Digital Signal Processing 106 (2020) 102809

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Face recognition: Past, present and future (a review) ☆

Murat Taskiran [a,*], Nihan Kahraman [a], Cigdem Eroglu Erdem [b]

[a] Yildiz Technical University, 34220, Istanbul, Turkey
[b] Marmara University, 34722, Istanbul, Turkey

A R T I C L E   I N F O

Article history:
Available online 16 July 2020

Keywords:
Face recognition
Face identification
Facial dynamics
Image-based face recognition
Video-based face recognition

A B S T R A C T

Biometric systems have the goal of measuring and analyzing the unique physical or behavioral characteristics of an individual. The main feature of biometric systems is the use of bodily structures with distinctive characteristics. In the literature, there are biometric systems that use physiological features (fingerprint, iris, palm print, face, etc.) as well as systems that use behavioral characteristics (signature, walking, speech patterns, facial dynamics, etc.) Recently, facial biometrics has been one of the most preferred biometric data since it generally does not require the cooperation of the user and can be obtained without violating the personal private space. In this paper, the methods used to obtain and classify facial biometric data in the literature have been summarized. We give a taxonomy of image-based and video-based face recognition methods, outline the major historical developments, and the main processing steps. Popular data sets that have been used for face recognition by researchers are also reviewed. We also cover the recent deep-learning based methods for face recognition and point out possible directions for future research.

1. Introduction

The increase of human factors in new generation technologies gives rise to the need for biometric systems for person identification and verification systems. There are biometric systems that use static physiological features such as the fingerprint, iris, and palmprint, as well as systems that use behavioral characteristics such as the signature, gait, walking pattern, speech patterns and facial dynamics, some of which are also known as soft biometrics [68].

Face has been one of the main biometric traits, which has many application areas including security and law enforcement, health, education, marketing, finance, entertainment, and human-computer interaction. In Table 1, the main application areas and specific applications related to these areas are listed.

The human face carries information about the identity, age [80], gender [67], race and facial expressions reflecting the emotions and mental states [92,395,227,394]. The analysis of the human face and facial behavior is an interdisciplinary research area involving psychology, neuroscience, and engineering.

Face recognition, in contrast to several other biometric traits, does not necessarily require the cooperation of the person and can be performed in an unobtrusive way, making it particularly suitable for surveillance applications. Moreover, face recognition can be based on both physical (static) features and dynamic features of the face, making it suitable for behavioral biometrics.

The problem of face recognition in unconstrained environments is a challenging problem due to head pose, illumination, age, and facial expression related variations. There may also be changes in appearance due to make-up, facial hair or accessories (e.g. glasses, scarves). Another difficulty in face recognition is the similarity among individuals (e.g. relatives, twins) [158].

Perception of faces is a task performed successfully and almost effortlessly by humans but it is not a very easy task for computers. The human visual system accommodates complex neural paths for processing of static and dynamic features of the faces to recognize faces in relation with contextual knowledge [334]. There are many studies in psychology and neuroscience that address different issues about face perception. For example, it has been shown that both holistic and feature-based representations are used although characteristic features may be dominant [35]. The hypothesis suggested by Bruce and Young is that there are several independent sub-processes working together for face perception [36]. According to this hypothesis, various properties such as age, gender, and basic facial expressions are obtained from the simple physical features of the face as a result of independent processes, and the use of these properties enables the creation of a personal face model

☆ This research was supported by The Scientific and Technological Research Council of Turkey (TUBITAK) under project 116E088.
* Corresponding author.
E-mail addresses: mrttskrn@yildiz.edu.tr, mrttskrn1071@gmail.com (M. Taskiran), nicoskun@yildiz.edu.tr (N. Kahraman), cigdem.erdem@marmara.edu.tr (C.E. Erdem).

https://doi.org/10.1016/j.dsp.2020.102809
1051-2004/

Table 1
Application areas of face recognition methods.

Application areas	Specific applications
Security, Health	Information security, user authentication, Login to electronic devices Human robot interaction, access to buildings Recognizing faces from surveillance videos Health related applications: smart homes, smart cars Robotic assistants
Law Enforcement	Border monitoring, illegal event analysis, tracking suspects Passports, national ID cards, driver's licences Immigration
Entertainment, Education, Marketing	Gaming, virtual reality Photo management Video analytics, video retrieval Online learning, student follow-up, user engagement Advertising campaigns, social network moderation

structure. By using this personal face model structure, face perception even under different conditions is provided by the brain. There are also studies, which try to understand which features (eyes, mouth, nose etc.) are more important in recognition of faces [304]. In [140], it was shown that lighting from the top was important for recognition of faces. There are also studies, which demonstrate that familiar faces are recognized easier if they are shown in motion even under challenging conditions such as negation, inversion or thresholding [182,35].

1.1. Brief history and previous surveys

The history of face recognition goes back to the 1950s and 1960s, but research on automatic face recognition is considered to be initiated in the 1970s [409]. In the early works, features based on distances between important regions of the face were used [164]. Research studies on face recognition flourished since the beginning of the 1990s following the developments in hardware and the increasing importance in security-related applications.

The progress of image-based face recognition techniques since the beginning of 1990s has been roughly divided into four major conceptual development phases by Wang and Deng [346], which is not a full taxonomy but reflects the historical development of the major methods: i) *Holistic* or *appearance-based* approaches use the face region as a whole and use linear or non-linear methods to map the face to a lower dimensional subspace [27,363]. One of the first successful methods was developed by Turk and Pentland [333,332] and is known as Eigenfaces. There have been other approaches that use linear subspaces [77], manifold learning [378,139] and sparse representations [73,75]. ii) *Local-feature based* face recognition methods became popular after 2000s and they use hand-crafted features to describe the face such as Gabor features [203,373], and local binary patterns (LBP) and variants [237,6,76]. iii) Methods which use *learning-based local descriptors* [41,190] emerged after the 2010s and they learn the discriminant image filters using shallow techniques. iv) *Deep Learning based methods* gained popularity after the great success of AlexNet in the ImageNet competition in 2012 [184], and brought a new perspective to face recognition problem. An unprecedented stability has been achieved for face recognition so that their performance is similar to humans on large-scale datasets collected under unconstrained settings [320,262].

A full taxonomy of image and video-based face recognition methods in the literature is given in the following sections of the paper.

There are also a number of survey papers summarizing the work done on face recognition. The first survey papers were published in 1990s [293,49]. Later, other survey papers have been published [409,3,157,135,223], some of which focus on a specific aspect or method including,

- Pose [404,83] or illumination [418] invariant face recognition methods,
- Dynamic face recognition from video [326,342,234,124,26],
- Multimodal face recognition using 3D and infrared modalities [297,33,414],
- Presentation attack detection (face anti-spoofing) methods [276,298],
- Sparsity-based face recognition methods [372],
- Deep learning based face recognition methods [21,231,279, 346,119].

We can see that recent survey papers mainly focus on image-based deep learning methods. Although a video is a very rich source of facial texture and dynamics, and it has become easier to record and share videos, there is no recent survey paper focusing on video-based face recognition.

1.2. Contributions and outline

The major contributions of this review can be summarized as follows:

- We give an up-to-date, comprehensive and compact overview of the vast amount of work on **image and video based face recognition** in the literature including the image and video databases and evaluation methods. Approximately 300 papers, which were published between 1990s and the beginning of 2020 have been reviewed. Our goal is to inform interested new researchers about the main developments in the past and to point out relevant references for further details.
- We provide a taxonomy of image and video-based methods, which also contains recent methods such as sparsity and deep learning based methods. The purpose of creating a taxonomy is to provide an overview of the methods in the literature for face recognition.
- We give an up-to-date review of the image and video-based data sets used for face recognition. We not only tabulate these data sets, but also give a timeline to show how the collected data sets evolved in time in terms of the number of subjects and the number of samples per subject.
- We review the recent deep-learning based methods, which have shown remarkable results on large scale and unconstrained challenging data sets. In this way, readers have been provided with detailed information about deep-learning based methods that have brought a new perspective to face recognition since the beginning of 2010s.
- We provide information on both image and video-based methods, with an emphasis on the video-based methods. We believe video-based face recognition has not yet reached its full potential in terms of utilizing facial dynamics information.

The organization of the paper is as follows. In Section 2, we give an overview of the main concepts related to face recognition, including taxonomy, main steps, databases, evaluation metrics, and face spoofing. In Section 3 and Section 4 we summarize the methods on image and video-based methods, respectively. Finally, in

Section 5, we provide main conclusions and directions for future research.

2. Overview of face recognition

Face recognition can be approached as an *identification* problem or a *verification* problem. Face identification is also referred to as the 1:N matching problem. The unknown face is compared with all the faces in the database of known identities and a decision is made as a result of all the comparisons. If the person is known to be in the database, the task is called as *closed-set*, otherwise, it is called as *open-set*. Face verification is known as the 1:1 matching problem. The identity of the query face is either confirmed or rejected by comparing it with the face data of the claimed identity in the database.

Below, we provide an overview of the face recognition systems in the literature focusing on the general taxonomy, main steps, image and video databases, and evaluation metrics used for face recognition.

2.1. Taxonomy of face recognition

Face recognition systems in the literature can be divided into two main groups as *image-based* and *video-based* methods. Image-based systems try to recognize a person by using the physical appearance. On the other hand, video-based systems use physical appearance as well as changes in appearance over time or dynamics of the face. The general taxonomy of the literature on face recognition is shown in Fig. 1.

Image-based face recognition (FR) methods can be classified into three main groups: i) appearance-based (or holistic) methods, ii) model-based methods and iii) texture (local appearance) based methods [26,158].

Video-based face recognition methods can be classified into two main groups: i) Set-based methods and ii) Sequence-based methods. Set-based methods treat the frames of a video sequence as a collection of images, without paying attention to the temporal order of the frames. On the other hand, sequence-based methods use the frames by keeping their temporal order. Hence, the dynamics of the face over time also plays a role in the recognition of the person.

It is very difficult to give a clear-cut taxonomy of all the work on face recognition in the literature. Hence, the proposed taxonomy in Fig. 1 is a coarse grouping of the methods in the literature and the algorithms in some groups may have overlapping properties.

2.2. Main steps of face recognition

Face recognition systems traditionally consist of six main stages (see Fig. 2):

i) Input image or video of the face is acquired.
ii) Face anti-spoofing module ensures the security of the system by employing presentation or adversarial attack detection (via liveness tests etc.).
iii) Face and/or facial landmarks are detected in the image or each video frame.
iv) Pre-processing is performed on the image or video, which may consist of alignment, video frame selection, noise reduction, contrast enhancement or similar operations.
v) Facial feature extraction from the image or video. Image-based methods use holistic, model-based or texture-based feature extraction approaches, whereas video-based methods use set-based or sequence-based approaches.
vi) Face identification or verification is performed.

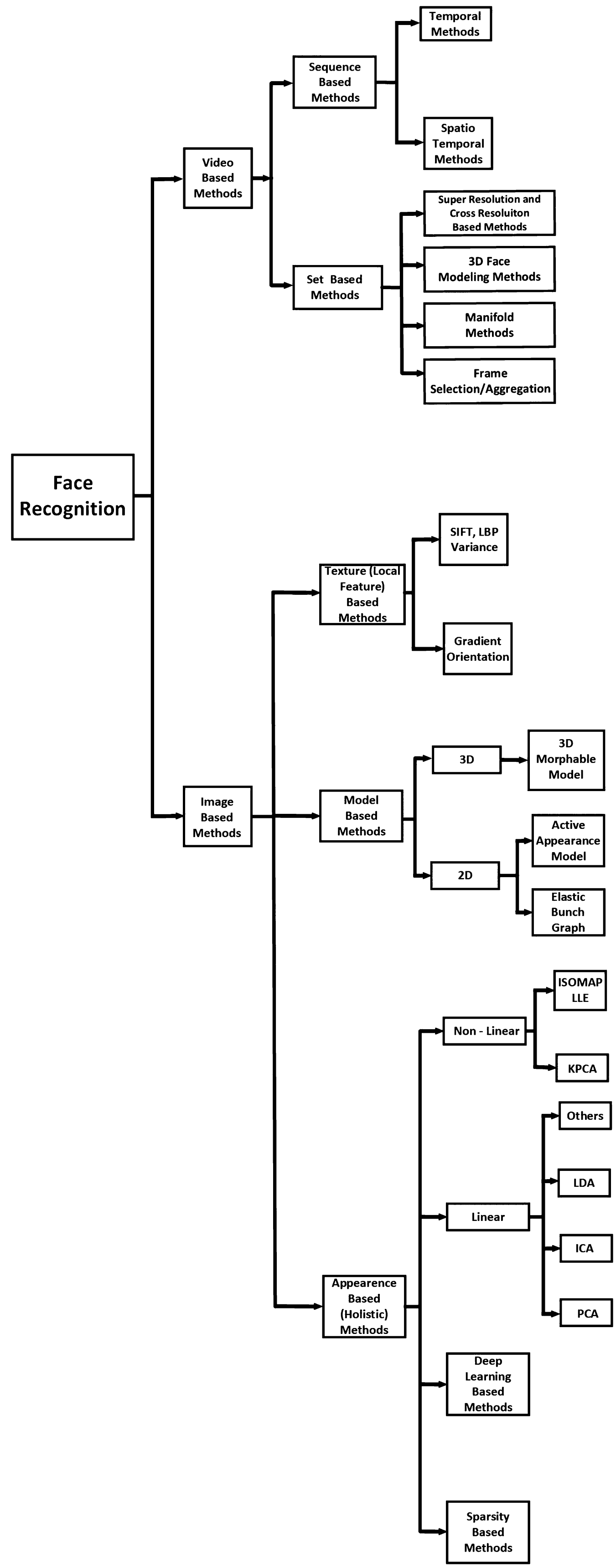

Fig. 1. The taxonomy of image-based and video-based face recognition systems in the literature.

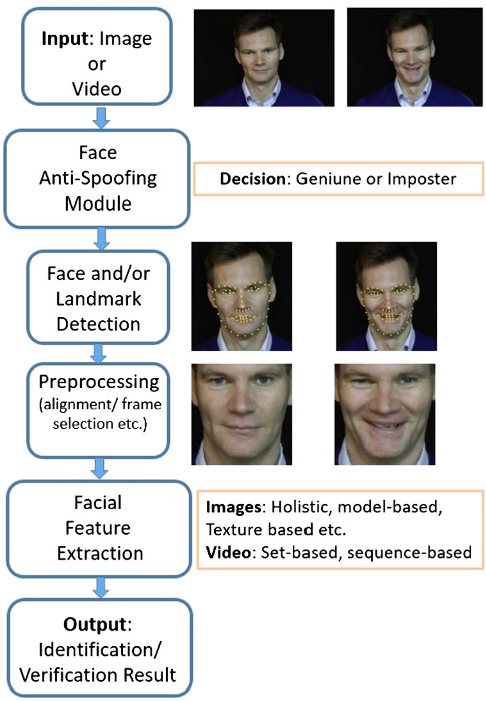

Fig. 2. The main steps of face recognition systems. Images are taken from the UvA-NEMO database [81,82]. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Below, we give a brief review of the face detection and facial landmarking methods in the literature. Accurate and effective face detection and facial landmarking algorithms increase the accuracy of face recognition systems.

2.2.1. Face detection

Face detection is estimating the bounding-box of the face in a given image or frames of a video. If there are multiple faces in the images, all of them are detected. The face detection should be robust to pose, illumination and scale differences and should eliminate the background as much as possible [279].

The Viola-Jones face detector [338] is a widely used face detector, which works well for frontal faces. It is based on Haar-like features and works in real-time. Other approaches have also been proposed, which use the color information as well [56,322,93].

Recently, deep learning based face detectors have provided successful results [279,196,146]. In a recent method, faster R-CNN, which uses the region proposals approach, has been used and was initially proposed for object detection [162,285]. There are also other deep learning based face detection methods, which use the sliding-window idea [97,382,193]. The single shot detector (SSD), which was first proposed for object detection [204], has also been successfully used for face detection [383,401].

2.2.2. Facial landmarking

After the face is detected, facial landmarks on the face (corners of the eyes, eyebrows, and the mouth, the tip of the nose, etc.) can be estimated to be used for face alignment. Aligning the face to a canonical position has been shown to be beneficial for face recognition [22]. Examples of facial landmarks are shown with yellow points in Fig. 2, which have been estimated using the ensemble of regression trees approach [169].

In the beginning of 2010s, different methods were proposed in order to perform face alignment procedures and it was shown in the studies that these methods showed high performance [371,397,284]. There are survey papers that summarize the studies on facial landmarking [42,283,163,295,60,348,366,31]. Wu and Ji have classified facial landmark detection methods into three categories as holistic methods, Constrained Local Model (CLM) methods, and regression-based methods [366]. Another possible categorization is to group them as generative methods and discriminative methods [163].

In order to evaluate the landmark localization performance two different metrics can be used: ground truth based localization error and task-oriented performance [283]. Due to the most recent advances in deep-learning techniques, the performance of facial landmark extraction methods have been substantially improved, even on in-the-wild datasets [375,365,31]. There are methods developed for multi-task learning, which combine face detection, and landmark localization with other tasks pose estimation and gender recognition [398,280,278]. Recently, single face tracking on mobile devices using deep learning has also been investigated [199].

2.3. Databases

Early works on face recognition were based on rather small-scale databases, recorded in laboratories under controlled conditions. One of the first image-based databases is ORL [294] contained 400 images from 10 subjects. Similarly, one of the first video-based face databases released in 1997 [30] consisted of 70 videos from 40 subjects. In recent years, face recognition databases have become large-scale with millions of images recorded under uncontrolled conditions or tens of thousands of videos.

The face databases used for face recognition can be grouped as *image-based* and *video-based* face databases. We summarize the main image-based and video-based face recognition databases in the literature in Table 2, 3, 4, and Table 5, respectively. A graphical temporal representation of image-based, video and 3D face recognition databases is also provided in Fig. 3.

2.4. Evaluation metrics

With the rapid increase in the usage of face recognition systems in our daily life, the performance of these systems has become a critical issue. In order to measure the performance of biometric systems, a number of evaluation metrics have been suggested by researchers.

Face recognition can be performed using an *identification* or *verification (authentication)* approach. The evaluation metrics and charts, which are commonly used for face verification are [88,271]:

- *False Match Rate (FMR)* (also known as *False Accept Rate (FAR)*): It is the percentage of impostor (intruder) samples, which are incorrectly recognized as the claimed identity.
- *False Non-Match Rate (FNMR)* (also known as *False Reject Rate (FRR)*): It is the percentage of genuine samples, which are incorrectly rejected.
- *Accuracy*: It is the percentage of samples, which are correctly classified.
- *Geniune Accept Rate (GAR)* (also known as *True Acceptance Rate (TAR)*): It is the percentage of genuine samples, which are correctly accepted (i.e. TAR = 1-FNMR).
- *Detection Rate*: It is percentage of intruders (not the samples) that are correctly detected.
- *Equal Error Rate (EER)*: It is the error rate at which FMR and FNMR are equal.
- *The Receiver operating characteristics (ROC) Curve*: This curve is the plot of FRR versus FAR obtained at different detection threshold values. The ROC curve can be also obtained by plotting TAR versus FAR. The Area under the ROC Curve (AUROC) is the metric that represents the system's performance, which takes the values between 0.5 (random selection) and 1 (perfect classification).

