We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Automated Face Recognition: Challenges and Solutions

Joanna Isabelle Olszewska

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/66013

Abstract

Automated face recognition (AFR) aims to identify people in images or videos using pattern recognition techniques. Automated face recognition is widely used in applications ranging from social media to advanced authentication systems. Whilst techniques for face recognition are well established, the automatic recognition of faces captured by digital cameras in unconstrained, real-world environment is still very challenging, since it involves important variations in both acquisition conditions as well as in facial expressions and in pose changes. Thus, this chapter introduces the topic of computer automated face recognition in light of the main challenges in that research field and the developed solutions and applications based on image processing and artificial intelligence methods.

Keywords: face recognition, face identification, face verification, face authentication, face labelling in the wild, computational face

1. Introduction

Automated face recognition (AFR) has received a lot of attention from both research and industry communities since three decades [1] due to its fascinating range of scientific challenges as well as rich possibilities of commercial applications [2], particularly in the context of biometrics/forensics/security [3] and, more recently, in the areas of multimedia and social media [4, 5].

Face recognition is the field trying to bring an answer to the question: ‘Whose face is it?’ For this purpose, people have natural abilities through their human perceptive and cognitive systems [6], whereas machines need complex systems involving multiple, advanced algorithms and/or large, adequate face databases. Studying, designing and developing such methods and technologies are the domain of automated face recognition (AFR).
AFR could be distinguished further into the computer automated face identification and the computer automated face verification. Hence, on the one hand, automated face identification consists in a one-to-many (1:N) search of a face image among a database containing many different face images in order to answer questions such as ‘Is it a known face?’ [7]. On the other hand, automated face verification is a one-to-one (1:1) search to solve the matter of ‘Is it the face of …?’ search [8].

Moreover, AFR could be the basis to the solution of the ‘Who is in the picture?’ problem, leading to the computer automated face labelling/face naming [9].

The general AFR process is illustrated in Figure 1. Usually, it first applies techniques addressing questions such as ‘Is there a face in the image?’ (face detection) and ‘Where is the face in the image?’ (face location) and next, it handles the computer-automated recognition mechanism itself [10].

Figure 1. Overview of the face detection and recognition processes.

In particular, this chapter is dedicated to the ‘why’ and ‘how’ of the computer-automated face recognition in constrained and unconstrained environments. The remaining parts of this chapter are structured as follows: in Section 2, we describe AFR’s today challenges, while corresponding scientific solutions and industrial applications are presented in Sections 3 and 4, respectively. Section 5 draws up new trends and future directions for automated face recognition performance improvements and evolution.

2. Challenges

The study and analysis of faces captured by digital cameras address a wide range of challenges, as detailed in Sections 2.1–2.7, which all have a direct impact on the computer automated face detection and recognition.

2.1. Pose variations

Head’s movements, which can be described by the egocentric rotation angles, i.e. pitch, roll and yaw [11], or camera changing point of views [12] could lead to substantial changes in face
appearance and/or shape and generate intra-subject face’s variations as illustrated in Figure 2, making automated face recognition across pose a difficult task [13].

Since AFR is highly sensitive to pose variations, pose correction is essential and could be achieved by means of efficient techniques aiming to rotate the face and/or to align it to the image’s axis as detailed in reference [13].

2.2. Presence/absence of structuring elements/occlusions

The diversity in the intra-subject face’s images could also be due to the absence of structuring elements (see Figure 3a) or the presence of components such as beard and/or moustache (see Figure 3b), cap (see Figure 3c), sunglasses (see Figure 3d), etc. or occlusions of the face (see Figure 3e) by background or foreground objects [14].

Thus, face’s images taken in an unconstrained environment often require effective recognition of faces with disguise or faces altered by accessories and/or by occlusions, as dealt by appropriate approaches such as texture-based algorithms [15].

2.3. Facial expression changes

Some more variability in face appearance could be caused by changes of facial expressions induced by varying person’s emotional states [16] which are displayed in Figure 4.

Hence, efficiently and automatically recognizing the different facial expressions is important for both the evaluation of emotional states and the automated face recognition. In particular, human expressions are composed of macro-expressions, which could express, e.g., anger, disgust, fear, happiness, sadness or surprise, and other involuntary, rapid facial patterns, i.e. micro-expressions; all these expressions generating non-rigid motion of the face. Such facial dynamics can be computed, e.g., by means of the dense optical flow field [17].
2.4. Ageing of the face

Another reason of face appearance’s changes could be engendered by the ageing of the human face, and could impact on the entire AFR process if the time between each image capture is significant [18], as illustrated in Figure 5.

