognition, for example, matching a face in a low resolution surveillance footage against a set of higher quality gallery sequences enrolled in a database. One common drawback of SR based face recognition approaches is that SR does not directly contribute to recognition. The identity information may be contaminated by some artifacts attributed to the SR process.

Another category of approaches do not apply the SR preprocessing to low resolution images. Li *et al.* [33] proposed a method to learn coupled mappings (CMs), which minimizes the difference between the low-resolution image and its high-resolution counterpart. Then the low resolution image is projected onto a unified feature space where higher recognition performance can be achieved. Biswas *et al.* [10] proposed a method using Multi-dimensional Scaling (MDS) to transform the low resolution gallery and probe images into an Euclidean space such that the distances between them approximates the best distances. Shekhar *et al.* [43] propose a generative approach to low-resolution image based on learning class specific dictionaries, which is also robust to illumination variations.

There are two types of effect attributed to the blur problem: *focus blur* and *motion blur*. A focus is the point where lights originating from a point on the object converge. When the light from object points is not well converged, an out-of-focus image with the blur effect will be generated by the sensor (e.g., camera). The work in [24] analyzed the impact of out-of-focus blur on face recognition performance. On the other hand, motion blur is due to the rapid object movement or camera shaking. Blurring affects the appearance of faces in images, causing two main problems [42]: 1) the appearance of face images from the same individual changes significantly due to blur, and 2) different individuals tend to appear more similar when blurred due to the loss of discriminative features. There are two main categories of approaches to improve the quality of the blurred face images: 1) blurred image modelling through subspace analysis [42] or sparse representation [54], and 2) blur-tolerant descriptors which attempt to extract blur insensitive features such as Local Phase Quantization (LPQ) [2, 22] to represent the face images.

5 Future trends

As introduced in the last section, the appearance variations caused by the unconstrained conditions are still challenging for face recognition from images and videos. This section will discuss several specific face recognition problems, which are the new trends of research in both biometrics and forensics communities.

5.1 Combining with other biometric traits

When faces are heavily occluded or degraded due to extreme conditions, face recognition technologies become ineffective. In unconstrained environments, the face is not the only trait used by humans to recognise each other. It is natural to combine face and other biometric data to improve the recognition performance. These data is either from other modalities such as voice, gait [20], or soft biometric features such as height, gender, hair, skin and clothing colour. The advantages of using such features are: 1) they can be captured without constraint in uncontrollable environments, and 2) they can be captured along with the face using the same sensor such as a CCTV camera. How to represent the features from different modalities and how to fuse these features and matching scores will be the important issues for future investigations.

5.2 Contending with the face ageing issue

Facial ageing is a complex process that affects both the shape and texture (e.g., skin tone or wrinkles) of a face. The typical scenario of face recognition across ageing is to detect if a particular person is present in a previously recorded database. Applications include missing children identification, suspect watch-list check, etc. For still image based recognition, ageing effect has been studied in two directions: 1) developing *age estimation techniques* to classify face images based on age [18, 21] and 2) developing *ageing robust systems* to perform recognition. However, the ageing issue are seldom considered in video based face recognition algorithms. One most challenging aspect of face recognition involving the ageing issue is that it must address all other 'historical' unconstrained variations as well. Pose, expression, illumination changes and occlusions can occur when images are taken years apart.

Compared to still images, videos contain the temporal information which is of great value for face recognition. It is interesting to investigate into the ways of utilizing the temporal information effectively to deal with the ageing issue.

5.3 Different imaging modalities

Face recognition across different imaging modalities, also called *heterogeneous face recognition* [29], is another interesting area for further explorations. It involves matching two face images from different imaging modalities, which is of great practical value in forensic scenarios. The images of suspects may come from various sources, e.g., still images captured from CCTV, footages taken by the police helicopters or images snapped by members of the public. In addition, in some extreme situations, only a particular modality of a face image is available. For example, in night-time environments, infrared imaging may be the only modality for acquiring a useful face image of a suspect. But the mug-shots held by the police are visible band images. Another example is the sketch-photograph matching. When no photograph of a suspect is available, a forensic sketch is often generated according to the description of an eye-witness. Matching sketches against face photographs is very important for forensic investigation. On the other hand, 2D-3D face matching is expected to attract intensive research efforts in the near future since face can be represented by heterogeneous features in the 3D and 2D modalities in the real-world cases.