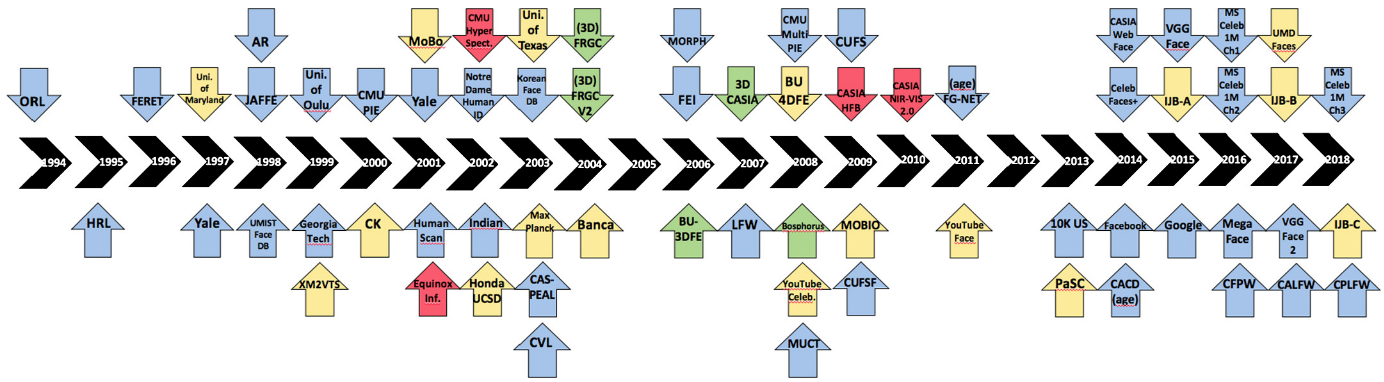

Fig. 3. The databases used for face recognition. 2D databases are shown with blue color, video databases are shown with yellow color, 3D databases are shown with green color and hyper-spectral/infra-red databases are shown with red color.

Table 2
Image-based databases for face recognition published between 1994-2002. Abbreviations should be interpreted as follows: V (various), N (No), Y (Yes), FE (number of different facial expressions), IL (illuminations), PO (head poses), OC (occlusions, e.g. hand, hair, eyeglasses, beard...), TI (recording times), AC (accessories), BG (backgrounds), ET (ethnicities). Depth means the number of images for each subject is high, and breadth means the number of subjects is high with as many images as possible for each subject.

Database	Year	Img. size, color	♯ subjects, ♯ images	Properties
ORL [294]	1994	92 × 110, N	10, 400	All pictures are frontal
HRL [130]	1995	193 × 254, N	10, ∼ 800	IL: 77-84.
Color FERET [268]	1996	256 × 384, Y	1199, 14.051	FE: 2, IL: 2, PO: 9-20, TI: 2.
Yale [107]	1997	320 × 243, N	15, 165	With and without glasses, IL, FE
JAFFE [219]	1998	256 × 256, N	10, 213	FE: 7
UMIST [114]	1998	220 × 220, N	20, 564	Various angles from left profile to right profile.
AR [226]	1998	576 × 768, Y	116, 3.288	FE: 4, IL: 4, OC: 2, TI: 2.
Georgia Tech [250]	1999	640 × 480, Y	50, ∼ 1.500	Frontal and/or tilted faces with FE, IL.
Univ. of Oulu [312]	1999	428 × 569, Y	125, 2.000	All are frontal images with 16 IL.
CMU PIE [306]	2000	640 × 486, Y	68, 41.368	PO: 13, IL: 43, FE: 3.
Human Scan [161]	2001	384 × 286, N	23, ∼ 1.500	Mainly frontal views.
Equinox Inf. [309]	2001	240 × 320, N	91, 25.000	Camera spectral range: 8-12 micrometer visible, IL: 3, FE: 3.
Yale B [108]	2001	640 × 480, N	10, 5.760	PO: 9, IL: 64.
Notre Dame HumanID [265]	2002	1600 × 1200, Y	≥ 300, ≥ 15.000	IL: 3, FE: 2, TI: 10-13.
Indian Face [159]	2002	640 × 480, N	40, 440	Images taken with a bright homogeneous background and subjects in an upright, frontal position.

Table 3
Image-based databases for face recognition published between 2003-2009. Abbreviations should be interpreted as follows: V (various), N (No), Y (Yes), FE (number of different facial expressions), IL (illuminations), PO (head poses), OC (occlusions, e.g. hand, hair, eyeglasses, beard...), TI (recording times), AC (accessories), BG (backgrounds), ET (ethnicities). Depth means the number of images for each subject is high and breadth means the number of subjects is high with as many images as possible for each subject.

Database	Year	Img. size, color	♯ subjects, ♯ images	Properties
Korean Face (KFDB) [32]	2003	640 × 480, Y	1000, 52.000	FE: 5, IL: 16, PO: 7.
CVL Face [310]	2003	640 × 480, Y	114, ∼ 800	Profile left/right, 45 degrees left/right, frontal, frontal smile, frontal smile with teeth.
CAS-PEAL [104]	2003	360 × 480, Y	2.747, 30.900	FE: 6, IL: 9-15, PO: 21, AC: 6, BG: 2-4, TI:2.
FRGC(3D) [266]	2004	1704 × 2272 or 1200 × 1600, Y	466, 50.000	Images taken under controlled and uncontrolled conditions.
FEI Face [325]	2006	640 × 480, Y	200, 2.800	Images are taken with a white homogeneous background in an upright frontal position or profile rotation up to 180 degrees.
BU-3DFE (Static) [389]	2006	V, Y	100, 2.500	FE:7, Ages: 18-70 years, ET: 6.
MORPH [286]	2006	V, Y	13.618, 55.134	Gender:%81 Male, %19 Female Ages: 16-77 years, ET: 4.
LFW [150]	2007	150 × 150, Y	5.749, 13.233	Un-posed photos, mainly frontal views.
CASIA 3D [413]	2007	V, Y	123, 4.624	FE: 6
MUCT [243]	2008	480 × 640, Y	276, ∼ 3.500	Frontal and three-quarter views. IL, manual landmarks.
Bosphorus [296]	2008	V, Y	105, 4.652	FE: 34, PO: 13, OC: 4.
CMU Muti-PIE [116]	2008	V, Y	337, ≥ 750.000	PO:15, IL:19, Some high resolution frontal images.
CUFS [353]	2009	V, Y	606, ∼ 1.200	One frontal image and one sketch for each subject.
CASIA HFB [194]	2009	V, Y	100, 800	4 VIS and 4 NIR face images per subject.
CUFSF [354]	2009	V, N	1.194, ∼ 2.400	LI, Sketch with exaggeration drawn by an artist.

Table 4
Image-based databases for face recognition published between 2010-2018. Abbreviations should be interpreted as follows: V (various), N (No), Y (Yes), FE (number of different facial expressions), IL (illuminations), PO (head poses), OC (occlusions, e.g. hand, hair, eyeglasses, beard...), TI (recording times), AC (accessories), BG (backgrounds), ET (ethnicities). Depth means the number of images for each subject is high, and breadth means the number of subjects is high with as many images as possible for each subject.

Database	Year	Img. size, color	♯ subjects, ♯ images	Properties
CASIA NIR-VIR 2.0 [195]	2010	640 × 480, Y	725, ∼ 36.000	1-22 VIS and 5-50 NIR face images per subject.
FG-NET [52]	2011	V, Y	82, 1.002	Ages:0-69 years, %58 Male, %42 Female.
10K US Adult Face [20]	2013	72 × 256, Y	2.222, 10.168	Memorability scores, computer vision and psychology attributes, and landmark point annotations.
CASIA WebFace [87]	2014	V, Y	10.575, 494.414	Celebrities.
Cross-Age Celebrity [50]	2014	V, Y	2.000, ≥ 160.000	Celebrities with ages 16-62.
CelebFaces+ [315]	2014	V, Y	10.177, 202.599	Celebrities, Private.
Facebook [320]	2014	V, Y	4.400, 4.4M	Private database.
Google [299]	2015	V, Y	$\geq 10M, \geq 500M$	Private database.
VGG Face [263]	2015	V, Y	2.622, 2.6M	Celebrities, face annotations with bounding boxes and pose.
MS-Celeb-1M(Ch1) [121]	2016	V, Y	100.000, 10M	Breadth; celebrities; knowledge base.
MS-Celeb-1M(Ch2) [121]	2016	V, Y	20.000, 1.5M	Breadth; celebrities; knowledge base.
Mega Face [249]	2016	V, Yes	672.052, 4.7M	Breadth; the whole long tail; commonalty.
CFPW [300]	2016	V, Y	500, 7.000	Frontal-profile images of celebrities.
VGG Face2 [40]	2017	V, Y	9.131, 3.31M	Depth; PO, age, ET; celebrities.
MS-Celeb-1M(Ch3) [121]	2018	V, Y	180.000, 6.8M	Breadth; celebrity.
CPLFW [410]	2018	250 × 250, Y	3.968,11.652	PO, Celebrities.

Table 5
Video-based databases for face recognition published between 1997-2018. Abbreviations: V (various), N (No), Y (Yes).
[H]

Database	Year	Frame size, color	♯ subjects, ♯ videos	Properties
Univ. of Maryland [30]	1997	560 × 240, N	40, 70	6 different facial expressions
XM2VTS Video [240]	1999	576 × 720, Y	295, 1.180	Frontal, profile, speech.
CK [165]	2000	640 × 480, N	97, 486	Each sequence begins with a neutral expression and ends with a peak expression, FE: 6.
CMU Motion of Body (MoBo) [117]	2001	V, Y	25, 100	All subjects are captured using six high resolution color cameras distributed evenly around the treadmill.
Honda/UCSD [189]	2002	640 × 480, Y	20, 1.500	Indoor environment, 15 fps, video length > 15 seconds.
Texas Univ. [256]	2003	720 × 480, N	284, ∼ 2.500	Four different categories: still facial mug shots, dynamic facial mug shots, dynamic facial speech, and dynamic facial expression.
Max Planck [34]	2003	786 × 576, Y	246	Facial Action Units recorded from six different viewpoints.
Banca Multi-Modal [19]	2004	720 × 576, Y	208, ∼ 2.500	Four languages; each subject was recorded during 12 different sessions over a period of 3 months.
YouTube Celeb. [174]	2008	V, Y	47, 1910	All videos are encoded in MPEG4 at 25 fps rate.
BU-4DFE [388]	2008	1040 × 1329, Y	101, 606	3D data at 25 fps; 6 different facial expressions.
MOBIO [236]	2009	V, Y	152, 1.824	Collected between 2008-2010 at six different sites from five countries; 12 sessions for each subject.
YouTube Face [361]	2011	V, Y	1.595, 3.425	Videos downloaded from YouTube. An average of 2.15 videos for each subject.
PaSC Video [29]	2013	V, Y	265, 2.802	Balanced with respect to distance to the camera, alternative sensors, frontal versus not-frontal views, and different locations.
IJB-A [180]	2015	V, Y	500, 2.085	Manually localized face images.
IJB-B [359]	2017	V, Y	1.845, 7.011	Manually localized face images
UMDFaces [22]	2017	V, Y	3.107, 22.075	Breadth; video.
IJB-C [235]	2018	V, Y	3.531, 11.779	This dataset has 10.000 non-face images.

The evaluation metrics and charts commonly used for face identification are: i) rank-1 accuracy and ii) Cumulative Match Characteristic (CMC) curve, which is the plot of identification rate at rank-k (correct identity is among the top-k results).

The aforementioned evaluation metrics and the charts are used for benchmarking and comparison purposes in face recognition challenges [267,239,266] and protocols [287,235].

2.5. Face anti-spoofing

Although face recognition is an easy to use a biometric trait, a major problem is its vulnerability to spoofing attacks done by photos, videos or 3D masks [94,276,301]. Spoofings attacks are most common during the recording of biometric data, feature extraction, or the decision phase. There are also other types of attacks on the network or the database where biometric data is stored [17].

Anti-spoofing in face recognition usually means liveness detection, or presentation attack detection, which can be done by sensing physiological movements such as eye blinking [362,9,308,188,62,191], facial expression changes, mouth movements [183,59,136,166], or head movements [187]. Detecting the heart rate from a face video is another method for liveness detection [196,254]. This technique is called as remote (non-contact) photoplethysmography, which utilizes the subtle color changes of the skin, which occur each time the heart beats and pumps blood to the body [70,57].

Other countermeasures can include different biometric modalities such as gait and speech. Indeed, multi-modal systems [308] are intrinsically more difficult to spoof than uni-modal systems [172]. More information about overcoming 2D photo spoofing attacks can be found in [33,260,61,253,123,99].

Deep CNN based methods have recently become popular face anti-spoofing [192,301]. In [248], the performances of different

CNN architectures for face anti-spoofing were investigated. In [212], a deep tree learning (DTL) method was proposed for zero-shot face anti-spoofing (ZSFA). ZSFA is the detection of spoof attacks that do not exist in training data, such as partial paper or transparent mask attacks. A spatio-temporal anti spoofing network (STASN) was proposed in [386], which can focus on subtle cues such as borders and moire patterns to detect spoof faces. They also presented data collection and synthesis solutions. A multi-modal face anti-spoofing challenge was also performed recently [400] using the CASIA-SURF multi-modal dataset. This study summarized the results of the most successful teams and gave important information about future research directions [201].

3. Image-based face recognition

Image-based face recognition approaches mainly involve recognition by using features from a single frame. Image-based FR methods can be grouped as appearance-based (or holistic), model-based, and texture (or local feature) based methods. Below, more details about these approaches will be given.

3.1. Appearance-based (holistic) face recognition

The expression ‘appearance-based’ was introduced by Murase and Nayar [247]. Appearance-based methods use the detected face region as a whole and try to represent it in a lower-dimensional subspace. The lower-dimensional subspace is learned from the training set using linear or non-linear methods.

3.1.1. Linear methods

One of the most well-known appearance-based face recognition methods is Eigenfaces [333], which uses the principal component analysis (PCA) to linearly project the images onto a lower dimensional space learned by using the images in the training set. The test image to be recognized is first projected onto this lower dimensional space, then the identity is determined by comparing it with the projections of the gallery images in the training set. In the study conducted by Yang et al. [377], a two-dimensional basic component analysis (2DPCA) method was proposed. The difference between this method and the principal component analysis is that the image matrix is not transformed into a one-dimensional vector prior to feature extraction. Instead, the image covariance matrix is generated directly from the original image matrix. As a result of the experimental studies conducted on three datasets, it was observed that the recognition rates obtained from 2DPCA were better than PCA.

In another linear projection method [27] Fisher’s Linear discriminant was used with the goal of generating well-separable classes in the lower dimensional space. The method is known as FisherFaces and was shown to be better than the Eigenfaces method on the Harvard and Yale Face Databases. The most common problem in the traditional linear discriminant analysis (LDA) comes up with small sample size (SSS) of datasets. In order to overcome this problem, Wang and Tang proposed A dual-space Linear Discriminant Analysis [352] approach. It was observed that the method yielded more successful results with less number of features. Another approach which uses regularized Fisher’s discriminant criterion, was proposed by Lu et al. [215]. The experimental studies using FERET database showed that the proposed method yields more successful results in face recognition than the Eigenfaces method and LDA-based variations which are proposed to solve the SSS problem.

Zhao et al. [407], proposed the method of singular value decomposition updating based on incremental principal component analysis (SVDU-IPCA) for face recognition. Due to computational cost and memory-requirement burden problems encountered in the PCA algorithm, IPCA algorithm was proposed for face recognition. Since the IPCA algorithm used in the literature does not give a guarantee about the error rate, the authors used SVDU based IPCA. Zhang et al. [396] proposed a new method of obtaining optimal projective vectors from diagonal face imagery without an image to vector transformation and called the method of Diagonal PCA (DiaPCA). In experimental studies, it was observed that DiaPCA yielded more successful results as compared to 2DPCA and PCA for face recognition.

Independent Component Analysis (ICA) performs a linear transformation so that the statistical independence between the components is guaranteed. In the study by Liu and Wechsler [202], the usefulness of the ICA algorithm for face recognition was investigated. An evaluation was made on the sensitivity of the ICA to the size of the space it was applied and also its discriminant performance was compared to other criteria such as Bayes or Fisher. Discriminant analysis showed that when ICA criteria were performed in a properly compacted and whitened area, it performed better than Eigenfaces and Fisherfaces methods for face recognition. In 2003, Liu and Wechsler introduced the independent Gabor features (IGFs) method and its applications in face recognition [203]. The innovations that this study brings to the literature are the derivation of Gabor features in the feature extraction stage and formation of IGF features-based probabilistic reasoning model (PRM) classification method. As a result, 180 features were obtained using IGF method for Face Recognition Technology (FERET) face database [268] and 88 features for ORL (Olivetti Research Laboratory) database [294]. Experimental studies on these two datasets gave face recognition accuracies of 98.5% and 100%, respectively. Deniz et al. [79] proposed a method which uses ICA as a feature extractor and SVM as a classifier for face recognition. In this study, ICA/SVM and PCA/SVM were applied on two different face databases and it was observed that the accuracies were similar. It was concluded that PCA/SVM is better for face recognition since the duration of training with ICA lasts much longer than PCA. In a study conducted by Zhi and Liu [412], a method was proposed that PCA was used to extract features from gray-scale face images, Genetic Algorithm was used to optimize the network’s weights and SVM was used for classification. As a result of the experimental studies conducted with the Cas-Peal database collected in 2003, 99% facial recognition success was achieved.

Recently, a multi-fold cross convolution method has been proposed for condensing the Gabor, PCA and ICA filters, which result in superior image descriptors. In another work [102], a linear mapping is learned using Bayesian sample steered discriminative regression (BSDR) for each class to extract the image class label features, which are then classified using a nearest neighbor classifier.

3.1.2. Non-linear methods

A group of non-linear appearance-based methods uses kernel based approaches. Kernel PCA was proposed as a non-linear extension of PCA [173]. The aim is to make a non-linear mapping of the data and then calculate the principal components of the features after this mapping. Kim et al. [173] observed that the system using Kernel PCA and SVM classifier had a smaller error rate as compared to other methods on the ORL database. A method based on kernel-based discriminant analysis was proposed by Lu et al. [214] to reduce the complexity caused by different emotions and other difficult conditions. It was observed that the proposed method is more successful than kernel principal component analysis (KPCA) and generalized discriminant analysis (GDA).

Locally Linear Discriminant Analysis (LLDA) method was proposed by Kim and Kittler [178] in order to align local structures linearly within global non-linear data structures. They reported that LLDA has low computational cost in face recognition processes compared to Kernel Linear Discriminant Analysis (KLDA) and GDA. ISOMAP [323,381] and Locally Linear Embedding (LLE) [288] are methods in which non-linear manifolds are learned from low dimensional input-space. These methods gave promising results when compared to other methods.