To overcome face ageing issue in AFR, methods need to take properly into account facial ageing patterns [18]. Indeed, over time, not only face characteristics such as its shape or lines are modified [19], but other aspects are changing as well, e.g. hairstyle [20].

2.5. Varying illumination conditions

Large variations of illuminations could degrade the performance of AFR systems. Indeed, for low levels of lighting of the background or foreground, face detection and recognition are much harder to perform [21], since shadows could appear on the face and/or facial patterns could be (partially) indiscernible. On the other hand, too high levels of lights could lead to over-exposure of the face and (partially) indiscernible facial patterns (see Figure 6).

Robust automated face detection and recognition in the case of (close-to-) extreme or largely varying levels of lighting apply to image-processing techniques such as illumination normalization, e.g. through histogram equalization [22]; or machine-learning methods involving the actual image global image intensity average value [21].
2.6. Image resolution and modality

Other usual factors influencing AFR performance are related to the quality and resolution of the face image and/or to the set-up and modalities of the digital equipment capturing the face [23]. For this purpose, ISO/IEC 19794-5 standard [24] has been developed to specify scene and photographic requirements as well as face image format for AFR, especially in the context of biometrics. However, real-world situations of face image acquisition imply the use of different photographic hardware, including one or several cameras which could be omnidirectional or pan-tilt-zoom [25], and which could include, e.g. wide-field sensors [25], photometric stereo [26], etc. Cameras could work in the range of the visible light or use infra-red sensors, leading to multiple modalities for AFR [6]. Hence, faces acquired in real-world conditions lead to further AFR challenges.

2.7. Availability and quality of face datasets

Each AFR technology requires an available, reliable and realistic face database in order to perform the 1:N or 1:1 face search within it (see Figure 1). Hence, the quality such as com-
pleteness (e.g. including variations in facial expressions, in facial details, in illuminations, etc.) as well as accuracy (e.g. containing ageing patterns, etc.) and the characteristics (e.g. varying image file format and colour/grey level, face resolution, constrained/unconstrained environment, etc.) of a face dataset are crucial to the AFR process [31]. Moreover, when dealing with face data, people’s consent and privacy should be respected as AFR systems should comply with the Data Protection Act 2010 [32].

For research purpose, several face databases have been developed and are publicly available. Well-established, online face databases are as follows:

- ORL [33] is a 400-picture dataset of 40 distinct subjects, in portable grey map (pgm) format and with a 92 × 112 pixel resolution, 8-bit grey level. Men and women’s faces are taken against a dark homogeneous background, under varying illumination conditions. The subjects are in up-right, frontal position, with variations in face expressions, facial details and poses within ±20% in yaw and roll.

- Caltech Faces [34] dataset consists of 450 jpeg images with a resolution of 896 × 592 pixels. Each image shows the frontal view of a face (single pose) of one out of 27 unique persons, under different lighting, expressions and backgrounds.

- The Face Recognition Technology (FERET) [35] database has been built with 14,126 face images from 1199 individuals, defining sets of 5–11 greyscale images per person. Each set contains mugshots with different facial expressions and facial details, acquired using various cameras and varying lighting.

- BioID Face database [36] has 1521 frontal face images of 23 people. Images of 384 × 286 pixel resolution are in pgm format and have been captured in real-world conditions, i.e. with a large variety of illumination, background and face size.

- Yale face database [37] has 165 greyscale, gif images of 15 individuals. There are 11 images per subject, one per different facial expression or configuration, i.e. left/centre/right-light, with or without glasses and with different expressions.

- Caltech 10,000 web faces [38] have collected 10,524 human faces of various resolutions and in different settings (e.g. portrait images, group of people, etc.) from Google Image. Coordinates of eyes, nose and the centre of the mouth for each frontal face are provided in order to be used as ground truth for face detection algorithms, or to align and/or crop the human faces for AFR.

Some databases contain both 2D and 3D face data, e.g. Face Recognition Grand Challenge (FRGC) dataset [39] recorded such 50,000 un-/controlled images from 4003 subject sessions.