5.4 Other issues in forensic tasks

The face recognition technologies discussed in previous sections mainly focus on how to improve the recognition accuracy from face images and videos. This is essential in forensic investigation and case linking. Besides that, there are other requirements in other forensic tasks [40]. For example, in forensic identification, such as identifying missing people, besides recognition accuracy, the other challenges lie in the development and management of reference databases. How to increase the integrity, quality and interoperability of the template data with the help of face image processing or analysis technologies is an important issue. For forensic evidence evaluation, the challenges are not only about the development of automatic methods, but also the integration of expert-based and automatic methods into hybrid methods.

6 Summary

The face is one of the most popular biometric traits used in the daily life for human recognition. The widespread use of CCTV cameras for surveillance and security applications have stirred extensive research interests in video based face recognition. Face recognition can play an important role in identifying perpetrators of crime activities as well as missing peoples. Automatic face recognition technology is becoming an indispensable tool for modern forensic investigations.

In this chapter we have introduced the advanced face recognition technologies. The past decades have seen significant progress in automatic face recognition. But the performance of the face recognition from videos taken in unconstrained environments is still unsatisfactory. Uncontrollable illumination, pose changes, low image quality, and occlusions pose acute challenges to face recognition techniques. Therefore, intensive research efforts to contend with these interweaving factors are required in the years to come.

7 Acknowledgement

This work is supported by the EU Horizon 2020 - Marie Skłodowska-Curie Actions through the project Computer Vision Enabled Multimedia Forensics and People Identification (Project No. 690907, Acronym: IDENTITY).

References

1. Andrea F. Abate, Michele Nappi, Daniel Riccio, and Gabriele Sabatino. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885 – 1906, 2007. Image: Information

and Control.
2. T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila. Recognition of blurred faces using local phase quantization. In *International Conference on Pattern Recognition (ICPR)*, pages 1–4, Dec 2008.
3. Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037 – 2041, Dec. 2006.
4. Ognjen Arandjelović and Roberto Cipolla. A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution. In *IEEE International Conference Computer Vision (ICCV)*, pages 1 – 8, Oct. 2007.
5. Ognjen Arandjelović and Roberto Cipolla. A pose-wise linear illumination manifold model for face recognition using video. *Computer Vision and Image Understanding*, 113(1):113–125, 2009.
6. Jeremiah R. Barr, Kevin W. Bowyer, Patrick J. Flynn, and Soma Biswas. Face recognition from video: a review. *International Journal of Pattern Recognition and Artificial Intelligence*, 26(5), 2012.
7. Peter N. Belhumeur, Jo*ã*o P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711 – 720, Jul. 1997.
8. P.N. Belhumeur and D. Kriegman. What is the set of images of an object under all possible lighting conditions? In *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, pages 270 – 277, Jun. 1996.
9. H.S. Bhatt, R. Singh, and M. Vatsa. On recognizing faces in videos using clustering-based re-ranking and fusion. *IEEE Transactions on Information Forensics and Security*, 9(7):1056 – 1068, Jul. 2014.
10. S. Biswas, K.W. Bowyer, and P.J. Flynn. Multidimensional scaling for matching low-resolution facial images. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pages 1 – 6, Sep. 2010.
11. C.D. Castillo and D.W. Jacobs. Using stereo matching with general epipolar geometry for 2d face recognition across pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2298 – 2304, Dec. 2009.
12. T. Chen, W. Yin, Xiang Sean Zhou, D. Comaniciu, and T.S. Huang. Total variation models for variable lighting face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1519 – 1524, Sep. 2006.
13. Yi-Chen Chen, VishalM. Patel, P.Jonathon Phillips, and Rama Chellappa. Dictionary-based face recognition from video. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *European Conference Computer Vision (ECCV)*, volume 7577 of *Lecture Notes in Computer Science*, pages 766–779. Springer Berlin Heidelberg, 2012.
14. Ming Du, A.C. Sankaranarayanan, and R. Chellappa. Robust face recognition from multi-view videos. *IEEE Transactions on Image Processing*, 23(3):1105 – 1117, Mar. 2014.
15. Shan Du and R. Ward. Wavelet-based illumination normalization for face recognition. In *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages II–954–7, Sep. 2005.
16. Niall Firth. Face recognition technology fails to find uk rioters. New Scientist, Aug. 2011.
17. Yongsheng Gao and M. K H Leung. Face recognition using line edge map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):764 – 779, Jun. 2002.
18. Xin Geng, Zhi-Hua Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2234 – 2240, 2007.
19. A.S. Georghiades, P.N. Belhumeur, and D. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643 – 660, Jun. 2001.