3.1.3. Deep learning based methods

Deep neural networks (DNN) have been successfully used in face recognition systems since 2014 due to the increase in processing power and the compilation of large databases containing (multiply) labeled samples [320,315,317,319,299]. The DeepFace method [320] based on DNN yielded a face recognition rate of about 97.35% on the Labeled Faces in the Wild (LFW) database containing thousands of face images taken under unrestricted conditions [150], which is very close to the human level (97.53%). Since then, the accuracy on the LFW dataset has reached 99.80% [346]. Although deep-learning based methods are non-linear appearance-based methods, deep learning based methods are reviewed under a new subsection in this survey study, because of the fact that deep-learning based methods have been used in the majority of facial recognition studies in the literature lately and facial recognition achievements are much higher than the other methods.

Although DNN methods give very high accuracy rates for both face identification and verification using images, their robustness under adverse conditions is being investigated [115]. They provide high recognition accuracy using a large number of high-quality images recorded under uncontrolled environments. However, if there are severe illumination variations, noise or when the images have low resolution, lower recognition accuracy has been reported [115]. Therefore, under adverse conditions, video-based approaches may provide helpful facial dynamics information.

Deep learning based face recognition methods consist of mainly three stages [346]: Face pre-processing, deep feature extraction, and face matching. We provide brief information about each of these steps below.

Face pre-processing: Deep learning based face recognition methods provide a certain robustness in recognizing facial images with different illumination, pose and facial expressions in uncontrolled environments. However, a recent study [109] showed that different illumination, exposure and facial expression have negative effects on the performance of the network and showed the need for face pre-processing to increase the performance.

In *one-to-many augmentation* pre-processing approach, the aim is to make the deep CNN pose-invariant by generating images in different poses from a single image during training. This is done since it is expensive and time-consuming to collect large numbers of images for creating a training database. The first approach for one-to-many augmentation uses data augmentation methods such as photometric transformations [184] and geometric transformations [368,376]. The second approach is based on reconstructing a 3D face model first to create 2D images in different poses [229,230]. In the third approach, CNN models are used to generate 2D images directly instead of creating 3D models from 2D images and projecting them back into 2D images of different poses [273,387]. Bao et al. [24] have synthesized a new face from the features obtained from an input image and any other input image, and have shown that this synthesis is very successful in producing realistic face images even if it is not within the training data set. Generative Adversarial Networks (GAN) are also used for this purpose [23,46,303,408].

In *many-to-one normalization* pre-processing approach, the goal is to generate the canonical view of the face image by using face images obtained from different angles in uncontrolled environments. Stacked Autoencoders (SAE) [376], CNN [417] and GAN [328] structures are have been used to obtain a frontal face image using patches from multiple images with different angles. In a recent study [405], a face frontalization method based on appearance-flow-based CNN is proposed. In [13], an adaptive pose alignment method is proposed, which adaptively learns alignment templates using facial poses, and then aligns the test and training images using these templates.

Recently, a method for illumination robust pre-processing method has been proposed [402], which removes soft and hard shadows and retains identity-related information.

Deep feature extraction: The most important decisions for designing deep CNNs for feature extraction are the choice of network architecture and loss function.

The architecture of deep CNNs can be grouped as *backbone networks* and *multiple networks*. After their high performance in ImageNet [290] competitions, deep learning networks such as AlexNet [184], VGGNet [307], GoogleNet [318], ResNet [137] and SENet [145], which are known as typical CNN architectures, have attracted the attention of researchers. These networks and their variations have also been used for face recognition. In addition to the mainstream deep neural networks, networks with multiple structures have been proposed for multi-task learning including face recognition [280,131].

The choice of the loss function is also important for training the deep CNN for face recognition. It was observed that for face recognition, the softmax loss function is not sufficient for separating the features since within-class variations are larger than between-class variations. Therefore, to make the features more discriminative, other loss functions have been proposed such as Euclidean-distance-based loss [370,356], triplet loss [299], angular/cosine-margin-based loss [206], and variations of soft-max loss [205,207].

Feature extraction is especially challenging for the one-shot face recognition problem. One-shot (or low-shot) face recognition refers to the case when there is only a single (or a few) images of some subject in the gallery [156,120,367,345,85], which is still an open research problem, since it is difficult to represent the variance of data with a few samples. There are various approaches to tackle the one-shot face recognition problem. In [156], intermediate deep attribute representations were used, which were obtained by fine-tuning a DCNN for specific attributes such as gender and face shape, which were shown to perform better than purely face-based feature representations. In [120], a regularization function is used with the cross-entropy loss function, and a new underrepresented-classes promotional loss is introduced. The best results were achieved in the MS-Celeb-1M Low-shot Face Recognition Challenge at ICCV 2017. Another method that uses a new regularization term is [345]. A different approach for one-shot face recognition was proposed in [85], which was based on using a generative adversarial network (GAN) to synthesize meaningful data for one-shot classes, which will enable the classifier to learn the one-shot classes better. The new samples were generated by adapting utilizing the data variances from other classes with more samples.

Face matching: After the selected architecture is trained using the determined loss function and the training data, deep feature extraction can be performed for test data, and face identification and verification operations can be performed. Face matching can be performed using cosine distance or L_2 distance. There are also methods, such as metric learning and sparse representation-based classifier (SRC) [346].

Main results: One of the first studies [320] showed that deep-learning networks, which are successful for object recognition would be very successful in face recognition. In this work known as **DeepFace**, Alexnet was used as the network architecture with the softmax loss function. The test performance on the Facebook dataset was 97.35%, which was close to the human performance. Another study that was done in 2014 [315] was named as **DeepID2**, which used AlexNet as network architecture and contrastive loss. The accuracy on the CelebFaces+ dataset was 99.15%. One year later **DeepID3** was proposed [316], which used VGG-Net10 instead of AlexNet as the network architecture with a test performance of 99.53%.

Schroff et al. [299] introduced **FaceNet**, which used GoogleNet-24 as the network architecture with the triplet loss function and Google database for training. Google database was the largest dataset used for face recognition until then. The triplet loss function involved an anchor image, a positive example of the same class and a negative example of a different class. The goal was to update the CNN such that the distance between the matching pair was minimized and the distance between the non-matching pair was maximized. The verification performance of faceNet on the LFW database was 99.63%.

Parkhi et al. [262] proposed the **VGGFace**, which used VGGNet-16 network architecture, the triplet loss function and the VGGFace dataset for training. The test performed on LFW database yielded 98.95% verification performance. In 2016, other methods were presented [358,206], one of which [358] used center loss as loss function and achieved higher test performance on the LFW database. In 2017, most of the proposed methods preferred the ResNet architecture and its variations [403,277,341,210,134,72]. The method proposed by Liu et al. [210], used the MS-Celeb-1M dataset [121] for training and the CoCo (congenerous cosine) loss function and achieved the highest verification performance of 99.86% on the LFW dataset.

Among the several methods proposed in 2018 [272,340,343], the one proposed by Wang et al. [343] provided the highest verification performance on the LFW dataset (99.33%) using variations of the ResNet architecture, cosface loss function and the CASIA-WebFace dataset [87] as the training data. The method proposed by Deng et al. [71] achieved 99.83% verification performance on LFW using ResNet-100 network architecture, arcface loss function and MS-Celeb-1M data set as training data [71,411].

When the test performance on the LFW database is investigated, it can be observed that there has been an extraordinary increase in the performance of face recognition systems in the last 5 years (from 97.35% to 99.86%). In the study conducted by Wu et al. [364], Intraspectrum discrimination and interspectrum correlation analysis deep network (IDICN) approach was proposed in order to increase the success in recognition of multi-spectral facial images. In the experimental studies conducted with 3 different multi-spectral face image databases, facial recognition successes ranging from 99.70% to 100% were achieved. Recently, a method for handling extreme out-of-plane pose variations have been proposed, which uses pose-specific deep CNNs [228]. Huang et al. [148] conducted extensive experiments to investigate the effect of re-sampling or cost-sensitive methods on learning success in imbalanced-class data. They created more balanced boundaries by using a deep learning network, especially in order to sustain the margins between classes. The successes achieved by combining this method with a simple kNN cluster algorithm are 99.62% and 96.5% for LFW and YTF databases, respectively. In [106], a two-stream CNN is proposed to recognize low-resolution faces. The teacher stream consists of a complex CNN for high-accuracy recognition and the student stream a simpler CNN for low-cost recognition. A new coupled mapping method using a two-branch deep learning network to perform facial recognition from low-resolution images was proposed by Zangeneh et al. [393]. The proposed structure is based on mapping low and high-resolution face images of 2-branch DCNN in a common space with non-linear transformation.

3.1.4. Sparsity based face recognition methods

After the advancements in sparse signal coding, sparsity based applications in computer vision became popular. Learning a dictionary from training data is an important part of sparse representation based approaches and has been reviewed in several papers [327,289,90,58,100,406].

Sparse representation based classifier (SRC) [363] was one of the first successful applications of sparsity for face recognition. The main approach in SRC is to use an overcomplete dictionary with elements chosen from the training samples, and to represent the test samples as a linear combination of the training samples from the same class.

In a recent study [372] dictionary learning algorithms for face recognition has been reviewed. Dictionary learning algorithms for face recognition have been grouped into five categories. The first category is based on *shared dictionary* learning and uses variants of the well-known K-SVD algorithm [5], and they are useful when inter-class variations are not large. When there are large intra-class variations *class-specific dictionary* learning approaches have been used [380]. The third group of algorithms use a *commonality dictionary* for handling intra-class variance, and a *particularity dictionary* to handle inter-class variance [339,379]. When the sample size is small *auxiliary dictionary* learning algorithms have been proposed [74,355]. When the training and test samples come from different domains *domain adaptive dictionary* learning algorithms have been used [415,274].

Recently, a method has been proposed to improve the performance of SRC classifier [384] for low-quality images. A new dictionary is learned based on extracting the low-rank components on the training data. Another recent work [144] addresses the problem of single sample face recognition under varying illumination. After handling the illumination variations using QRCP method, an SRC based classifier is employed. In [11], the performance of sparsity-based and CNN algorithms have been compared. The disadvantage of CNN based approaches is that they require large databases for training. In [138], convolutional neural networks and SRC classifier are combined to overcome the difficulties of partial face recognition. Similarly, in [43], dictionary based sparse representation is constructed using deep features obtained from a CNN, with the goal of making the CNN features more robust to occlusion. In [101], a low-rank matrix recovery approach is proposed, for face recognition with occlusion.

3.2. Model-based face recognition

In this section we summarize the model-based face recognition methods as 2D methods and 3D methods.

3.2.1. 2D methods

The most well-known 2D model-based approach is the Active Appearance Model [64]. In AAM, the statistical appearance model is matched by controlling a set of model parameters for the shape and gray level variation modes learned from a training set. AAM uses an efficient iterative matching algorithm by learning the relationship between perturbations in the model parameters and the induced image errors. In 2009, a face recognition system was proposed with constrained AAM [357]. In this work, 58 predetermined feature points were manually marked and 40 pictures were taken for the training process. Constrained AAM/NNC (Nearest Neighbor Classifier) achieved 92.5% face recognition performance.

In order to model previously unrecognized face images more accurately, Multi-Model AAM has been proposed that distinguishes faces from similar shapes before conventional AAM and allows the generation of several AAM models [171,220]. It has been observed that, this new model significantly improves the performance in images that are problematic in both shape and texture. Elastic bunch graph matching (EBGM) is another important model-based method for face recognition [186,360]. The aim is to recognize the identity in a large data set even a with single image. In this method, the fiducial points of the face are described using wavelets and converted into an image graph in order to reduce the variations in the image. In order to improve the performance of the EBGM algorithm, "retinex and color constancy" preprocessing algorithm was proposed to solve the illumination problem [170]. In order to increase the performance of the EBGM algorithm, features based on Gabor wavelets, which are called Gabor Jets [373] were used with EBGM, and beta filters were associated with the EBGM [89].

3.2.2. 3D methods

People can recognize the faces even when all the details are no longer resolved. The remaining information in the sensation of face is basically geometrical and represents what is remembered in a coarse resolution [37]. In [335,336], landmarks measures and geometrical features including curvature and shape are analyzed for 3D human face description, which are also used for face recognition. Three dimensional face registration and recognition can be performed using 3D facial shape indexes based on facial curvature characteristics and other features such as surface normals [311,112]. In [111] new features based on discrete Fourier transform, discrete cosine transform, nonnegative matrix factorization, and principal curvature directions to represent shape are investigated.

In the literature, 3D morphable models were widely used for expression invariant face recognition systems [12,10]. In [18], multiple images of a face were obtained from a single image using a 3D morphable model in order to be classified using the Fisherface method. Experimental results using the ORL face database [294] and UMIST face database [114] showed that the proposed method was more successful than the conventional Eigenface method. In FRVT 2006 (Face Recognition Vendor Test), the performance of identification systems using front-facing images, 3D information and iris biometric information were compared [269]. Dual camera systems were also used together with active appearance models (AAM) to reconstruct 3D models of the face [47], which show that 3D models are robust to facial expressions and photo spoofing attacks.

3.3. Texture (local feature) based face recognition

Texture based face recognition methods in the literature used local feature descriptors. Below we summarize several widely used approaches for describing local features and their applications for face recognition.

3.3.1. Gradient orientation based methods

Histogram of Oriented Gradients (HOG) was used for in face recognition as well as other recognition and detection tasks in computer vision [66]. In a study conducted by Albiol et al. [1], EBG was used to find the facial landmarks. Then, the HOG descriptor was calculated for each facial landmark in the graph and the nearest neighbor algorithm was used as the classifier. In another work, HOG descriptors were calculated from regular grids and these were used for face recognition [78]. An extension of the HOG algorithm was presented for use in face recognition problems in [86]. Each sub-region of the face image was given a specific weight and a method called Co-occurrence of Oriented Gradients (CoHOG) was tested using the Yale Face database [107] and the ORL Face database. It was concluded that CoHOG algorithm is more successful than HOG algorithm and the use of gradient magnitude increases the CoHOG face recognition rate. A new open-set face recognition system has been proposed by Al-Obaydy and Suandi [8] to perform facial recognition under different lighting and facial expressions. In the proposed method, while histograms of oriented gradient and Gabor wavelets are used for feature extraction, fuzzy ARTMAP classifier is used for recognition. In experimental studies, the correct classification rate for AR database was achieved as 94.22%.

3.3.2. SIFT, local binary patterns and variants

Another method for face recognition using the orientation descriptor is the Scale-invariant Feature Transform (SIFT) [213]. Although the method is generally resistant to scale and rotation changes, the computation time increases proportionally with the number of feature points. Local binary patterns (LBP) was initially proposed for texture analysis [255], and has been adopted for many applications in computer vision [149].

The method proposed by Shen and Chiu [302] used the orientation information of SIFT and the texture information of the local binary patterns (LBP) together in order to increase the recognition performance and reduce the computation time. The experimental results on the FERET database showed that the proposed LBP orientation descriptors decreased computation time by 30% compared to the original SIFT descriptors. There are also other studies in the literature that use orientation information and texture information together [344,221,330].

A study by Liao and Chung [197] proposed a new descriptor named Local Gradient Orientation Binary Pattern (LGOBP) and a new saliency measure function based on Generalized Survival Exponential Entropy (GSEE) to identify the most prominent regions in face images. In the experimental studies, it was observed that GSEE + LGOBP method gave better results than Bee Baseline and LBP method.

Meena and Suruliandi [237] performed a comparative test for face recognition using Local Binary Patterns and its variants. In this study, which compared LBP, Multivariate LBP (MLBP), Center Symmetric LBP (CS-LBP) and LBP Variance (LBPV) algorithms, it was observed that CS-LBP was more successful than other LBP variants as a result of the tests performed on three different face databases. There are also other papers, which improve LBP features [197,251,385,344,221,330] or apply them to infrared face images [369]. Xie et al. [369] proposed a method, which divides the infrared face image into non-overlapping local regions and applies pattern selection to the LBP features obtained from these regions. In [251], the steps of face detection, face localization and face recognition were performed using a single LBP transformation. This work also presents an innovative approach using LBP transformation for eye-pupil detection.

In a study by Wang et al. [344], a fusion method was proposed, which used Local Difference Binary (LDB) descriptors [385] to extract local features from a face image, and HOG descriptors to extract edge features. In the experiments using the ORL and Yale databases [107], it was observed that the proposed method was more successful than the LBP/HOG fusion method and the computation time was shorter. In a study conducted by Mady and Hilles [221], a method which used Viola-Jones algorithm for face detection, HOG, and LBP algorithms for feature extraction and Random Forest as a classifier was proposed. In the experimental studies on the mediu staff database [222], the average test accuracy of the proposed method was 97.6%. Compressive binary patterns (CBP) have been proposed, which improves LBP using random-field eigenfilters [76].

Recently, The Triangular Coil Pattern of Local Radius of Gyration Face (TCPLRGF) method, which is a variation of the Local Radius of Gyration Face (LRGF), has been proposed by Kar and Neogi [167]. In experimental studies using AR, CMU-PIE and Extended Yale B Database, 100%, 98.27% and 96.35% facial recognition success were observed, respectively.

Table 6
A summary and comparison of main **set-based video face recognition** methods in literature published between 2000-2007.

Author and year	Used methods for facial features	Classification algorithm	Database and face recognition rate	Main results
de Campos et al. (2000) [39]	Gabor Wavelet Network, Karhunen-Love	k-Nearest Neighbor voting scheme	174 images from 29 people, 97.7% Recognition rate	Proposed method is better than traditional PCA for face recognition.
Liu et al. (2003) [211]	Facial asymmetry information and conventional eigenface and fisherface	Linear Discriminant Analysis	110 subjects from FERET Database (classification error rate reduced by 38%), 55 subjects from Cohn-Kanade Database (classification error rate reduced by 45%)	The information obtained from face asymmetry can be used for face recognition.
Gunturk et al. (2003) [118]	Super-resolution reconstruction and Eigenface	Euclidean Distance	68 video sequences from CMU database, 79% recognition rate	Super resolution reconstruction provides higher face recognition performance compared to low resolution images.
Park and Jain (2007) [261]	3D modeling	Match score	CMU's Face In Action (FIA) video database with 221 subjects, 70% Recognition rate	The 3D model has increased the match performance by 40%.
Liu and Chen (2007) [208]	3D ellipsoid images (Face mosaic model)	Distance Measurement	29 subject from FIA database [110], 4.14% Error rate	Mosaic model works better than PCA method for face recognition.
Stallkamp et al. (2007) [313]	Distance to-model (DTM), distance-to-second-closest (DT2ND) and their combination.	GMM and kNN	2,292 video sequences of 41 subjects recorded during 6 months (own database), EER 18% with GMM-I, EER 21% with kNN	the combination of DTM and DT2ND methods gave the most successful result for face recognition.
Arandjelovic and Cipolla (2007) [15]	Extended version of the Genetic Shape Illumination Manifold (gSIM)	Matching with artefact model	Cambridge Face Database (CamFace) and the Toshiba Face Database (ToshFace), 1300 video sequence from 160 individuals, average error rate is 3.4%	In experimental studies, it was stated that the error rate in face recognition decreased by 50%.