Other datasets have multiple modalities such as XM2VTSDB multi-modal face database [40] which is the Extended M2VTS database. It is a large, multi-modal database captured onto high-quality, digital video. It contains four recordings, each with a speaking head shot and a rotating head shot, of 295 subjects taken over a period of 4 months. This database includes high-quality colour images, 32 kHz 16-bit sound files, video sequences and also a 3D model.
Another multi-modal database is the Surveillance Cameras Face (SCFace) [41] dataset. It has recorded 4160 static human faces of 130 subjects, in the visible and infrared spectrum, in an unconstrained indoor environment, using a multi-camera set-up consisting of five video-surveillance cameras which various qualities mimic real-world conditions.

Recent developments of face databases focus on capturing faces in the wild, i.e. in unconstrained environments. For example, Face Detection Data Set and Benchmark (FDDB) [42] is a dataset of 2845 images, both greyscale and colour ones, with 5171 faces in the wild, which could include occlusions, poses variations, low resolution and out-of-focus faces.

Labelled Faces in the Wild (LFW) [43] database is a popular dataset for studying multi-view faces in an unconstrained environment. It has recorded 13,233 foreground face images; other faces in the images being assimilated to the background. It has targeted 5749 different individuals, which could have one or more images in the database, and presents variations in pose, lighting, expression, background, race, ethnicity, age, gender, clothing, hairstyles, camera quality, colour saturation, focus, etc. Images have a $250 \times 250$ pixels resolution and are in jpeg format; they are mostly in colour, although few are greyscale only.

Some other available face datasets have been designed for specific purposes. Hence, Spontaneous MICro-expression database (SMIC) [44] is used for facial micro-expressions recognition, while the Acted Facial Expression in the Wild (AFEW) database [45], which has semi-automatically collected face images with acted emotions from movies, is dedicated to macro-expression recognition in close-to-real conditions. On the other hand, FG-NET Ageing database (FG-NET) [46] could be applied for age estimation, age-invariant face recognition and age progression.

3. Solutions

Major pattern recognition techniques as well as main machine-learning methods used for AFR systems are presented in Section 3.1, while classic approaches for AFR in still images or video databases/live video streams are mentioned in Section 3.2.

3.1. Face recognition systems

Most of the AFR systems consist in a two-step process (see Figure 8) based firstly on facial feature extraction, as explained in Section 3.1.1, and second, on facial feature classification/matching against an available face database, as mentioned in Section 3.1.2.

3.1.1. Feature extraction

Facial features are representing the face in a codified way which is computationally efficient for further processes such as matching, classification or other machine-learning techniques, in order to perform AFR. On the other hand, computing facial features in an image could serve to detect a face and to locate it within the image, as illustrated in Figure 9.
Facial feature representations could be of different nature from sparse to dense ones, and could be focused on face appearance, face texture or face geometry [15].

Commonly computed facial features are Haar-like features [47] (Figure 10(a, b)); linear binary patterns (LBP) [48] (Figure 10(c)), which have been extended to local directional pattern (LDP) [49] for micro-expressions recognition in particular; edge maps (Figure 10(d)) and their
extension to line edge maps (LEM) [50]; active shape or active contours [51] (Figure 10(e)); SIFT points [52] (Figure 10(f)), etc.

The detected facial features, e.g. with SIFT points usually correspond to some or all elements of the set of facial anthropometric landmarks, i.e., facial fiducial points (FPs) (see Figure 11), which are defined as follows: FP1—top of the head, FP2—right eyebrow right corner, FP3—right eyebrow left corner, FP4—left eyebrow right corner, FP5—left eyebrow left corner, FP6—right eye right corner, FP7—right eye centre of pupil, FP8—right eye left corner, FP9—left eye right corner, FP10—left eye centre of pupil, FP11—left eye left corner, FP12—nose right corner, FP13—nose centre bottom, FP14—nose left corner, FP15—mouth right corner, FP16—mouth left corner, FP17—chin corner, FP18—right ear top corner, FP19—right ear bottom corner, FP20—left ear top corner and FP21—left ear bottom corner [53].

Computer automated face recognition relies on facial features, in the same way forensic examiners focus their attention not only on the overall similarity of two faces regarding their shape, size, etc. [54], but also on morphological comparisons region by region, e.g. nose, mouth, eyebrows, etc. [53]. Some AFR methods evaluate also discriminative characteristics such as the distance from people’s mouth to the nose, nose to eyes, mouth to eyes, etc. [55]. This adds robustness into AFR systems in the case of modification of some facial patterns over the course of time or occlusions.