20. Yu Guan, Xingjie Wei, Chang-Tsun Li, G.L. Marcialis, F. Roli, and M. Tistarelli. Combining gait and face for tackling the elapsed time challenges. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pages 1 – 8, Sep. 2013.
21. Guodong Guo, Yun Fu, C.R. Dyer, and T.S. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178 – 1188, 2008.
22. A. Hadid, M. Nishiyama, and Y. Sato. Recognition of blurred faces via facial deblurring combined with blur-tolerant descriptors. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1160–1163, Aug 2010.
23. Gabriel Hermosilla, Javier Ruiz del Solar, Rodrigo Verschae, and Mauricio Correa. A comparative study of thermal face recognition methods in unconstrained environments. *Pattern Recognition*, 45(7):2445 – 2459, 2012.
24. Fang Hua, P. Johnson, N. Sazonova, P. Lopez-Meyer, and S. Schuckers. Impact of out-of-focus blur on face recognition performance based on modular transfer function. In *IAPR International Conference Biometrics (ICB)*, pages 85 – 90, Mar. 2012.
25. Hongjun Jia and Aleix M. Martínez. Face recognition with occlusions in the training and testing sets. In *IEEE International Conference Automatic Face and Gesture Recognition (FG)*, pages 1 – 6, Sep. 2008.
26. Hongjun Jia and Aleix M. Martínez. Support vector machines in face recognition with occlusions. In *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, pages 136 – 141, Jun. 2009.
27. Dalong Jiang, Yuxiao Hu, Shuicheng Yan, Lei Zhang, Hongjiang Zhang, and Wen Gao. Efficient 3d reconstruction for face recognition. *Pattern Recognition*, 38(6):787 – 798, 2005. Image Understanding for Photographs.
28. Takeo Kanade. Picture processing system by computer complex and recognition of human faces. In *Doctoral dissertation, Kyoto University*. Nov. 1973.
29. B. Klare and A.K. Jain. Heterogeneous face recognition: Matching NIR to visible light images. In *International Conference on Pattern Recognition (ICPR)*, pages 1513 – 1516, 2010.
30. Joshua C. Klontz and Anil K. Jain. A case study on unconstrained facial recognition using the boston marathon bombings suspects. *Technical Report MSU-CSE-13-4*, 2013.
31. J. Lambert. Photometria sive de mensura et gradibus luminus. *Colorum et Umbrae", Eberhard Klett*, 1760.
32. Baoxin Li and R. Chellappa. A generic approach to simultaneous tracking and verification in video. *IEEE Transactions on Image Processing*, 11(5):530 – 544, May 2002.
33. Bo Li, Hong Chang, Shiguang Shan, and Xilin Chen. Low-resolution face recognition via coupled locality preserving mappings. *IEEE Signal Processing Letters*, 17(1):20 – 23, Jan. 2010.
34. S.Z. Li, RuFeng Chu, Shengcai Liao, and Lun Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):627 – 639, Apr. 2007.
35. Yongmin Li, Shaogang Gong, Jamie Sherrah, and Heather Liddell. Support vector machine based multi-view face detection and recognition. *Image and Vision Computing*, 22(5):413 – 427, 2004.
36. Shengcai Liao, Anil K. Jain, and Stan Z. Li. Partial face recognition: Alignment-free approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5):1193 – 1205, 2013.
37. Xiaoming Liu and Tsuhan Chen. Video-based face recognition using adaptive hidden markov models. In *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–340–I–345 vol.1, Jun. 2003.
38. David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91 – 110, 2004.
39. Aleix M. Martínez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748 – 763, Jun. 2002.