4. Video-based face recognition

Humans use both rigid facial features and dynamic facial features to recognize other people around them [125]. The results of psychological and neurological studies can be summarized as follows [182,257,181,124]:

- The rigid features of the face give more reliable results than the dynamic features.
- Dynamic features contribute more to the success rate of the system under stringent conditions (such as low lighting, low resolution, recognition from a distance).
- Facial dynamics are less sensitive to illumination changes and other appearance changes (beard, glasses, makeup, etc.).
- Learning of facial dynamic features is slower than rigid features.
- Facial dynamics is helpful to recognize the face since it helps to capture the 3D features.
- The facial dynamics make it easy for people to recognize the faces they are familiar with. For unfamiliar faces, the face video is perceived only as a sequence of rigid multiple images.
- Face dynamics is useful for gender estimation.
- Dynamics of emotional expressions are independent of age, and they are consistent over the years.

In some studies, facial dynamics refer to both non-rigid movements on the face as well as rigid movements of the head. This review focuses on facial expressions and facial dynamics resulting from speech movements and different emotions. Video-based face recognition systems in the literature can be broadly grouped as: i) Set-based methods, and ii) Sequence-based methods. Below, studies in the literature under these two headings are summarized.

We also tabulate main set-based methods in Table 6, Table 7, Table 8 and Table 9. Sequence-based methods are tabulated in Table 10, Table 11, Table 12, Table 13 and Table 14 which give a summary of facial features used, classification method, database and recognition rate, and main results.

4.1. Set-based methods

In the image set-based approach for face recognition, frames of a video are treated as a set of image samples and the temporal order is not considered. Set-based approaches can be classified as methods that use fusion before matching and after matching. *Fusion before matching* involves combining features obtained from each face image before the recognition process. *Fusion after matching* technique combines the recognition results obtained from each image. This combination can be done using score, rank or decision level fusion [26]. de Campos et al. [39] used the Gabor Wavelet Network [185] to locate the face and segment the face into mouth, eyes and nose regions. Then, feature extraction was carried out using Karhunen-Loeve transform and the features obtained from the regions for each frame were classified by kNN. Finally, the voting scheme classifier [2] was used to obtain a final score from the scores which are obtained from every frame. In the proposed method, the recognition rate was 97.7%, which was 3% higher than the traditional PCA method.

In a survey paper [26], image set-based face recognition methods have been grouped under four major groups: i) Super-resolution based methods [118,15,91], ii) 3D Modeling based methods [261,208,324,347], iii) Manifold modeling and iv) Frame selection based methods [168,313].

4.1.1. Super-resolution and cross-resolution methods

In a *super-resolution based* method [118], the low-resolution images obtained from surveillance cameras are used to obtain the information contained in the high-resolution image. Instead of increasing the resolution and extracting the feature from a high-resolution image, the super-resolution process was applied to the feature vectors obtained from many low-resolution pictures and only facial information was obtained. Thus, a significant reduction in computational complexity was observed. Arandjelovic and Cipolla [15] proposed an extended version of the Genetic Shape-Illumination Manifold (gSIM) method [14], in order to match the high-resolution images in the gallery with the low resolution

Table 7
A summary and comparison of main **set-based video face recognition** methods in literature published between 2008-2013.

Author and year	Used methods for facial features	Classification algorithm	Database and face recognition rate	Main results
Thomas et al. (2008) [324]	Background substraction, gestalt clusters, temporal continuity	Weighted match scores	Own database with 12,981 frames from 57 subjects	The combination of the proposed methods increased the face recognition performance.
Mian (2008) [241]	SIFT Features	Similarity Measure	Honda/UCSD Database (Acc: 99.5%)	Frame selection is a method that can provide high performance for face recognition.
Kim et al. (2010) [179]	CCA (Principal Angles)	Nearest Neighbor	700 Face image sets	It has been observed that online learning provides lower calculation costs and locally orthogonal method improves performance.
Chen et al. (2011) [55]	Multi-Region Histogram (MRH) and LBP	Feature averaging (Avg-Feature), Mutual Subspace Method (MSM), Manifold to Manifold Distance (MMS) and Affine Hull Method (AHM).	LFW (recognition rate 92.59%) and MOBIO(Half total error rate 23.13%)	Avg-Feature method yields more successful and faster results than the remaining methods when it is necessary to identify people with few images.
Hu et al. (2012) [147]	Sparse Approximated Nearest Point (SANP)	RBFNN	Honda/UCSD DB(Acc:94.02%), CMU MoBo DB(Acc:97.91%), YouTube Celeb. DB (Acc:65.46%)	In the experiments conducted with three datasets gave the best results compared to other studies in that period.
Kashyap et al. (2012) [168]	Frequencies of action units (AU) or combinations of emotional movements on the face. AUs of the speech on the lip were not taken into account.	The sum of absolute differences between histograms.	Own database, interview videos of 20 people, Recognition rate: 50-55%	Face movements carry biometric data.
Huang et al. (2013) [155]	Coupling Alignments with Recognition (CAR)	Min, Voting, C-Voting	YouTube-S2V (recognition rates 24.57%, 30.17%, 36.21%) and COX-S2V (recognition rates 52.57%, 54.24%, 55%)	CAR method outperforms the state-of-the-art methods.

Table 8
A summary and comparison of main **set-based video face recognition** methods in literature published between 2014-2017.

Author and year	Used methods for facial features	Classification algorithm	Database and face recognition rate	Main results
Huang et al. (2015) [154]	Projection Metric Learning (PML)	Grassmannian Graph-embedding Discriminant Analysis (GGDA) [133]	YTF DB (Acc:70.04%) and PaSC DB (Acc:43.63%)	The proposed method has led to a reduction in computational costs.
Huang and Chen (2015) [151]	Local Vector Pattern (LVP) with the weighting mechanism	Feature-point Bilateral Recognition (BR)	FERET Database 99.68% recognition performance for Fb database, 95.23% recognition performance for Dup1 database	WLVP method was more successful than LVP method.
Miaoli (2015) [242]	RGB Histogram (FH), DCTmod2 features, DCTb [270]	Bayes, FLD, MLP, Mean, SVM, PSVM	XM2VTS databases. PSVM Half Total Error Rate 1.083% (for Lp1 with different feature)	PSVM has reduced the effects of low resolution images.
Cevikalp and Serhan Yavuz (2017) [44]	Extended Polyhedral Conic Functions (EPCF)	Extended Polyhedral Conic Classifier (EPCC)	COX DB (Acc:64.00%), YouTube Celeb. DB (Acc:72.1%)	The test performance increased by 18%.
ElSayed et al. (2017) [91]	SRCNN and LBP variations	Distance Measurement (chi-square metric)	Labeled Faces in the Wild Database (LFW) (average rank-1 recognition rate of 26.51% with High-Dim LBP)	Improving the resolution of the pictures increased the performance of face recognition.
Sun et al. (2017) [314]	HOG Features, CNN Features	Deep Match Kernels (DMK)	Honda/UCSD DB(Acc:98.00%), YouTube Celeb. DB (Acc:80.3%)	The DMK achieved the highest performance compared to the methods used for image-set classification.
Lu et al. (2017) [217]	CNN Features	Simultaneous Feature and Dictionary Learning (SFDL), Deep-SFDL (D-SFDL)	Honda/UCSD DB(SFDL Acc:100%,D-SFDL Acc:100%), MoBo DB (SFDL Acc:96.7%,D-SFDL Acc:98.5%), YouTube Celeb DB (SFDL Acc:76.7%,D-SFDL Acc:79.5%), IJB-A DB (SFDL Acc:26.6%,D-SFDL Acc:28.2%)	It has been observed that D-SFDL method is more successful than SFDL method.
Rao et al. (2017) [281]	Discriminative aggregation network(DAN) Features	Fully Connected Layer	YTF DB (Acc:94.28%), PaSC DB (Acc:92.06%), YTC DB (Acc:97.32%)	Proposed method can be integrate information from video frames successfully.

probe images, which are obtained from low quality video. Another method proposed by ElSayed et al. [91] used super-resolution CNN (SRCNN), which first improves the image resolution and then extracts features by a variant of LBP. Chi-square metric was used to determine the similarity between the gallery and probe images.

Coupling Alignments with Recognition (CAR) method, focused on the problem of matching a high resolution image in the gallery with the low resolution images captured under unconstrained environments for still-to-video (S2V) face recognition [155]. In the experimental studies conducted on Youtube-S2V [361] and COX-

Table 9
A summary and comparison of main **set-based video face recognition** methods in literature published between 2017-2020.

Author and year	Used methods for facial features	Classification algorithm	Database and face recognition rate	Main results
Rao et al. (2017) [282]	CNN Features	Fully Connected Layer	YTF DB (Acc:96.52%), PaSC DB (Acc:95.67%), YTC DB (Acc:97.82%)	An attention-aware deep reinforcement learning is used to eliminate the frames, which are not useful.
Wang (2018) [347]	3D dynamic features	Distance Measurement	ORL Database (300 images from 30 subjects) (precision probability 100%)	Proposed method can be improve the accuracy of face recognition.
Wang et al. (2018) [351]	Discriminant Analysis on Riemannian Manifold of Gaussian Distributions (DARG)	Set Classification	COX DB (Acc:90.13%(max) (COX13)), YouTube Celeb. DB (Acc:77.09%), YTF DB (Acc:73.01%), PaSC DB (Acc:49.37%)	The highest performance on four different face databases has been achieved.
Ding and Tao (2018) [84]	Trunk-Branch Ensemble CNN model (TBE-CNN)	Trunk-Branch Ensemble CNN model (TBE-CNN)	PaSC Database (verification rate 96.12%), COX Face Database (identification rate 98.96%), Youtube Database (verification rate 94.96%)	The proposed CNN network structure is more successful than the recently proposed CNN networks.
Mokhayeri et al. (2019) [246]	Domain-specific face synthesis (DSFS)	SRC Classifier	COX-S2V DB (pAUC:0.916, AUPR:0.775)	The proposed method has provided in a significant reduction in computational complexity.
Mokhayeri and Granger (2020) [245]	Synthetic plus variational model	SRC Classifier	COX-S2V DB (pAUC:0.905, AUPR:0.776)	A face recognition system resistant to images obtained at different angles has been proposed.

Table 10
Comparison of main **sequence-based video face recognition** studies in literature published between 2000-2007.

Author and year	Used methods for facial features	Classification algorithm	Database and face recognition rate	Main results
Chen et al. (2001) [54]	Optical flow vectors	PCA, LDA	Own database with 9 people, video from 28 people, 85% Recognition rate	Face recognition using facial motion is less sensitive to light changes.
Cohn et al. (2002) [63]	Frequency of face action units (AU)	k-Nearest Neighbor	Own database, Videos containing natural disgust, humiliation and smile recorded from 85 people, 50% Recognition rate	Personalized facial expressions are stable over time and can be used as biometric data.
Liu and Cheng (2003) [209]	Temporal Dynamics	Hidden Markov Model	Own database, 4 video sequence from 21 subjects, 1.2% Error rate	The proposed method has been more successful than image-based methods.
Aggarwal et al. (2004) [4]	Spatio-temporal method	ARMA	Honda/UCSD dataset Recognition rate more than 79%	The ARMA algorithm has achieved 90% performance in video based face recognition.
Hadid and Pietikainen (2005) [125]	Spatio-temporal method	HMM and ARMA	MoBo, 93.4% (ARMA) Honda/UCSD, 91.2% (HMM) (Note: more head movement in these databases)	Combining facial dynamics with appearance features does not systematically give better results.
Saeed et al. (2006) [292]	Geometric features of head and mouth movements	GMM	Own database, 130 videos of 9 TV speakers, 97% Recognition rate	Head and mouth movements can be used for person recognition.
Saeed et al. (2006) [292]	Head, mouth and eye movements	GMM	Own database with 144 videos of 9 TV speakers, recognition rate 97.75% (with PCA and lip movements)	Facial expressions (especially lip movements) are useful for person recognition.

S2V [153] databases, the accuracy was approximately 70%. Recently, in [103] a different approach has been used for cross-resolution face recognition. First, discriminative features, which are robust to pose variations are learned in low-resolution and high-resolution spaces using multilayer locality-constrained structural orthogonal Procrustes regression. Then, recognition is performed using these resolution-robust features.

4.1.2. 3D modeling methods

Methods that use *3D face modeling* have been proposed to overcome the difficulties of face surveillance images/videos, which are generally low resolution, have poor contrast and non-frontal. In a study by Park and Jain [261], a 3D face model was created from multiple non-frontal frames in a video, and person recognition was performed using a commercial 2D face recognition system. In the experimental studies, it was observed that the use of 3D models increased the match ratio by 40%. Liu and Chen [208] proposed an approach for face recognition, which uses the facial appearance and facial geometry to create face mosaicing. A 3D ellipsoid model was created by using local regions from different images, and these 3D ellipsoid images were used for classification.

4.1.3. Manifold methods

Different from super-resolution and 3D face modeling methods, *manifold methods* try to model the face subspaces directly, without paying attention to the underlying physical image formation process [16,127,200]. In [177], an approach based on canonical correlations of linear subspaces was proposed for comparing image

Table 11
Comparison of main **sequence-based video face recognition** studies in literature published between 2007-2008.

Author and year	Used methods for facial features	Classification algorithm	Database and face recognition rate	Main results
Matta and Dugelay (2007) [233]	Head movements	GMM	Own database, 192 videos from 12 people, TV speakers, recognition rate 92.71% with temporal features	Combining physiological and behavioral information increases the recognition rate.
Tulyakov et al. (2007) [331]	Motion vector of the characteristic points between the peak of the emotion and the neutral expression + PCA	1-NN (Euclidean distance)	CK database videos (including 9 happiness and 13 sadness videos) - Big Brother 3 (own database), 46 sadness and 8 happiness videos, 0.4% Equal error rate (EER)	The movements of the characteristic points carry information for face recognition.
Faraj and Bigun (2007) [96]	Quantification of optical flow vectors around the lips originated by speech	GMM	XM2VTS database (Recognition rate 98%)	The changes in lip movements while the person is talking have characteristic features about the person.
Al-Jawhar et al. (2008) [7]	Wavelet subbands, Optical Flow, PCA and ICA	Similarity measurement	FERET Database from 157 people, Recognition rate 73.24% with PCA and 90.45% with ICA	The structure created with ICA was more successful than the structure created with PCA.
Ning and Sim (2008) [252]	Dense optical flow fields	kNN	Own smile video database (341 video from 10 subjects)	Laughing has characteristic information about identity.

Table 12
Comparison of main **sequence-based face recognition** studies in literature published between 2009-2011.

Author and year	Used Methods for facial features	Classification algorithm	Database and face recognition rate	Main results
Hadid and Pietikainen (2009) [126]	Extended volume LBP (EVLBP), feature selection with AdaBoost.	1-NN, Compared Methods: PCA, LDA, LBP, HMM, ARMA	MoBo database (recognition rate 97.9%), [65] (recognition rate 98.5%), Honda/UCSD (recognition rate 96%)	Combining face and motion information on the face is beneficial for face and gender recognition.
Tistarelli et al. (2009) [326]	Pixels in the face area	Pseudo Hierarchical HMM	Own database, videos from 21 subjects, (Recognition rate 100%)	Fusion modeling gives better results than static modeling.
Paleari et al. (2009) [259]	14 distances from 12 characteristic points on the face	GMM	eNTERFACE [225] (1300 videos reflecting 6 basic emotions of 44 people) (recognition rate 40% for all emotions)	Facial expressions carry biometric information.
Hsieh et al. (2009) [142]	Combination of constrained optical flow and synthesized image	Similarity measurement	BU-3DFE Database [390] (2400 images from 100 subjects), (Average recognition rate 94.44%)	The proposed method is suitable for face recognition independent of facial expressions.
Tsai et al. (2009) [329]	17 Euclidean distances from the characteristic points on the face for the image at the peak of the emotion	PCA and LDA	JAFFE Database [219] (Recognition rate 80%,) CMU-AMP Database (975 images from 13 people), JAFFE + CMU-AMP (Recognition rate 65%)	Facial expressions carry biometric data. The recognition rate increases when combined with the features of the appearance.
Hadid et al. (2011) [124]	VLBP + Adaboost	SVM	CRIM database, Recognition rate 98.1%	Face movements are useful for face recognition as well as gender estimation age estimation ethnicity estimation
Zafeiriou and Pantic (2011) [392]	Change/deformation occurring during spontaneous smile/laugh + PCA/LDA	Distance measurement	Own DB (563 spontaneous smiles/laughter episodes, 849 speech utterances, 51 posed laughs, 67 speech-laughs and 167 other human noises)	During smile/laugh, feature biometric data is revealed.

sets. In [179], a method based on maximizing the orthogonality of subspaces, which represent different classes. A continuous improvement in recognition performance is provided by using online learning.

There are also set-based face recognition methods that focus on solving the face recognition problem from low-resolution images, which are obtained from surveillance cameras. In a study by Chen et al. [55], a comparison of face recognition systems using low-resolution images taken from Close Circuit TeleVision (CCTV) systems was performed. This benchmarking study was carried out by comparing four different methods on the image-based face database (LFW) and video-based face database (MOBIO) [224]: i) Feature averaging (Avg-Feature) [160], ii) Mutual subspace method (MSM) [374], iii) Manifold to manifold distance (MMS) [349] and iv) Affine hull method (AHM) [45]. In the experimental results, it has been observed that video-based face recognition methods are more successful than image-based ones in harsh conditions. In addition, the Avg-Feature method yields more successful and faster results than the other methods when it is necessary to identify faces with few images.

In [216], multiple order statistics of image set were used as features and an adaptive weight multiple kernel learning algorithm is proposed. Wang et al. [350,351], proposed the Discriminant Analysis on Riemannian Manifold of Gaussian Distributions (DARG) method to perform facial recognition using image sets. The aim of this study is to provide better classification by determining the basic data distribution in each class. For this purpose, discriminative Gaussian components in different classes were determined in the image sets as Gaussian Mixture Model (GMM). In a study by [154], a method which learns the projection metric fea-

Table 13
Comparison of main **sequence-based face recognition** studies in literature published between 2012-2016.

Author and year	Used methods for facial features	Classification algorithm	Database and face recognition rate	Main results
Mendez-Vazquez et al. (2013) [238]	Volume Structured Ordinal Features	SVM, AdaBoost	Youtube Face Database (Recognition rate 67.76% with SVM, 79.72% with AdaBoost)	Face recognition using local discriminative information achieves more successful results.
Chen (2014) [53]	LBP variation	Dual Linear Regression Classification (DLRC)	PIE DB (Error rate 20.59%), LFW DB (Error rate 1.61%), Honda/UCSD DB (Recog. rate 92.31%), CMU Mobo DB (Recog. rate 91.60%), Youtube DB (Recog. rate 66.18%)	DLRC is an appropriate method for classification.
Franco et al. (2014) [98]	Spatio-temporal key points, Hough Transform	Temporal Key points Matching	MoBo DB (Recog. rate 95.92%), Honda/UCSD DB (Recog. rate 100%)	Spatio-temporal key points can be used for face recognition.
Gavrilescu (2016) [105]	Action units (AU) and amplitude values on the face, distances between characteristic points	Artificial neural network (for facial movements) + PCA-FR (for appearance)	Own DB (videos recorded from 64 people watching 4 emotions), Honda/UCSD DB (Recog. rate 94.5%), YTF DB (Recognition rate 93%)	The use of facial movements increases face recognition rates and reduces the success of misleading attempts.
Shreve et al. (2016) [305]	Face action units (AU) and amplitudes	Histogram similarity and DTW (dynamic time warping)	Own database (96 people recorded when interacting with the tablet), 62% (rank-1 recognition rate)	Person recognition is possible with facial movements.
Kim et al. (2016) [175]	CNN features and human attributes	3DCNN	UVA-NEMO smile [81] database (Error rate 1.7%)	When the spatio-temporal feature and human attributes were used together, the system gave more successful results.