Once the face is detected/located and the facial features are extracted, actions to crop the face, to correct its alignment by rotating it, etc., could be performed to address the challenges mentioned in Section 2, before passing the facial features into the next stage described in Section 3.1.2.
3.1.2. Feature classification/matching

For the recognition stage itself of the face recognition process, classification is often used as shown in Figure 12. Indeed, it is a machine-learning technique [56] that has the task of first learning and then applying a function that maps the facial features of an individual to one of the predefined class labels, i.e. class 1 (face of the individual) or class 2 (not the face of the individual), leading in this case to a binary classifier. Classifiers could be applied to the entire set of the extracted facial features or to some specific face attributes, e.g. gender, age, race, etc. [57]. More recently, methods like neural networks are used as classifiers [58].

![Figure 12. Overview of the model computation.](image)

On the other hand, some AFR systems use the matching technique that could be applied on facial geometric features or templates [59]. This approach is also useful for multimodal face data [60].

3.2. Examples of methods

Among hundreds of techniques developed in this field [1-10], Sections 3.2.1–3.2.4 explain briefly some well-established methods for automated face recognition.

3.2.1. Eigenfaces

The eigenface approach [61] is a very successful AFR method. It involves pixel intensity features and uses the principal component analysis (PCA) of the distribution of faces, or eigenvectors, which are a kind of set of features characterizing faces’ variations where each face image contributes more or less to each eigenvector. Thus, an eigenvector can be seen as a ghostly face, or eigenface. Recognition of a test face is determined by applying the nearest-
neighbour technique to the probe face projection in the face space [13]. Fisherfaces extend the eigenface approach by using linear discriminant analysis (LDA) instead of PCA [62, 63].

3.2.2. Active appearance models

The active appearance model (AAM) [64] combines shape and texture features; thus it is slower but more robust for AFR than active shape models (ASM). AAM is built as a multi-resolution model based on a Gaussian-image pyramid. For each level of the pyramid, a separate texture model is computed using 400 face images. Each face is labelled with 68 points around the main features, and the facial region is sampled by c. 10,000 intensity values. AFR is performed by matching the test face with the AAM, following a multi-resolution approach that improves speed and robustness of this method [64].

3.2.3. Local binary patterns

In reference [48], local binary patterns (LBP), which are texture features, have been introduced for AFR. In particular, the face image is divided into independent regions where the LBP operator is applied to codify every pixel of each region by thresholding the $3 \times 3$-neighbourhood of each pixel with the centre pixel value and by binarizing it, and then, creating a local texture descriptor with the histogram of the codes for each face region. A global description of the face is formed by concatenating the local descriptors. Next, the nearest-neighbour classifier is used [48]. LBP approach has been widely adopted for AFR, and several enhancements have been proposed, e.g. the local directional patterns (LDP) [49].

3.2.4. SIFT

The discriminative deep metric-learning (DDML) [52] approach for AFR in unconstrained environment uses facial features such as SIFT descriptors and trains a deep neural network as a classifier to learn a Mahalanobis distance metric in order to maximize face’s inter-class variations and minimize face’s intra-class variations, simultaneously [52].

4. Applications

Nowadays, industry integrates cutting-edge, face recognition research into the development of the latest technologies for commercial applications such as mentioned in Sections 4.1–4.2.

4.1. Security

Face recognition is one of the most powerful processes in biometric systems [8] and is extensively used for security purpose in tracking and surveillance [65, 66], attendance monitoring, passenger management at airports, passport de-duplication, border control and high security access control as developed by companies like Aurora [67].

AFR is applied in forensics for face identification [68], face retrieval in still image databases or CCTV sequences [69], or for facial sketch recognition [70]. It could also help law enforcement
through behaviour and facial expression observation [71], lie detection [72], lip tracking and reading [73].

Moreover, AFR is now used in the context of ‘Biometrics as a Service’ [74], within cloud-based, online technologies requiring face authentication for trustworthy transactions. For example, MasterCard developed an app which uses selfies to secure payments via mobile phones [75]. In this MasterCard’s app, AFR is enhanced by facial expression recognition as the application requires the consumer blinks to prove that s/he is human.

4.2. Multimedia

In our today’s life, AFR engines are embedded in a number of multi-modal applications such as aids for buying glasses or for digital make-up and other face sculpting or skin smoothing technologies, e.g. designed by Anthropics [76].