40. Didier Meuwly and Raymond Veldhuis. Forensic biometrics: From two communities to one discipline. In *International Conference of the Biometrics Special Interest Group BIOSIG*, pages 1–12, Darmstadt, Germany, 2012.
41. Rui Min, Abdenour Hadid, and Jean-Luc Dugelay. Improving the recognition of faces occluded by facial accessories. In *IEEE International Conference Automatic Face and Gesture Recognition (FG)*, pages 442 – 447, 2011.
42. M. Nishiyama, A. Hadid, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi. Facial deblur inference using subspace analysis for recognition of blurred faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):838 – 845, Apr. 2011.
43. S. Shekhar, V.M. Patel, and R. Chellappa. Synthesis-based recognition of low resolution faces. In *IEEE International Joint Conference on Biometrics (IJCB)*, pages 1 – 6, Oct. 2011.
44. Markus Storer, Martin Urschler, and Horst Bischof. Occlusion detection for ICAO compliant facial photographs. In *IEEE Conference Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 122 – 129, 2010.
45. Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, and Jun Liu. Face recognition under occlusions and variant expressions with partial similarity. *IEEE Transactions on Information Forensics and Security*, 4(2):217 – 230, Jun. 2009.
46. Alan Travis. Police trying out national database with 750,000 mugshots, MPs told. The Guardian, Mar. 2008.
47. Matthew A. Turk and Alex P. Pentland. Face recognition using eigenfaces. In *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, pages 586 – 591, Jun. 1991.
48. Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. Subspace learning from image gradient orientations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12):2454 – 2466, 2012.
49. Ruiping Wang, Shiguang Shan, Xilin Chen, and Wen Gao. Manifold-manifold distance with application to face recognition based on image set. In *IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Jun. 2008.
50. Xingjie Wei and Chang-Tsun Li. Fixation and saccade based face recognition from single image per person with various occlusions and expressions. In *IEEE Conference Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 70 – 75, 2013.
51. Xingjie Wei, Chang-Tsun Li, and Yongjian Hu. Robust face recognition under varying illumination and occlusion considering structured sparsity. In *International Conference Digital Image Computing Techniques and Applications (DICTA)*, pages 1 – 7, 2012.
52. Xingjie Wei, Chang-Tsun Li, Zhen Lei, Dong Yi, and Stan Z. Li. Dynamic image-to-class warping for occluded face recognition. *IEEE Transactions on Information Forensics and Security*, 9(12):2035 – 2050, Dec. 2014.
53. John Wright, Allen Y. Yang, Arvind Ganesh, Shankar S. Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210 – 227, Feb. 2009.
54. Haichao Zhang, Jianchao Yang, Y. Zhang, N.M. Nasrabadi, and T.S. Huang. Close the loop: Joint blind image restoration and recognition with sparse representation prior. In *IEEE International Conference Computer Vision (ICCV)*, pages 770 – 777, Nov. 2011.
55. Lei Zhang, Meng Yang, and Xiangchu Feng. Sparse representation or collaborative representation: Which helps face recognition? In *IEEE International Conference Computer Vision (ICCV)*, pages 471 – 478, 2011.
56. Shaohua Zhou, Volker Krueger, and Rama Chellappa. Probabilistic recognition of human faces from video. *Computer Vision and Image Understanding*, 91(1-2):214–245, 2003. Special Issue on Face Recognition.
57. S.K. Zhou, R. Chellappa, and B. Moghaddam. Visual tracking and recognition using appearance-adaptive models in particle filters. *IEEE Transactions on Image Processing*, 13(11):1491 – 1506, Nov. 2004.
58. Jianfei Zhu, Dong Cao, Sifei Liu, Zhen Lei, and Stan Z. Li. Discriminant analysis with Gabor phase for robust face recognition. In *IAPR International Conference Biometrics (ICB)*, pages 13 – 18, 2012.

59. W.W. Zou, P.C. Yuen, and R. Chellappa. Low-resolution face tracker robust to illumination variations. *IEEE Transactions on Image Processing*, 22(5):1726–1739, May 2013.
60. Xuan Zou, J. Kittler, and K. Messer. Illumination invariant face recognition: A survey. In *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pages 1 – 8, Sep. 2007.