Table 14
Comparison of main **sequence-based face recognition** studies in literature published between 2017-2018.

Author and year	Used Methods for facial features	Classification algorithm	Database and face recognition rate	Main results
Haque et al. (2017) [132]	FACS	Artificial Neural Network	PAINFUL DATA: The UNBC-McMaster Shoulder Pain Expression Archieve Database (Recog. rate 87.42%)	The frame of the person suffered were not always visually distinctive with the frame of they did not suffer.
Hajati et al. (2017) [129]	Spatio-temporal information (Dynamic texture)	Derivative Sparse Representation (DSR)	Honda/UCSD DB (Recog. rate 96.31%), CMU MoBo DB (Recog. rate 83.69%), Youtube DB (Recog. rate 90.45%) (when dynamics texture length is equal to 10 frames)	DSR was more successful than other methods for short-length videos.
Haamer et al. (2018) [122]	CNN features and geometric features	Support Vector Machine (SVM).	Own database (630 videos from 61 people) (Recog. rate 96.2%)	The experimental results showed that the transition frames outperform the peak emotion frames in face recognition.

ture was directly on Grassmann manifold instead of Hilbert space for video-based facial verification was proposed. The aim of this study was to avoid the high computational costs encountered in kernel-based methods. Cevikalp and Serhan Yavuz [44] proposed a new method for calculating the distance between gallery images and the convex-hull of query images for large-scale face recognition applications using image sets. In addition, a new polyhedral conic classifier (Extended Polyhedral Conic Classifier (EPCC)) was proposed. In another work [314], deep match kernels (DMK) were proposed for image-set classification. DMKs were recommended to overcome the challenges of high inter-class ambiguity and high intra-class variability in image sets. Lu et al. [217] proposed simultaneous feature and dictionary learning (SFDL) and deep-SFDL (D-SFDL) methods for image-set based face recognition under challenging conditions (different pose, illumination, expression, etc.). The D-SFDL method was proposed to improve the recognition performance by overcoming the difficulties caused by the non-linearity between image samples. In the experiments conducted on five different face data sets, it was observed that SFDL and D-SFDL achieved high performance for image-set based face recognition.

4.1.4. Frame-selection-based methods

In *frame selection* methods for face recognition, the goal is to select the most informative and diverse subset of the images/frames in a face video or set of images [241]. A two-stage face recognition method was proposed by Huang and Chen [151]. In the first stage, a new feature extraction approach called Local Vector Pattern (LVP) [95] was used with a weighting mechanism to calculate the distance between the probe images and the enrolled images, and M candidates are determined from the enrolled pictures. In the second stage, the final classification process is carried out by using the feature-point Bilateral Recognition (BR) approach. Stallkamp et al. [313] performed a study to create a face recognition system from videos recorded at the entrance of a laboratory for six months. Because images were obtained from an uncontrolled environment, many challenging situations were encountered. In the study, which used distance to model (DTM) and distance-to-second-closest (DT2ND) weight functions, the K-Nearest Neighbor (KNN) and Gaussian Mixture Model (GMM) approaches, it was observed that the combination of DTM and DT2ND method gave the most successful result. Thomas et al. [324] proposed a method that used background subtraction and gestalt clusters to improve the face detection performance in difficult images. Weighting match scores based on the results of previously seen frames were used. Although the data used in the study are real-world data, it has been observed that the performance has been increased by combining the methods.

4.1.5. Frame-aggregation-based methods

Recently, deep-learning based *frame aggregation* methods have been utilized for face recognition from video. In [281], an aggregation network is learned to synthesize discriminative images from video using metric learning and adversarial learning. In [375], features of video frames are first extracted using CNNs and then the features are aggregated using attention blocks. In [282], an attention-aware deep reinforcement learning is used to the eliminate frames, which are not useful based on the information from the image space and the feature space. The experimental results on Youtube Face, Point-and-Shoot Challenge and Youtube datasets show competitive performance.

Sparse representations have also been employed for image-set based face recognition. In [147], a Sparse Approximated Nearest Point (SANP) method was proposed in order to calculate the between-set distance, which used an accelerated proximal gradient method for optimization. In 2015, a study was conducted by Huang et al. [152], which compared the set-based face recognition methods using the COX face database. Also, point-to-set correlation learning (PSCL) method was proposed for video-to-still (V2S)/S2V face recognition. In a recent work [246], a domain-specific face synthesis method has been proposed for face recognition. A small representative subset of face images is selected for 3D model reconstruction, which is then used for designing a discriminative dictionary for SRC classifier. In another study conducted by Mokhayeri and Granger [245], a model called synthetic plus variational model, in which a common probe picture was created by using the variational dictionary and the gallery dictionary, using synthetic facial images augmented with different exposure angles, was proposed.

Recently, trunk-branch ensemble deep CNN (TBE-CNN) model has been proposed by Ding and Tao [84] to solve the illumination and low-resolution problems, when face recognition is performed using images obtained from surveillance cameras. In order to increase the performance of the network, artificially blurred images were given to the training and improved triplet loss function was proposed. In experimental studies performed with Point-to-shoot camera(PaSC) [29], COX Face and Youtube Face Database, 96.12% verification rate, 98.96% identification rate and 94.96% verification rate were obtained respectively.

In summary, in set-based methods, three video data sets were commonly used for comparison purposes. These data sets are: Youtube Face DB, Youtube Celeb DB and PaSC DB. Although the face identification performance on these datasets was around 80% at the end of 2015, it has been observed that identification performance has exceeded 95% with the use of deep neural networks since 2017 [281,282,84].

4.2. Sequence-based face recognition

Sequence-based methods employ the temporal information that exists in a video, and hence the order of frames is important. Sequence-based methods can be grouped as temporal methods and spatio-temporal methods. Temporal methods use the facial dynamics information separately from the texture information, whereas spatio-temporal methods model the texture and the motion information together.

4.2.1. Temporal methods

Studies about recognition of people from facial movements first appeared in the beginning of the 2000s and have attracted the attention of many researchers [326,124]. Facial movements can be captured using facial landmarks, action units (AU) or optical flow vectors.

Cohn et al. [63] was one of the first studies showing that facial dynamics can be used for person recognition. First, facial expressions during solitary viewing of films were analyzed and the rate of positive expressions in the responses of people to a video they watched were used. Then, facial expression during two-person interviews is analyzed, which showed that the frequency of occurrence of some facial action units on the face can be used for person recognition. Facial expression was measured using convergent measures, including facial EMG, automatic feature-point tracking, and manual FACS. Chen et al. [54] proposed a method for face recognition using a high dimensional vector obtained from a sequence of motion flow fields. This feature vector includes mainly temporal information of the face. It has been observed in experimental studies that this method can give successful results under difficult conditions.

Studies in the literature indicate that emotional expressions on the face are independent of age and therefore remain constant over the years and are less sensitive to light changes and other appearance changes (beard, glasses, makeup, etc.) [259,234]. Paleari et al. Paleari et al. [259] conducted experiments on the eNTERFACE [225] dataset containing 1300 videos of 44 people. Videos in this data set contain short sentences that reflect the six basic emotions (anger, fear, disgust, happiness, sadness and astonishment). Using the distances between the characteristic points on the face, 14 attributes were obtained and the averages of these attributes on the video were normalized and used for GMM-based classification. Considering the number of instances in the ENTERFACE database, the average of 1-best recognition accuracy is 16 times better than random; In the worst case, the system is performed 7 times better than random.

Matta and Dugelay used the behavioral approach for face recognition in some studies [232,292,233]. Saeed et al. [292] proposed a new person recognition system based on temporal signals from rigid head displacements and non-rigid mouth movements. Gaussian Mixture Model (GMM) approach and Bayes classifier were used as classifiers. In this study, a data set consisting of 130 videos from 9 different people was used and the identification rate was estimated as 97%. One year later, a multi-modal system [233] was proposed, in which behavioral information and physiological information were used together by Matta and Dugelay. In order to obtain the behavioral information, statistical features obtained from the displacement of the head were used. The physiological information was obtained using the probabilistic extension of the Eigenface approach. Classification was done using the Gaussian Mixture Model and the Bayes classifier. This work showed that combining physiological and behavioral information increased the identification rate. The face dynamics recognition method [347] used adaptive template matching for segmentation and three dimensional dynamic scanning method for extracting 3D dynamic features from the pupil and the eyelid.

A method which tries to combine the appearance attributes of the face with emotional facial expressions was proposed by Tsai et al. [329]. The attributes of the appearance were obtained by using PCA and the facial expressions were obtained from the 17 distances between the characteristic points on the face. Experiments on JAFFE and CMU-AMP data sets have shown that both facial appearance attributes and facial expression attributes can be used for biometric recognition. The results were also confirmed by confidence interval analysis. In another fusion work, Saeed and Dugelay [291] created a new biometric system by combining eye dynamics with other biometric dynamics. Firstly, the global and local features obtained in the system were merged and then classified by using a Bayesian Classifier. The identification rate of the system was 97.75% with the help of PCA and mouth dynamics.

Kashyap et al. [168] used facial asymmetry for face recognition. They used the frequency of facial action units' (AU) related to facial movements and facial asymmetry as soft biometric data. Experimental results obtained on a data set consisting of interviews indicate that facial movements and facial asymmetry can be used as behavioral biometric data and a combination of facial asymmetry and facial action units were more successful than using the two methods separately. In another work, the facial action units (AUs) exhibited by a person were used as distinctive features when performing a task on a tablet computer in a semi-restricted environment [305]. AUs were measured from videos of 96 different participants in a show-like quiz game that included a reward. They proposed a method that took advantage of the activation properties and the temporal dynamics of facial behavior.

The display of the velocity field caused by the relative movement between an object and the camera is called the **optical flow**. Optical flow, which is one of the methods used in motion analysis in the video, is also used for face spoofing detection, facial expression detection, and face asymmetry measurement as well as face recognition. Face spoofing is an important problem for face recognition, which is mostly done using photography. A liveness detection method, which used the lighting differences between 2D object movement and 3D object movement in optical flow, was proposed by Bao et al. [25]. Chen et al. [51] proposed a method of face asymmetry measurement based on optical flow, which was used for face recognition and face image beautification.

In the study conducted by Faraj and Bigun [96], it was revealed that lip movements during the speech in addition to voice can be used for person recognition. In this study, quantification of optical flow vectors around the lips originated by speech was used for feature extraction. It was concluded that the lip movements caused by the speech carry biometric information. In another study by Al-Jawhar et al. [7], a method using wavelets, optical flow and PCA was proposed. First, wavelet transform and PCA were applied to the reference image and the test image. Then, optical flow residue image was obtained and recognition was done by comparing the residue images with the same emotions and different emotions. Ning and Sim conducted a study to investigate whether smile dynamics conveyed information about the identity [252]. Dense optical flow fields were calculated and features were extracted. In experimental studies using videos of 10 subjects, it was observed that smiles have characteristic information about identity. Hsieh et al. [142,143] proposed a method, which combined computed intra-person optical flow with synthesized face images in a probabilistic framework in order to create more robust face recognition systems against facial expressions.

Recently, neural networks have become a part of fusion methods since they are used for both feature extraction and classification. In a study by Gavrilescu [105], a method was proposed using individual differences in facial expressions in order to strengthen face recognition systems in misleading situations. In this study, the face is analyzed in four regions (eyebrows, eyes, mouth, cheeks) in order to create individual differences of facial expressions for use with the standard PCA-based face recognition algorithm. Then, a facial expression behavior map was created by using these regions and recognition was performed by using artificial neural networks. Finally, the standard PCA-based face recognition and individual facial expression recognition processes were combined. When the combined method was applied to the Honda/UCSD and Youtube Face databases, the test performance was 94.5% and 92.9%, respectively. Another recent work extracts facial dynamics features from smile videos for face recognition [321]. The extracted features utilize statistical geometric characteristics of the face during the onset, apex, and offset phases of the emotional expression.

4.2.2. Spatio-temporal methods

Spatio-temporal methods for face recognition from videos [209,4] utilize the motion and texture information together. A method using adaptive Hidden Markov Models (HMM) [275] was proposed by [209] to learn the temporal dynamics from video sequences of each subject. The study by Tistarelli et al. [326] presents a dynamic facial model based on Hidden Markov Models to capture both facial appearance and facial dynamics. The classification of the different emotions in facial expressions was performed by an unsupervised clustering method. In the experimental studies, data of 21 subjects were used and the proposed pseudo hierarchical HMM method achieved 100% accuracy. In the study by Aggarwal et al. [4], the autoregressive moving average (ARMA) model was used, which provided face recognition from video under different poses and expressions. However, these methods use holistic information on the face and do not take into account local characteristics. In order to eliminate these disadvantages, a method was proposed, which used local information in facial videos and selects only face recognition aids in facial dynamics [128,126].

In the study by Hadid et al. [128], the Extended Volume Local Binary Patterns (EVLBP) was proposed and the most discriminating EVLBP attributes were selected with the AdaBoost learning algorithm. The face video was subdivided into local rectangular prisms to extract local characteristics. The characteristic points on the face were not utilized when extracting local attributes. In this study, three different public video face databases (Motion of Body (MoBo), CRIM, and Honda / UCSD) were used and five methods (PCA, LDA, LBP, HMMs, and ARMA) were used for comparison. The proposed method was found to be more successful than the other benchmark methods and it achieved 97.9%, 96% and 98.5% test performance on MoBo, Honda / UCSD, and CRIM data sets, respectively. One of the important results of this study was that some of the facial movements (intra-personal) are not useful for face recognition, but may be useful when the attribute is selected to model inter-personal differences. Mendez-Vazquez et al. [238] have proposed a new spatio-temporal descriptor based on structured ordinal features to capture local discriminative information, which is overlooked in most spatio-temporal methods. In the experiments on the Youtube Face Database, it was observed that the proposed method was more successful than the LBP variants.

A method of face recognition using the spatio-temporal dynamics in eyes was proposed by Vinette et al. [337]. Bubbles procedure [113] was used to examine the information in the first part of the video. In a later study [125], spatio-temporal methods (HMM and ARMA) and image-based methods were compared for different video lengths and resolutions on datasets containing facial expressions and head movements. Spatio-temporal methods were found to have worse results for shorter videos and better results at lower resolutions. These results support the conclusion that facial dynamics in the aforementioned low resolution cases contribute more to the performance of the recognition system. However, it was also concluded that the combination of facial appearance and facial dynamics do not always have a positive effect. It is stated that this result may be due to head movement rather than facial expression in the data sets used. Hadid et al. [124] investigated the spatio-temporal features obtained from the head and facial parts as an adjunct to the LBP-based methods.

In a study performed by Zafeiriou and Pantic [392], it was shown that the change/ deformation occurring during a spontaneous smile/laugh was a biometric feature. During spontaneous laughing, the moment when the change/deformation was maximum (apex) was used for feature extraction. In the experiments, 563 spontaneous laugh videos collected from 22 subjects were used. Chen [53] proposed Dual Linear Regression Classification (DLRC) method in order to solve the problem of not having spatio-temporal connection between multiple images in cases where multiple images are present in both the gallery and probe sets for each subject. In experimental studies on the Youtube Face Database, DLRC algorithm was the fastest algorithm, although the face recognition performance rate was not very different from the compared methods at that time.

Franco et al. [98] proposed a video-based face recognition system using spatio-temporal key points. In the proposed method, key points were analyzed in a temporal window and key points in fixed positions and scales were selected using Hough Transform. Then, a template was created using the spatio-temporal descriptor computation and binary representation computation. In the experimental studies with MoBo and Honda dataset, 100% recognition performance was achieved.

Recently, deep CNNs have been used for feature extraction and classification for sequence-based face recognition. Kim et al. [175] designed a 3D convolutional neural network (3DCNN) for spatio-temporal representation obtained from facial motion and appearance. In order to train the network with a small number of images, some human attributes were used. In the experimental studies, it was observed that 3DCNN, which was designed using human attributes, was more successful than 3DCNN without human attributes.

In the study conducted by Haque et al. [132], a biometric person recognition system was proposed using the features obtained from the pain expression model. The pain database, which was collected by [218] was used. The database includes face videos of participants suffering from shoulder pain and performing a series of active and passive motion tests. FACS was used for feature extraction and ANN was used for classification.

One of the problems encountered in video-based face recognition algorithms is that videos can be very short-length. In order to solve this problem, Hajati et al. [129] have proposed a new derivative sparse representation approach for face recognition in short-length videos. In the experiments conducted with four different databases, it was observed that proposed method was more successful than other methods for short-length videos.

In a study by [122], a person recognition system was proposed using transition frames of emotions for dynamic feature extraction. A fine-tuned VGG-Face CNN [262] was used and geometric features were obtained from facial landmark points. The study focuses on two different approaches. In the first approach, a Long-Short Term Memory (LSTM) network [141] was trained by using the features obtained from the CNN and the geometric features obtained from the facial landmarks. In the second approach, an LSTM is independently trained by using the features of the CNN and the geometric features obtained from the facial landmarks and then used with an SVM. As a result of the study, it was observed that the system, which used SVM was more successful than the others.

In summary, when the results of face recognition using sequence-based methods were examined, it was observed that Franco et al. [98] achieved a 100% accuracy on Honda UCSD data set and 95.92% accuracy on MoBo DB using spatio-temporal key points and Hough Transforms. Moreover, CNN-based algorithms have appeared among the sequence-based methods since 2016 and it has been observed that face recognition accuracy has increased [175,122].

5. Conclusions and future work

In this survey, the vast literature on face recognition is reviewed and the main experimental results using different databases are provided. In the first part of the survey, general information about face recognition systems and their development throughout the history is given. In the second section, a taxonomy of facial recognition methods and a summary of popular facial data sets used for the training and testing facial recognition systems are provided. Image-based and video-based face recognition methods are reviewed in detail, and tables are used to compare different methods.

The first studies on face recognition mainly utilize images taken in controlled environments. Since the performance for face recognition methods using images may be limited, researchers thought that using temporal or spatio-temporal information obtained from the videos would increase the accuracy of face recognition. As a result of this, between 2000 and 2010 facial several video data sets were collected. In 2010, the success of deep neural networks for object recognition attracted the attention of researchers and deep neural networks began to be applied to problems in many areas. The use of Deep Neural networks for face recognition can be considered a milestone for face recognition. Facial recognition systems using Deep Neural Networks have achieved over 99% accuracy even when very large face data sets collected in the wild are used. On the other hand, several recent studies after 2018 [346,115,258] showed that the performance of face recognition systems using deep neural networks decrease when face images collected under adverse conditions are used such as images with low resolution, severe illumination variations, blur, and noise, which are also referred to as semantic adversarial attacks [244]. Hence, research efforts towards making deep learning based methods more robust under adverse conditions is needed. Methods for verifying the robustness of the deep learning models against semantic perturbations are also emerging [244].