In social media, many collaborative applications within Facebook [77], Google [78] or Yahoo! [79] are calling upon AFR. Applications such as Snapchat require AFR on mobile [80]. With 200 million users of which half of those engage on daily basis [81], Snapchat is a popular image messaging and multimedia mobile application, where ‘snaps’, i.e. a photo or a short video, can be edited to include filters and effects, text caption and drawings. Snapchat has features such as the ‘Lens’, which allows users to add real-time effects into their snaps by using AFR technologies, and ‘Memories’ which searches content by date or using local recognition systems [82].

Other multimedia applications are using AFR, e.g. in face naming to generate automated headlines in Video Google [83], in face expression tracking for animations and human-computer interfaces (HCI) [84], or in face animation for socially aware robotics [85]. Companies such as Double Negative Visual Effects [86] or Disney Research [87] propose also AFR solutions for face synthesis and face morphing for films and games visual effects.

5. Conclusions

Since constraints shape the path for innovative solutions, we focused this chapter on scientific and technical challenges brought by computer automated face recognition, and we explained current solutions as well as potential applications. Moreover, there are a number of challenges ahead and plenty of room for innovations in this field of automated face recognition. In particular, three emerging directions are discussed in Sections 5.1–5.3.

5.1. Deep face

On the one hand, the proliferation of mobile devices such as smartphones and tablets, which are world-wide available for consumers and which allow users to easily record digital pictures, and on the other hand, the outbreak of mobile and web applications, which manipulate and store thousands of pictures, have paved the way to the Big Data, and, among others, to the necessity to analysis large-scale, face databases. This phenomenon has given rise to
questions such as AFR technology scalability and computational power, and it has led to the
development of a new AFR approach called deep face recognition [88], which involves deep-
learning techniques using convolutional neural networks [89], well fitted for big datasets [90].
Indeed, deep face methods are using large databases for training their models, as by biomi-
metics, they rely on the familiarity concept [91], which is based on the fact that more people
are familiar with a person’s face, more easily they recognized his/her face, even in complex
situations like occlusions or low resolution. Moreover, the recent development of the deep face
approach has benefited from progress in parallel computing tools for acceleration and
enhancement of distributed computing techniques for scalability. In particular, for deep face
recognition, graphics processing units (GPUs), which are specialized processors for real-time,
high-resolution 3D graphics, are used as highly parallel multi-core systems for big data [92],
together with the Compute Unified Device Architecture (CUDA), which provides a simple
and powerful platform [93], making easier for specialists in parallel programming to utilize
GPU resources without advanced skills in graphics programming. Since the above-mentioned,
iterative computation consists of local parallel processing, CUDA implementation is employed
for reducing the computation time of the AFR system [93]. However, deep face-based methods
generate themselves further challenges, e.g. face frontalization [94] that is the process of
synthesizing frontal facing views of faces appearing in single unconstrained photos, in order
to boost AFR performance within intelligent systems.

5.2. Wild face

Another challenge that has appeared with the generation of a large amount of visual data
captured ‘in the wild’, i.e. in an unconstrained environment, by commercial cameras is the
automated recognition of faces in the wild. It involves the enhancement of AFR methods [95]
in order they efficiently deal with complex, real-world backgrounds [96], multiple-face scenes
[51], skin-colour variations [97], gender variety [98] and with inherent challenges such as image
quality, resolution, illumination or facial pose correction [23, 27, 99].

5.3. Dynamic face

In the recent years, handling facial dynamics efficiently is crucial for AFR systems, because
people have recorded a large amount of faces as still digital images, e.g. selfies or as video
streams, e.g. CCTV sequences or online movies. Indeed, on the one hand, the different
variations in facial micro/macro expressions [100], which generate fast, facial dynamics and
the different processes such as ageing, which is an extremely slow, dynamic problem since the
face evolves over large periods of time [18], have all an impact on AFR techniques. On the other
hand, face acquisition in videos intrinsically creates facial dynamics due to camera motion,
change of point of view, as well as head’s movements or pose variations. Such situations require
AFR engines perform in real time [84], apply image/frames pre-processing such as face
alignment [101], cope with intra-class variations/inter-class similarities [102] and are able to
process single/multiple camera views [41] or synthesize a 3D face model from a single camera
[103], leading to the wider study of the computational face.
Author details

Joanna Isabelle Olszewska
Address all correspondence to: joanna.olszewska@ieee.org
School of Computing and Technology, University of Gloucestershire, Cheltenham, UK

References