Video-based face recognition methods are more successful under challenging conditions as compared to image-based face recognition approaches since behavioral features obtained from facial dynamics also can be used as auxiliary features and have a positive effect on the recognition rate. Moreover, facial dynamics are less sensitive to illumination and other appearance changes (beard, glasses, makeup, aging etc.). However, using only the facial dynamics can not achieve sufficient performance for person recognition and not every feature obtained from the face/head dynamics has been shown to be useful for person recognition. Therefore, decomposition and utilization of the identity-related facial dynamics information is a future direction for research.

It is foreseen that studies related to improving facial recognition systems will include the following concepts in the future:

Image enhancement: The goal here is to apply super resolution algorithms or 3-D image generation algorithms to low resolution face images in order to increase the performance of face recognition systems. Since there is a large volume of data obtained from security cameras, face recognition using low-resolution images is an important and challenging research problem to investigate [223,229,103]. New algorithms for resolution-robust feature extraction methods are needed for reducing the gap between low-resolution and high-resolution images.

Loss functions: Since the performance of face recognition systems using deep neural networks decrease under adverse conditions, another research direction is to increase the performance of these systems by utilizing new loss functions [346,343,119]. Currently, there are approximately 20 different loss functions used in the literature for deep-learning based face recognition and face anti-spoofing systems. It is predicted that the number of loss functions used in face recognition systems will increase in the future.

Data set design: In order to improve the robustness of face recognition using deep neural networks, images with different illumination, pose and noise effects could be used during training [228,84]. Since it is very difficult to obtain large annotated databases, another approach could be to decrease the number of training images using active learning and similar approaches [198]. Another interesting research direction would be to apply curriculum learning which uses the training data set starting from easier images for better generalization performance [28,38]. New multi-modal video-based datasets are also needed, which contain various facial expressions of the same person to recognize the identity from facial dynamics information only [321]. Multi-modal datasets, which include RGB, depth and infrared data as well as 3D masks are also important for face anti-spoofing research [201].

Soft biometrics: Soft biometric data can be extracted from facial dynamics using the spatio-temporal information in facial videos. Although research results show that soft biometric features alone are not sufficient for face recognition, they can be used together with appearance-based methods to increase the face recognition accuracy of the system under adverse conditions [122,321]. It has been recently shown that face authentication with high accuracy is possible using facial dynamics of smile expression [176]. Investigating whether the facial dynamics of other emotional expressions (e.g. anger, sadness, surprise, disgust, fear) carry identity information is an interesting direction of research. It may also be interesting to investigate the use of facial dynamics for face anti-spoofing, which also requires collecting new datasets, as mentioned above.

Face anti-spoofing: Although face recognition systems, which utilize deep neural networks have shown to exceed human performance in various scenarios, it has been observed that deep learning networks are more easily deceived than humans. Deepfake challenge has been recently organized [69] to encourage the research on development of more robust deep networks. Therefore, making face recognition more robust to spoofing attacks is an interesting research direction. Recently, many deep neural network based methods have been proposed for face anti-spoofing [248,212,386], which have shown successful performance against various types of spoofing attacks. In the future, it is expected that zero-shot face anti-spoofing approaches will be needed since new types of spoofing attacks are being created. Improving and validating the robustness of deep neural networks against adversarial and semantic attacks is also an active research area.

Multi-modal and cross-modal face recognition: In this survey, we focused on image and video-based face recognition using the visual (RGB) modality, since it is the most widely used and cost-effective way for capturing face information. However, we would like to mention that face recognition using other modalities (3D, near infrared, thermal infrared, sketches) and multi-modal face recognition and face anti-spoofing are active and interesting research areas [414]. Heterogeneous face recognition, which tries to match face images acquired using different modalities has been attracting the attention of researchers [264]. For example, matching face sketches to photos is an important problem for forensic security. Synthesizing a face photo from a sketch and vice-versa are related interesting and challenging problems to investigate [416,399,48,391].

The ultimate goal of all these academic studies is to develop an automated face recognition system that can reproduce/surpass the human vision system. This aim can be achieved by mutual and coordinated studies between computer-vision researchers and neuroscientists.

Abbreviations

LBP: Local Binary Patterns, FR: Face Recognition, R-CNN: Region with Convolutional Neural Network, SSD: Single Shot Detector, CLM: Constrained Local Model, FMR: False MAtch Rate, FAR: False Accept Rate, FNMR: False Non-Match Rate, FRR: False Reject Rate, GAR: Geniune Accept Rate, TAR: True Acceptance Rate, EER: Equal Error Rate, ROC: Receiver Operating Characteristics, AUROC: Area under the Receiver Operating Characteristics, V: Various, N: No, Y: Yes, FE: Facial Expression, IL: Illuminations, PO: Head Poses, OC: Occlusions, TI: Recording Times, AC: Accessories, ET: Ethnicities, CMC: Cumulative Match Characteristic, PCA: Principal Component Analysis, 2DPCA: Two Dimensional Principal Component Analysis, LDA: Linear Discriminant Analysis, SVDU-IPCA: Singular Value Decomposition Updatingbased on Incremental Principal Component Analysis, DiaPCA: Diagonal Principal Component Analysis, ICA: Independent Component Analysis, IGFs: Independent Gabor features, PRM: Probabilistic Reasoning Model, ORL: Olivetti Research Laboratory, SVM: Support Vector Machine, KPCA: Kernel Principal Component Analysis, LLDA: Locally Linear Discriminant Analysis, KLDA: Kernel Linear Discriminant Analysis, LLE: Locally Linear Embedding, DNN: Deep Neural Networks, CNN: Convolutional Neural Network, 2D: 2-Dimensional, 3D: 3-Dimensional, GAN: Generative Adversarial Networks, SAE: Stacked Autoencoders, SRC: Sparse Representation-based Classifier, AAM: Active Appearance Model, NNC: Nearest Neighbor Classifier, EBGM: Elastic Bunch Graph Matching, FRVT: Face Recognition Vendor Test, HOG: Histogram of Oriented Gradients, Co-HOG: Co-occurrence of Oriented Gradients, SIFT: Scale-invariant Feature Transform, LGOBP: Local Gradient Orientation Binary Pattern, GSEE: Generalized Survival Exponential Entropy, MLBP: Multivariate Local Binary Patterns, CS-LBP: Center Symmetric Local Binary Patterns, LDB: Local Difference Binary, gSIM: Genetic Shape-Illumination Manifold, SRCNN: Super-Resolution Convolutional Neural Network, CAR: Coupling Alignments with Recognition, Avg-Feature: Feature averaging, MSM: Mutual subspace method, MMS: Manifold to manifold distance, AHM: Affine hull method, GMM: Gaussian Mixture Model, DARG: Riemannian Manifold of Gaussian Distributions, EPCC: Extended Polyhedral Conic Classifier, DMK: Deep MAtch Kernels, SFDL: Simultaneous Feature and Dictionary Learning, D-SFDL: Deep Simultaneous Feature and Dictionary Learning, LVP: Local Vector Pattern, KNN: K-Nearest Neighbor, SANP: Sparse Approximated Nearest Point, V2S: Video-to-still, S2V: Still-to-Video, PSCL: Point-to-Set Correlation Learning, TBE-CNN: Trunk-Branch Ensemble Convolutional Neural Network, PaSC: Point-to-Shoot Camera, AU: Action Units, ARMA: Auto-regressive Moving Average, EVLBP: Extended Volume Local Binary Patterns, MoBo: Motion of Body, DLRC: Dual Linear Regression Classification, ANN: Artificial Neural Network, LSTM: Long-Short Term Memory.

CRediT authorship contribution statement

Murat Taskiran: Writing - Original Draft, Writing - Review & Editing. **Nihan Kahraman:** Writing - Review & Editing, Supervisor. **Cigdem Eroglu Erdem:** Conceptualization, Writing - Review & Editing, Supervisor, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Albiol, D. Monzo, A. Martin, J. Sastre, A. Albiol, Face recognition using HOG-EBGM, Pattern Recognit. Lett. 29 (2008) 1537–1543, https://doi.org/10.1016/j.patrec.2008.03.017.

[2] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 4–37, https://doi.org/10.1109/34.824819.

[3] A.F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2d and 3d face recognition: a survey, Pattern Recognit. Lett. 28 (2007) 1885–1906.

[4] G. Aggarwal, A.K.R. Chowdhury, R. Chellappa, A system identification approach for video-based face recognition, in: International Conference on Pattern Recognition, Cambridge, UK, 2004, pp. 175–178.

[5] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (2006) 4311–4322, https://doi.org/10.1109/TSP.2006.881199.

[6] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 2037–2041.
[7] W.A. Al-Jawhar, A.M. Mansour, Z.M. Kuraz, Multi technique face recognition using PCA/ICA with wavelet and optical flow, in: International Multi-Conference on Systems, Signals and Devices, 2008.
[8] W.N.I. Al-Obaydy, S.A. Suandi, Open-set single-sample face recognition in video surveillance using fuzzy artmap, Neural Comput. Appl. (2018) 1–8.
[9] A. Ali, S. Hoque, F. Deravi, Gaze stability for liveness detection, Pattern Anal. Appl. 21 (2018) 437–449.
[10] A.A. Alomari, F. Khalid, R.W.O.K. Rahmat, M.T. Abdallah, Expression invariant face recognition using multi-stage 3d face fitting with 3d morphable face model, in: International Conference on Computer Applications and Industrial Electronics, Kuala Lumpur, Malaysia, 2010, pp. 151–154.
[11] N. Alskeini, K. Nguyen, V. Chandran, W. Boles, Face recognition: sparse representation vs. deep learning, in: International Conference on Graphics and Signal Processing, ICGSP 2018, Sydney, Australia, 2018.
[12] B. Amberg, R. Knothe, T. Vetter, Expression invariant 3d face recognition with a morphable model, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2008, pp. 1–6.
[13] Z. An, W. Deng, J. Hu, Y. Zhong, Y. Zhao, APA: adaptive pose alignment for pose-invariant face recognition, IEEE Access 7 (2019) 14653–14670, https://doi.org/10.1109/ACCESS.2019.2894162.
[14] O. Arandjelovic, R. Cipolla, Face recognition from video using the generic shape-illumination manifold, in: European Conference on Computer Vision, Graz, Austria, 2006.
[15] O. Arandjelovic, R. Cipolla, A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution, in: IEEE International Conference on Computer Vision, Piscataway, New Jersey, USA, 2007.
[16] O. Arandjelović, R. Cipolla, A methodology for rapid illumination-invariant face recognition using image processing filters, Comput. Vis. Image Underst. 113 (2009) 159–171, https://doi.org/10.1016/j.cviu.2008.06.008, http://www.sciencedirect.com/science/article/pii/S1077314208000933.
[17] M. Bagga, B. Singh, Spoofing detection in face recognition: a review, in: 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom, 2016, pp. 2037–2042.
[18] X. Bai, B. Yin, Q. Shi, Y. Sun, Face recognition using extended fisherface with 3d morphable model, in: International Conference on Machine Learning and Cybernetics, 2005, pp. 4481–4486.
[19] E. Bailly-Bailliére, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, J.P. Thiran, The BANCA database and evaluation protocol, in: Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication, Springer-Verlag, Berlin, Heidelberg, 2003, pp. 625–638, http://dl.acm.org/citation.cfm?id=1762222.1762304.
[20] W.A. Bainbridge, P. Isola, A. Oliva, The intrinsic memorability of face photographs, J. Exp. Psychol. Gen. 142 (4) (2013) 1323–1334, https://www.wilmabainbridge.com/facememorability2.html.
[21] S. Balaban, Deep learning and face recognition: the state of the art, in: Proceedings, Biometric and Surveillance Technology for Human and Activity, 2015.
[22] A. Bansal, C.D. Castillo, R. Ranjan, R. Chellappa, The do's and don'ts for CNN-based face verification, CoRR abs/1705.07426, http://arxiv.org/abs/1705.07426, arXiv:1705.07426, 2017.
[23] J. Bao, D. Chen, F. Wen, H. Li, G. Hua, CVAE-GAN: fine-grained image generation through asymmetric training, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 2764–2773.
[24] J. Bao, D. Chen, F. Wen, H. Li, G. Hua, Towards open-set identity preserving face synthesis, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 6713–6722.
[25] W. Bao, H. Li, N. Li, W. Jiang, A liveness detection method for face recognition based on optical flow field, in: International Conference on Image Analysis and Signal Processing, Taizhou, China, 2009, pp. 233–236.
[26] J.R. Barr, K.W. Bowyer, P.J. Flynn, S. Biswas, Face recognition from video: a review, Int. J. Pattern Recognit. Artif. Intell. 26 (2012) 1–53, https://doi.org/10.1142/S0218001412660024.
[27] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 711–720.
[28] Y. Bengio, J. Louradour, R. Collobert, J. Weston, Curriculum learning, in: Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada, 2009.
[29] J.R. Beveridge, P.J. Phillips, D.S. Bolme, B.A. Draper, G.H. Givens, Y.M. Lui, M.N. Teli, H. Zhang, W.T. Scruggs, K.W. Bowyer, P.J. Flynn, S. Cheng, The challenge of face recognition from digital point-and-shoot cameras, in: IEEE International Conference on Biometrics: Theory, Applications and Systems, BTAS, Arlington, VA, USA, 2013, pp. 1–8, https://www.cs.colostate.edu/~vision/pasc/index.php.
[30] M.J. Black, Y. Yacoob, Recognizing facial expressions in image sequences using local parameterized models of image motion, Int. J. Comput. Vis. 25 (1997) 23–48.
[31] M. Bodini, A review of facial landmark extraction in 2d images and videos using deep learning, Big Data Cogn. Comput. 3 (2019) 14.
[32] H. Bon-Woo, H. Byun, R. Myoung-Cheol, L. Seong-Whan, Performance evaluation of face recognition algorithms on the Asian face database, KFDB, in: J. Kittler, M.S. Nixon (Eds.), Audio- and Video-Based Biometric Person Authentication, Springer, Berlin, Heidelberg, 2003, pp. 557–565.
[33] K.W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition, Comput. Vis. Image Underst. 101 (2006) 1–15.
[34] M. Breidt, D.W. Cunningham, C. Wallraven, Max Planck video database, https://vdb.kyb.tuebingen.mpg.de/.
[35] V. Bruce, P.J.B. Hancock, A.M. Burton, Human face perception and identification, in: H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie, T.S. Huang (Eds.), Face Recognition: From Theory to Applications, Springer-Verlag, Berlin, Germany, 1998, pp. 51–72.
[36] V. Bruce, A. Young, Understanding face recognition, Br. J. Psychol. 77 (1986) 305–327.
[37] R. Brunelli, T. Poggio, Face recognition through geometrical features, in: European Conference on Computer Vision, Springer, 1992, pp. 792–800.
[38] B. Buyuktas, C.E. Erdem, A.T. Erdem, Curriculum learning for face recognition, in: European Signal Processing Conference, EUSIPCO, 2020.
[39] T.E. de Campos, R.S. Feris, R.M.C. Jr, A framework for face recognition from video sequences using GWN and eigenfeature selection, in: Workshop on Artificial Intelligence and Computer Vision, 2000.
[40] Q. Cao, L. Shen, W. Xie, O.M. Parkhi, A. Zisserman, Vggface2: a dataset for recognising faces across pose and age, CoRR abs/1710.08092, http://arxiv.org/abs/1710.08092, arXiv:1710.08092, 2017.
[41] Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning based descriptor, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 2707–2714.
[42] O. Celiktutan, S. Ulukaya, B. Sankur, A comparative study of face landmarking techniques, Int. J. Image Video Process. 2013 (2013) 13, https://doi.org/10.1186/1687-5281-2013-13.
[43] F. Cen, G. Wang, Dictionary representation of deep features for occlusion-robust face recognition, IEEE Access 7 (2019) 26595–26605, https://doi.org/10.1109/ACCESS.2019.2901376.
[44] H. Cevikalp, H. Serhan Yavuz, Fast and accurate face recognition with image sets, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 1564–1572.
[45] H. Cevikalp, B. Triggs, Face recognition based on image sets, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 2010, pp. 2567–2573.
[46] W. Chai, W. Deng, H. Shen, Cross-generating GAN for facial identity preserving, in: 2018 13th IEEE International Conference on Automatic Face Gesture Recognition, FG 2018, 2018, pp. 130–134.
[47] C.Y. Chang, C.S. Huang, Application of active appearance model for dual-camera face recognition, in: International Conference on Information Security and Intelligence Control, ISIC, 2012, pp. 333–336.
[48] W. Chao, L. Chang, X. Wang, J. Cheng, X. Deng, F. Duan, High-fidelity face sketch-to-photo synthesis using generative adversarial network, in: 2019 IEEE International Conference on Image Processing, ICIP, 2019, pp. 4699–4703.
[49] R. Chellappa, C.L. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, Proc. IEEE 83 (1995) 705–740.
[50] B. Chen, C. Chen, W.H. Hsu, Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset, IEEE Trans. Multimed. 17 (2015) 804–815, https://doi.org/10.1109/TMM.2015.2420374, https://bcsiriuschen.github.io/CARC/.
[51] J. Chen, C. Yang, Y. Deng, G. Zhang, G. Su, Exploring facial asymmetry using optical flow, IEEE Signal Process. Lett. 21 (2014) 792–795, https://doi.org/10.1109/LSP.2014.2316918.
[52] K. Chen, S. Gong, T. Xiang, C.C. Loy, Cumulative attribute space for age and crowd density estimation, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2467–2474, http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html.
[53] L. Chen, Dual linear regression based classification for face cluster recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2673–2680.
[54] L. Chen, H.M. Liao, J. Lin, Person identification using facial motion, in: IEEE International Conference on Image Processing, ICIP, Thessaloniki, Greece, 2001, pp. 677–680.
[55] S. Chen, S. Mau, M.T. Harandi, C. Sanderson, A. Bigdeli, B.C. Lovell, Face recognition from still images to video sequences: a local-feature-based framework, Int. J. Image Video Process. (2011), https://doi.org/10.1155/2011/790598.
[56] W. Chen, K. Wang, H. Jiang, M. Li, Skin color modeling for face detection and segmentation: a review and a new approach, Multimed. Tools Appl. 75 (2016) 839–862, https://doi.org/10.1007/s11042-014-2328-0.
[57] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, Z.J. Wang, Video-based heart rate measurement: recent advances and future prospects, IEEE Trans. Instrum. Meas. 68 (2019) 3600–3615.

[58] H. Cheng, Z. Liu, L. Yang, X. Chen, Sparse representation and learning in visual recognition: theory and applications, Signal Process. 93 (2013) 1408–1425, https://doi.org/10.1016/j.sigpro.2012.09.011.
[59] G. Chetty, Robust audio visual biometric person authentication with liveness verification, in: Intelligent Multimedia Analysis for Security Applications, Springer, 2010, pp. 59–78.
[60] G.G. Chrysos, E. Antonakos, P. Snape, A. Asthana, S. Zafeiriou, A comprehensive performance evaluation of deformable face tracking "in-the-wild", Int. J. Comput. Vis. 126 (2018) 198–232, https://doi.org/10.1007/s11263-017-0999-5.
[61] M. Chrzan, Liveness detection for face recognition, Ph.D. thesis, Masarykova Univerzita, Fakulta Informatiky, 2014.
[62] C.H. Chu, Y.K. Feng, Study of eye blinking to improve face recognition for screen unlock on mobile devices, J. Electr. Eng. Technol. 13 (2018) 953–960.
[63] J.F. Cohn, K. Schmidt, R. Gross, P. Ekman, Individual differences in facial expression: stability over time, relation to self-reported emotion, and ability to inform person identification, in: IEEE International Conference on Multimodal Interfaces, Pittsburgh, PA, USA, 2002, pp. 491–496.
[64] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 6 (2001) 681–685, https://doi.org/10.1109/34.927467.
[65] CRIM, CRIM database, http://www.crim.ca/.
[66] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'05, vol. 1, 2005, pp. 886–893.
[67] A. Dantcheva, F. Bremond, Gender estimation based on smile dynamics, IEEE Trans. Inf. Forensics Secur. 12 (2017) 719–729.
[68] A. Dantcheva, P. Elia, A. Ross, What else does your biometric data reveal? A survey on soft biometrics, IEEE Trans. Inf. Forensics Secur. 11 (2016) 441–467.
[69] DeepFake, Deepfake detection challange, https://www.kaggle.com/c/deepfake-detection-challenge.
[70] H. Demirezen, C.E. Erdem, Remote photoplethysmography using nonlinear mode decomposition, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2018, pp. 1060–1064.
[71] J. Deng, J. Guo, S. Zafeiriou, Arcface: additive angular margin loss for deep face recognition, CoRR abs/1801.07698, http://arxiv.org/abs/1801.07698, arXiv:1801.07698, 2018.
[72] J. Deng, Y. Zhou, S. Zafeiriou, Marginal loss for deep face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 60–68.
[73] W. Deng, J. Hu, J. Guo, Extended SRC: undersampled face recognition via intraclass variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 1864–1870, https://doi.org/10.1109/TPAMI.2012.30.
[74] W. Deng, J. Hu, J. Guo, Extended SRC: undersampled face recognition via intraclass variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 1864–1870, https://doi.org/10.1109/TPAMI.2012.30.
[75] W. Deng, J. Hu, J. Guo, Face recognition via collaborative representation: its discriminant nature and superposed representation, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 2513–2521, https://doi.org/10.1109/TPAMI.2017.2757923.
[76] W. Deng, J. Hu, J. Guo, Compressive binary patterns: designing a robust binary face descriptor with random-field eigenfilters, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2019) 758–767, https://doi.org/10.1109/TPAMI.2018.2800008.
[77] W. Deng, J. Hu, J. Lu, J. Guo, Transform-invariant PCA: a unified approach to fully automatic facealignment, representation, and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014) 1275–1284, https://doi.org/10.1109/TPAMI.2013.194.
[78] O. Deniz, G. Bueno, J. Salido, F.D. la Torre, Face recognition using histograms of oriented gradients, Pattern Recognit. Lett. 32 (2011) 1598–1603, https://doi.org/10.1016/j.patrec.2011.01.004.
[79] O. Deniz, M. Castrillon, M. Hernandez, Face recognition using independent component analysis and support vector machines, Pattern Recognit. Lett. 24 (2003) 2153–2157, https://doi.org/10.1016/S0167-8655(03)00081-3.
[80] H. Dibeklioglu, F. Alnajar, A.A. Salah, T. Gevers, Combining facial dynamics with appearance for age estimation, IEEE Trans. Image Process. 24 (2015) 1928–1943.
[81] H. Dibeklioglu, A.A. Salah, T. Gevers, Are you really smiling at me? Spontaneous versus posed enjoyment smiles, in: European Conference Computer Vision, Springer, Berlin, Heidelberg, 2012, pp. 525–538.
[82] H. Dibeklioglu, A.A. Salah, T. Gevers, Recognition of genuine smiles, IEEE Trans. Multimed. 17 (2015) 279–294.
[83] C. Ding, D. Tao, A comprehensive survey on pose-invariant face recognition, ACM Trans. Intell. Syst. Technol. 7 (2016).
[84] C. Ding, D. Tao, Trunk-branch ensemble convolutional neural networks for video-based face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 1002–1014, https://doi.org/10.1109/TPAMI.2017.2700390.
[85] Z. Ding, Y. Guo, L. Zhang, Y. Fu, One-shot face recognition via generative learning, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2018, IEEE, 2018, pp. 1–7.
[86] T. Do, E. Kijak, Face recognition using co-occurrence histograms of oriented gradients, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Kyoto, Japan, 2012, pp. 1301–1304.
[87] Y. Dong, L. Zhen, L. Shengcai, Z.L. Stan, Learning face representation from scratch, CoRR abs/1411.7923, http://arxiv.org/abs/1411.7923, 2014.
[88] S. Eberz, K.B. Rasmussen, V. Lenders, I. Martinovic, Evaluating behavioral biometrics for continuous authentication: challenges and metrics, in: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, ACM, 2017, pp. 386–399.
[89] R. Ejbali, M. Zaied, C.B. Amar, Face recognition based on beta 2d elastic bunch graph matching, in: 13th International Conference on Hybrid Intelligent Systems, HIS 2013, 2013, pp. 88–92.
[90] M. Elad, Sparse and redundant representation modeling-what next?, IEEE Signal Process. Lett. 19 (2012) 922–928, https://doi.org/10.1109/LSP.2012.2224655.
[91] A. ElSayed, A. Mahmood, T. Sobh, Effect of super resolution on high dimensional features for unsupervised face recognition in the wild, in: IEEE Applied Imagery Pattern Recognition Workshop, AIPR, Washington, DC, USA, 2017.
[92] C.E. Erdem, C. Turan, Z. Aydin, Baum-2: a multilingual audio-visual affective face database, Multimed. Tools Appl. 74 (2015) 7429–7459.
[93] C.E. Erdem, S. Ulukaya, A. Karaali, A.T. Erdem, Combining Haar feature and skin color based classifiers for face detection, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2011, Prague, 2011, pp. 22–57.
[94] N. Erdogmus, S. Marcel, Spoofing face recognition with 3d masks, IEEE Trans. Inf. Forensics Secur. 9 (2014) 1084–1097, https://doi.org/10.1109/TIFS.2014.2322255.
[95] K. Fan, T. Hung, A novel local pattern descriptor-local vector pattern in high-order derivative space for face recognition, IEEE Trans. Image Process. 23 (2014) 2877–2891, https://doi.org/10.1109/TIP.2014.2321495.
[96] M.I. Faraj, J. Bigun, Audio-visual person authentication using lip-motion from orientation maps, Pattern Recognit. Lett. 28 (2007) 1368–1382, https://doi.org/10.1016/j.patrec.2007.02.017.
[97] S.S. Farfade, M.J. Saberian, L.J. Li, Multi-view face detection using deep convolutional neural networks, in: Proc. ACM Int. Conf. Multimedia Retrievals, 2015, pp. 643–650.
[98] A. Franco, D. Maio, F. Turroni, Spatio-temporal keypoints for video-based face recognition, in: International Conference on Pattern Recognition, Stockholm, Sweden, 2014, pp. 489–494.
[99] J. Galbally, S. Marcel, J. Fierrez, Biometric antispoofing methods: a survey in face recognition, IEEE Access 2 (2014) 1530–1552.
[100] M.J. Gangeh, A.K. Farahat, A. Ghodsi, M.S. Kamel, Supervised dictionary learning and sparse representation-a review, CoRR abs/1502.05928, http://arxiv.org/abs/1502.05928, arXiv:1502.05928, 2015.
[101] G. Gao, J. Yang, X.Y. Jing, F. Shen, W. Yang, D. Yue, Learning robust and discriminative low-rank representations for face recognition with occlusion, Pattern Recognit. 66 (2017) 129–143.
[102] G. Gao, J. Yang, S. Wu, X. Jing, D. Yue, Bayesian sample steered discriminative regression for biometric image classification, Appl. Soft Comput. 37 (2015) 48–59.
[103] G. Gao, Y. Yu, M. Yang, H. Chang, P. Huang, D. Yue, Cross-resolution face recognition with pose variations via multilayer locality-constrained structural orthogonal procrustes regression, Inf. Sci. 506 (2020) 19–36.
[104] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, The cas-peal large-scale Chinese face database and baseline evaluations, IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 38 (2008) 149–161, https://doi.org/10.1109/TSMCA.2007.909557, http://www.jdl.ac.cn/peal/.
[105] M. Gavrilescu, Study on using individual differences in facial expressions for a face recognition system immune to spoofing attacks, IET Biometrics 5 (2016) 236–242, https://doi.org/10.1049/iet-bmt.2015.0078.
[106] S. Ge, S. Zhao, C. Li, J. Li, Low-resolution face recognition in the wild via selective knowledge distillation, IEEE Trans. Image Process. 28 (2019) 2051–2062, https://doi.org/10.1109/TIP.2018.2883743.
[107] A. Georghiades, P. Belhumeur, D. Kriegman, Yale face database, [Online] http://cvc.yale.edu/projects/yalefaces/yalefa, 1997.
[108] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 643–660, https://doi.org/10.1109/34.927464, http://vision.ucsd.edu/content/extended-yale-face-database-b-b.
[109] M.M. Ghazi, H.K. Ekenel, A comprehensive analysis of deep learning based representation for face recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, 2016, pp. 102–109.
[110] R. Goh, L. Liu, X. Liu, T. Chen, The CMU face in action (FIA) database, in: International Workshop on Analysis and Modeling of Faces and Gestures, Berlin, 2005, pp. 255–263.
[111] B. Gokberk, H. Dutağacı, A. Ulaş, L. Akarun, B. Sankur, Representation plurality and fusion for 3-d face recognition, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 38 (2008) 155–173, https://doi.org/10.1109/TSMCB.2007.908865.
[112] B. Gokberk, M.O. Irfanoglu, L. Akarun, 3d shape-based face representation and feature extraction for face recognition, Image Vis. Comput. 24 (2006) 857–869, https://doi.org/10.1016/j.imavis.2006.02.009, http://www.sciencedirect.com/science/article/pii/S0262885606000928.

[113] F. Gosselin, P.G. Schyns, Bubbles: a technique to reveal the use of information in recognition tasks, Vis. Res. 41 (2001) 2261–2271, https://doi.org/10.1016/S0042-6989(01)00097-9.

[114] D.B. Graham, N.M. Allinson, Characterizing virtual eigensignatures for general purpose face recognition, in: H. Wechsler, P.J. Phillips, V. Bruce, F. Fogelman-Soulie, T.S. Huang (Eds.), Face Recognition: From Theory to Applications, in: NATO ASI Series F, Computer and Systems Sciences, vol. 163, 1998, pp. 446–456, http://images.ee.umist.ac.uk/danny/database.html.

[115] K. Grm, V. Struc, A. Artiges, M. Caron, H.K. Ekenel, Strengths and weaknesses of deep learning models for face recognition against image degradations, IET Biometrics 7 (2018) 81–89.

[116] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-PIE, Image Vis. Comput. 28 (2010) 807–813, https://doi.org/10.1016/j.imavis.2009.08.002, http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html, Special issue: Best of Automatic Face and Gesture Recognition 2008.

[117] R. Gross, J. Shi, The CMU Motion of Body (MoBo) Database, Technical Report CMU-RI-TR-01-18, Carnegie Mellon University, Pittsburgh, PA, 2001, https://www.ri.cmu.edu/publications/the-cmu-motion-of-body-mobo-database/.

[118] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M.H. Hayes, R.M. Mersereau, Eigenface-domain super-resolution for face recognition, IEEE Trans. Image Process. 12 (2003) 597–606, https://doi.org/10.1109/TIP.2003.811513.

[119] G. Guo, N. Zhang, A survey on deep learning based face recognition, Comput. Vis. Image Underst. 189 (2019) 102805.

[120] Y. Guo, L. Zhang, One-shot face recognition by promoting underrepresented classes, arXiv preprint arXiv:1707.05574, 2017.

[121] Y. Guo, L. Zhang, Y. Hu, X. He, J. Gao, MS-Celeb-1M: a dataset and benchmark for large-scale face recognition, CoRR abs/1607.08221, https://www.microsoft.com/en-us/research/publication/ms-celeb-1m-dataset-benchmark-large-scale-face-recognition-2/, 2016.

[122] R.E. Haamer, K. Kulkarni, N. Imanpour, M.A. Haque, E. Avots, M. Breisch, K. Nasrollahi, S. Escalera, C. Ozcinar, X. Baro, A.R. Naghsh-Nilchi, T.B. Moeslund, G. Anbarjafari, Changes in facial expression as biometric: a database and benchmarks of identification, in: IEEE Conf. on Automatic Face and Gesture Recognition Workshop, China, 2018, pp. 621–628.

[123] A. Hadid, Face biometrics under spoofing attacks: vulnerabilities, countermeasures, open issues, and research directions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 113–118.

[124] A. Hadid, J. Dugelay, M. Pietikainen, On the use of dynamic features in face biometrics: recent advances and challenges, Signal Image Video Process. 5 (2011) 495–506, https://doi.org/10.1007/s11760-011-0247-3.

[125] A. Hadid, M. Pietikainen, An experimental investigation about the integration of facial dynamics in video-based face recognition, Electron. Lett. Comput. Vis. Image Anal. 5 (2005) 1–13, https://doi.org/10.5565/rev/elcvia.80.

[126] A. Hadid, M. Pietikainen, Combining appearance and motion for face and gender recognition from videos, Pattern Recognit. 42 (2009) 2818–2827, https://doi.org/10.1016/j.patcog.2009.02.011.

[127] A. Hadid, M. Pietikáinen, Manifold learning for video-to-video face recognition, in: J. Fierrez, J. Ortega-Garcia, A. Esposito, A. Drygajlo, M. Faundez-Zanuy (Eds.), Biometric ID Management and Multimodal Communication, BioID 2009, in: Lecture Notes in Computer Science, vol. 5707, Springer, Berlin, Heidelberg, 2009.

[128] A. Hadid, M. Pietikainen, S. Li, Learning personal specific facial dynamics for face recognition from videos, in: Lecture Notes in Computer Science, 2007, pp. 1–15.

[129] F. Hajati, M. Tavakolian, S. Gheisari, Y. Gao, A.S. Mian, Dynamic texture comparison using derivative sparse representation: application to video-based face recognition, IEEE Trans. Human-Mach. Syst. 47 (2017) 970–982, https://doi.org/10.1109/THMS.2017.2681425.

[130] P.W. Hallinan, A Deformable Model for the Recognition of Human Faces Under Arbitrary Illumination, Ph.D. thesis, Harvard University, Cambridge, MA, USA, 1995, ftp://ftp.hrl.harvard.edu/pub/faces.

[131] E.M. Hand, R. Chellappa, Attributes for improved attributes: a multi-task network utilizing implicit and explicit relationships for facial attribute classification, in: Proc. AAAI Conf. Artificial Intelligence, 2017, pp. 4068–4074.

[132] M.A. Haque, K. Nasrollahi, T.B. Moeslund, Pain expression as a biometric: why patients' self-reported pain doesn't match with the objectively measured pain?, in: IEEE International Conference on Identity, Security and Behavior Analysis, ISBA, New Delhi, India, 2017.

[133] M.T. Harandi, C. Sanderson, S. Shirazi, B.C. Lovell, Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching, in: CVPR 2011, IEEE, 2011, pp. 2705–2712.

[134] M. Hasnat, J. Bohné, J. Milgram, S. Gentric, L. Chen, et al., von Mises-Fisher mixture model-based deep learning: application to face verification, arXiv preprint arXiv:1706.04264, 2017.

[135] M. Hassaballah, S. Aly, Face recognition: challenges, achievements and future directions, IET Comput. Vis. 9 (2015) 614–626.

[136] S.M. Hatture, P. Karchi, Prevention of spoof attack in biometric system using liveness detection, Int. J. Latest Trends Eng. Technol. (2013) 42–49.

[137] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 770–778.

[138] L. He, H. Li, Q. Zhang, Z. Sun, Dynamic feature matching for partial face recognition, IEEE Trans. Image Process. 28 (2019) 791–802, https://doi.org/10.1109/TIP.2018.2870946.

[139] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacian faces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 328–340, https://doi.org/10.1109/TPAMI.2005.55.

[140] H. Hill, V. Bruce, Effects of lighting on matching facial surfaces, J. Exp. Psychol. Hum. Percept. Perform. 22 (1996) 986–1004.

[141] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.

[142] C. Hsieh, S. Lai, Y. Chen, 2d expression-invariant face recognition with constrained optical flow, in: IEEE International Conference on Multimedia and Expo, New York, NY, USA, 2009, pp. 1058–1061.

[143] C. Hsieh, S. Lai, Y. Chen, An optical flow-based approach to robust face recognition under expression variations, IEEE Trans. Image Process. 19 (2010) 233–240, https://doi.org/10.1109/TIP.2009.2031233.

[144] C. Hu, X. Lu, P. Liu, X. Jing, D. Yue, Single sample face recognition under varying illumination via qrcp decomposition, IEEE Trans. Image Process. 28 (2019) 2624–2638, https://doi.org/10.1109/TIP.2018.2887346.

[145] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[146] P. Hu, D. Ramanan, Finding tiny faces, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 1522–1530.

[147] Y. Hu, A.S. Mian, R. Owens, Face recognition using sparse approximated nearest points between image sets, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 1992–2004.

[148] C. Huang, Y. Li, C.L. Chen, X. Tang, Deep imbalanced learning for face recognition and attribute prediction, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.

[149] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: a survey, IEEE Trans. Syst. Man Cybern., Part C, Appl. Rev. 41 (2011) 765–781, https://doi.org/10.1109/TSMCC.2011.2118750.

[150] G.B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments, in: Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Erik Learned-Miller and Andras Ferencz and Frédéric Jurie, Marseille, France, 2008, https://hal.inria.fr/inria-00321923.

[151] Y. Huang, S. Chen, A geometrical-model-based face recognition, in: IEEE International Conference on Image Processing, ICIP, Quebec City, QC, Canada, 2015, pp. 3106–3110.

[152] Z. Huang, S. Shan, R. Wang, H. Zhang, S. Lao, A. Kuerban, X. Chen, A benchmark and comparative study of video-based face recognition on Cox face database, IEEE Trans. Image Process. 24 (2015) 5967–5981, https://doi.org/10.1109/TIP.2015.249344.

[153] Z. Huang, S. Shan, H. Zhang, H. Lao, A. Kuerban, X. Chen, Benchmarking still-to-video face recognition via partial and local linear discriminant analysis on COX-S2V dataset, in: Asian Conference on Computer Vision, Daejeon, Korea, 2012, pp. 589–600.

[154] Z. Huang, R. Wang, S. Shan, X. Chen, 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, 2015, pp. 140–149.

[155] Z. Huang, X. Zhao, S. Shan, R. Wang, X. Chen, Coupling alignments with recognition for still-to-video face recognition, in: IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 2013, pp. 3296–3303.

[156] A. Jadhav, V.P. Namboodiri, K. Venkatesh, Deep attributes for one-shot face recognition, in: European Conference on Computer Vision, Springer, 2016, pp. 516–523.

[157] R. Jafri, H.R. Arabnia, A survey of face recognition techniques, J. Inf. Process. Syst. 5 (2009) 41–68.

[158] A.K. Jain, A.A. Ross, K. Nandakumar, Introduction to Biometrics, Springer, 2011.

[159] V. Jain, A. Mukherjee, The Indian face database, http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/, 2002.

[160] R. Jenkins, A.M. Burton, 100% accuracy in automatic face recognition, Science 319 (2008) 435, https://doi.org/10.1126/science.1149656.

[161] O. Jesorsky, K.J. Kirchberg, R.W. Frischholz, Robust face detection using the Hausdorff distance, in: J. Bigun, F. Smeraldi (Eds.), Audio- and Video-Based Biometric Person Authentication, Springer, Berlin, Heidelberg, 2001, pp. 90–95, http://www.humanscan.de/support/downloads/facedb.php.

[162] H. Jiang, E. Learned-Miller, Face detection with the faster R-CNN, in: 2017 12th IEEE International Conference on Automatic Face Gesture Recognition, FG 2017, 2017, pp. 650–657.

[163] X. Jin, X. Tan, Face alignment in-the-wild: a survey, Comput. Vis. Image Underst. 162 (2017) 1–22.

[164] T. Kanade, Computer Recognition of Human Faces, Birkhauser Verlag, Basel und Stuttgart, 1977.

[165] T. Kanade, J.F. Cohn, Y. Tian, Comprehensive database for facial expression analysis, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2000, pp. 46–53.

[166] C. Kant, N. Sharma, Fake face recognition using fusion of thermal imaging and skin elasticity, Int. J. Comput. Sci. Commun. 4 (2013) 65–72.

[167] A. Kar, P.P.G. Neogi, Triangular coil pattern of local radius of gyration face for heterogeneous face recognition, Appl. Intell. 50 (2020) 698–716.

[168] A.L. Kashyap, S. Tulyakov, V. Govindaraju, Facial behavior as a soft biometric, in: Proceedings of IAPR International Conference on Biometrics, ICB, New Delhi, India, 2012, pp. 147–151.

[169] V. Kazemi, J. Sullivan, One millisecond face alignment with an ensemble of regression trees, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1867–1874.

[170] N. Kela, A. Rattani, P. Gupta, Illumination invariant elastic bunch graph matching for efficient face recognition, in: IEEE International Conference on Computer Vision and Pattern Recognition Workshop, CVPRW, 2006.

[171] M.A. Khan, C. Xydeas, H. Ahmed, Multi-model aam framework for face image modeling, in: 18th International Conference on Digital Signal Processing, DSP, 2013.

[172] M. Killioglu, M. Taskiran, N. Kahraman, Anti-spoofing in face recognition with liveness detection using pupil tracking, in: 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics, SAMI, 2017, pp. 000087–000092.

[173] K.I. Kim, K. Jung, H.J. Kim, Face recognition using kernel principal component analysis, IEEE Signal Process. Lett. 9 (2002) 40–42, https://doi.org/10.1109/97.991133.

[174] M. Kim, S. Kumar, V. Pavlovic, H. Rowley, Face tracking and recognition with visual constraints in real-world videos, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[175] S.T. Kim, D.H. Kim, Y.M. Ro, Spatio-temporal representation for face authentication by using multi-task learning with human attributes, in: IEEE International Conference on Image Processing, ICIP, Phoenix, AZ, USA, 2016, pp. 2996–3000.

[176] S.T. Kim, Y.M. Ro, Attended relation feature representation of facial dynamics for facial authentication, IEEE Trans. Inf. Forensics Secur. 14 (2019) 1768–1778.

[177] T. Kim, J. Kittler, R. Cipolla, Discriminative learning and recognition of image set classes using canonical correlations, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 1005–1018, https://doi.org/10.1109/TPAMI.2007.1037.

[178] T.K. Kim, J. Kittler, Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 318–327, https://doi.org/10.1109/TPAMI.2005.58.

[179] T.K. Kim, J. Kittler, R. Cipolla, On-line learning of mutually orthogonal subspaces for face recognition by image sets, IEEE Trans. Image Process. 19 (2010) 1067–1074.

[180] B.F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, A.K. Jain, Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1931–1939, https://www.nist.gov/itl/iad/image-group/ijb-dataset-request-form.

[181] B. Knappmeyer, I.M. Thornton, H.H. Bulthoff, The use of facial motion and facial form during the processing of identity, Vis. Res. 43 (2003) 1921–1936, https://doi.org/10.1016/S0042-6989(03)00236-0.

[182] B. Knight, A. Johnston, The role of movement in face recognition, Vis. Cogn. 4 (1997) 265–273, https://doi.org/10.1080/713756764.

[183] K. Kollreider, H. Fronthaler, M.I. Faraj, J. Bigun, Real-time face detection and motion analysis with application in “liveness” assessment, IEEE Trans. Inf. Forensics Secur. 2 (2007) 548–558.

[184] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Conf. on Neural Information Processing Systems, NIPS, 2012, pp. 1097–1105.

[185] V. Kruger, A. Happe, G. Sommer, Affine real-time face tracking using a wavelet network, in: Proceedings International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems. In Conjunction with ICCV'99, Corfu, Greece, 1999, pp. 141–148.

[186] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. Malsburg, R. Wurtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, IEEE Trans. Comput. 42 (1993) 300–311.

[187] A. Lagorio, M. Tistarelli, M. Cadoni, C. Fookes, S. Sridharan, Liveness detection based on 3d face shape analysis, in: 2013 International Workshop on Biometrics and Forensics, IWBF, IEEE, 2013, pp. 1–4.

[188] G. Lavrentyeva, O. Kudashev, A. Melnikov, M. De Marsico, Y. Matveev, Interactive photo liveness for presentation attacks detection, in: International Conference Image Analysis and Recognition, Springer, 2018, pp. 252–258.

[189] K. Lee, J. Ho, M. Yang, D. Kriegman, Visual tracking and recognition using probabilistic appearance manifolds, Computer Vision and Image Understanding, http://vision.ucsd.edu/~leekc/HondaUCSDVideoDatabase/HondaUCSD.html, 2005.

[190] Z. Lei, M. Pietikainen, S.Z. Li, Learning discriminant face descriptor, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014) 289–302.

[191] G. Lenz, S.H. Ieng, R. Benosman, High speed event-based face detection and tracking in the blink of an eye, arXiv preprint arXiv:1803.10106, 2018.

[192] H. Li, P. He, S. Wang, A. Rocha, X. Jiang, A.C. Kot, Learning generalized deep feature representation for face anti-spoofing, IEEE Trans. Inf. Forensics Secur. 13 (2018) 2639–2652.

[193] H. Li, Z. Lin, X. Shen, J. Brandt, G. Hua, A convolutional neural network cascade for face detection, in: Proc. IEEE Conf. Computer Vision Pattern Recognition, 2015, pp. 5325–5334.

[194] S.Z. Li, Z. Lei, M. Ao, The HFB face database for heterogeneous face biometrics research, in: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009, pp. 1–8.

[195] S.Z. Li, D. Yi, Z. Lei, S. Liao, The CASIA NIR-VIS 2.0 face database, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013, pp. 348–353, https://pypi.org/project/bob.db.cbsr-nir-vis-2/.

[196] Y. Li, B. Sun, T. Wu, Y. Wang, Face detection with end-to-end integration of a ConvNet and a 3d model, in: Proc. European Conf. Computer Vision, 2016, pp. 420–436.

[197] S. Liao, A.C.S. Chung, Face recognition with salient local gradient orientation binary patterns, in: 16th IEEE International Conference on Image Processing, ICIP, Cairo, Egypt, 2009, pp. 3317–3320.

[198] L. Lin, K. Wang, D. Meng, W. Zuo, L. Zhang, Active self-paced learning for cost-effective and progressive face identification, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 7–19, https://doi.org/10.1109/TPAMI.2017.2652459.

[199] Y. Lin, S. Cheng, J. Shen, M. Pantic, Mobiface: a novel dataset for mobile face tracking in the wild, in: IEEE Conf. Automatic Face and Gesture Recognition, FG, 2019.

[200] Lina T. Takahashi, I. Ide, H. Murase, Incremental unsupervised-learning of appearance manifold with view-dependent covariance matrix for face recognition from video sequences, IEICE Trans. Inf. Syst. E92.D (2009) 642–652, https://doi.org/10.1587/transinf.E92.D.642.

[201] A. Liu, J. Wan, S. Escalera, H. Jair Escalante, Z. Tan, Q. Yuan, K. Wang, C. Lin, G. Guo, I. Guyon, et al., Multi-modal face anti-spoofing attack detection challenge at CVPR2019, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.

[202] C. Liu, H. Wechsler, Comparative assessment of independent component analysis (ICA) for face recognition, in: The Second International Conference on Audio and Video Based Biometric Oerson Authentication, 1999, pp. 1–6.

[203] C. Liu, H. Wechsler, Independent component analysis of Gabor features for face recognition, IEEE Trans. Neural Netw. 14 (2003) 919–928.

[204] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, SSD: single shot multibox detector, in: Proc. European Conf. Computer Vision, 2016, pp. 21–37.

[205] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, L. Song, Sphereface: deep hypersphere embedding for face recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 212–220.

[206] W. Liu, Y. Wen, Z. Yu, M. Yang, Large-margin softmax loss for convolutional neural networks, in: International Conference on Machine Learning, 2016.

[207] W. Liu, Y.M. Zhang, X. Li, Z. Yu, B. Dai, T. Zhao, L. Song, Deep hyperspherical learning, in: Advances in Neural Information Processing Systems, 2017, pp. 3950–3960.

[208] X. Liu, T. Chen, Face mosaicing for pose robust video-based recognition, in: Asian Conference on Computer Vision, Tokyo, Japan, 2007, pp. 662–671.

[209] X. Liu, T. Cheng, Video-based face recognition using adaptive hidden Markov models, in: IEEE Conference on Computer Vision and Pattern Recognition, Madison, WI, USA, 2003.

[210] Y. Liu, H. Li, X. Wang, Rethinking feature discrimination and polymerization for large-scale recognition, arXiv preprint arXiv:1710.00870, 2017.

[211] Y. Liu, K.L. Schmidt, J.F. Mitra, Facial asymmetry quantification for expression invariant human identification, Comput. Vis. Image Underst. 91 (2003) 138–159, https://doi.org/10.1016/S1077-3142(03)00078-X.

[212] Y. Liu, J. Stehouwer, A. Jourabloo, X. Liu, Deep tree learning for zero-shot face anti-spoofing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 4680–4689.

[213] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2004) 91–110, https://doi.org/10.1023/B:VISI.0000029664.99615.94.

[214] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Face recognition using kernel direct discriminant analysis algorithms, IEEE Trans. Neural Netw. 14 (2003) 117–126, https://doi.org/10.1109/TNN.2002.806629.

[215] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition, Pattern Recognit. Lett. 26 (2005) 181–191, https://doi.org/10.1016/j.patrec.2004.09.014.

[216] J. Lu, G. Wang, P. Moulin, Image set classification using holistic multiple order statistics features and localized multi-kernel metric learning, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 329–336.

[217] J. Lu, G. Wang, J. Zhou, Simultaneous feature and dictionary learning for image set based face recognition, IEEE Trans. Image Process. 26 (2017) 4042–4054.

[218] P. Lucey, J.F. Cohn, K.M. Prkachin, P.E. Solomon, I. Matthews, Painful data: the unbc-mcmaster shoulder pain expression archive database, in: IEEE Interna-

tional Conference on Automatic Face and Gesture Recognition Workshops, FG, Santa Barbara, CA, USA, 2011, pp. 57–64.
[219] M.J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, J. Budynek, The Japanese female facial expression (JAFFE) database, in: IEEE International Conference on Automatic Face and Gesture Recognition, 1998, pp. 14–16, https://zenodo.org/record/3451524#.Xrp3nWgzaUk.
[220] M.A. Khan, C. Xydeas, H. Ahmed, On the application of AAM-based systems in face recognition, in: European Signal Processing Conference, EUSIPCO, 2014, pp. 2445–2449.
[221] H. Mady, S.M.S. Hilles, Face recognition and detection using random forest and combination of LBP and HOG features, in: International Conference on Smart Computing and Electronic Enterprise, ICSCEE, Shah Alam, Malaysia, 2018.
[222] H.H. Mady, S.M. Hilles, Efficient real time attendance system based on face detection case study "mediu staff", Int. J. Contemp. Comput. Res. 1 (2017) 21–25.
[223] Z. Mahmood, N. Muhammad, N. Bibi, T. Ali, A review on state-of-the-art face recognition approaches, Fractals 25 (2017).
[224] S. Marcel, C. McCool, P. Matéjka, T. Ahonen, J. Černocký, S. Chakraborty, et al., On the results of the first mobile biometry (MOBIO) face and speaker verification evaluation, in: Recognizing Patterns in Signals, Speech, Images and Videos, ICPR 2010, 2010, pp. 210–225.
[225] O. Martin, I. Kotsia, B. Macq, I. Pitas, P. Levant, The enterface'05 audio-visual emotion database, in: International Conference on Data Engineering Workshops, Atlanta, GA, USA, 2006, pp. 2–9.
[226] A. Martinez, R. Benavente, The AR Face Database, Technical Report 24, CVC Technical Report, http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html, 1998.
[227] B. Martinez, M.F. Valstar, B. Jiang, Automatic analysis of facial actions: a survey, IEEE Trans. Affect. Comput. 13 (2017).
[228] I. Masi, F. Chang, J. Choi, S. Harel, J. Kim, K. Kim, J. Leksut, S. Rawls, Y. Wu, T. Hassner, W. AbdAlmageed, G. Medioni, L. Morency, P. Natarajan, R. Nevatia, Learning pose-aware models for pose-invariant face recognition in the wild, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2019) 379–393, https://doi.org/10.1109/TPAMI.2018.2792452.
[229] I. Masi, T. Hassner, A.T. Tran, G. Medioni, Rapid synthesis of massive face sets for improved face recognition, in: 2017 12th IEEE International Conference on Automatic Face Gesture Recognition, FG 2017, 2017, pp. 604–611.
[230] I. Masi, A.T. Trán, T. Hassner, J.T. Leksut, G. Medioni, Do we really need to collect millions of faces for effective face recognition?, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision – ECCV 2016, Springer International Publishing, Cham, 2016, pp. 579–596.
[231] I. Masi, Y. Wu, T. Hassner, P. Natarajan, Deep face recognition: a survey, in: SIBGRAPI – Conference on Graphics, Patterns and Images, 2018.
[232] F. Matta, J. Dugelay, A behavioural approach to person recognition, in: IEEE International Conference on Multimedia and Expo, 2006, pp. 1461–1464.
[233] F. Matta, J. Dugelay, Video face recognition: a physiological and behavioural multimodal approach, in: IEEE International Conference on Image Processing, San Antonio, TX, USA, 2007, pp. 497–500.
[234] F. Matta, J. Dugelay, Person recognition using facial video information: a state of the art, J. Vis. Lang. Comput. 20 (2009) 180–187, https://doi.org/10.1016/j.jvlc.2009.01.002.
[235] B. Maze, J. Adams, J.A. Duncan, N. Kalka, T. Miller, C. Otto, A.K. Jain, W.T. Niggel, J. Anderson, J. Cheney, P. Grother, Iarpa Janus benchmark - C: face dataset and protocol, in: 2018 International Conference on Biometrics (ICB), 2018, pp. 158–165, https://www.nist.gov/itl/iad/ig/ijb-c-dataset-request-form.
[236] C. McCool, S. Marcel, MOBIO Database for the ICPR 2010 Face and Speech Competition. Idiap-Com Idiap-Com-02-2009. Idiap, https://www.idiap.ch/dataset/mobio, 2009.
[237] K. Meena, A. Suruliandi, Local binary patterns and its variants for face recognition, in: International Conference on Recent Trends in Information Technology, ICRTIT, Chennai, Tamil Nadu, India, 2011, pp. 782–786.
[238] H. Mendez-Vazquez, Y. Martinez-Diaz, Z. Chai, Volume structured ordinal features with background similarity measure for video face recognition, in: International Conference on Biometrics, ICB, Madrid, Spain, 2013, pp. 1–6.
[239] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Czyz, L. Vandendorpe, et al., Face verification competition on the xm2vts database, in: International Conference on Audio-and Video-Based Biometric Person Authentication, Springer, 2003, pp. 964–974.
[240] K. Messer, J. Matas, J. Kittler, J. Luettin, G. Maitre, Xm2vtsdb: the extended m2vts database, in: Second International Conference on Audio and Video-Based Biometric Person Authentication, 1999, pp. 965–966, http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/.
[241] A. Mian, Unsupervised learning from local features for video-based face recognition, in: 2008 8th IEEE International Conference on Automatic Face Gesture Recognition, 2008, pp. 1–6.
[242] W. Miaoli, Face and speech recognition fusion method based on penalty coefficient and SVM, in: IEEE Advanced Information Technology, Electronic and Automation, Control Conference, IAEAC, Chongqing, China, 2015, pp. 6–10.
[243] S. Milborrow, J. Morkel, F. Nicolls, The MUCT landmarked face database, Pattern recognition association of South Africa, http://www.milbo.org/muct, 2010.
[244] J. Mohapatra, T.W. Weng, P.Y. Chen, S. Liu, L. Daniel, Towards verifying robustness of neural networks against semantic perturbations, in: IEEE Conf. Computer Vision and Pattern Recognition, CVPR, 2020.
[245] F. Mokhayeri, E. Granger, A paired sparse representation model for robust face recognition from a single sample, Pattern Recognit. 100 (2020) 107129.
[246] F. Mokhayeri, E. Granger, G. Bilodeau, Domain-specific face synthesis for video face recognition from a single sample per person, IEEE Trans. Inf. Forensics Secur. 14 (2019) 757–772, https://doi.org/10.1109/TIFS.2018.2866295.
[247] H. Murase, S.K. Nayar, Visual learning and recognition of 3-d objects from appearance, Int. J. Comput. Vis. 14 (1995) 5–24.
[248] C. Nagpal, S.R. Dubey, A performance evaluation of convolutional neural networks for face anti spoofing, in: 2019 International Joint Conference on Neural Networks, IJCNN, IEEE, 2019, pp. 1–8.
[249] A. Nech, I. Kemelmacher-Shlizerman, Level playing field for million scale face recognition, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 3406–3415, http://megaface.cs.washington.edu/.
[250] A. Nefian, M.H. Hayes III, A hidden Markov model-based approach for face detection and recognition, Ph.D. thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, 1999, http://www.anefian.com/research/face_reco.htm.
[251] O. Nikisins, M. Greitans, Reduced complexity automatic face recognition algorithm based on local binary patterns, in: 19th International Conference on Systems, Signals and Image Processing, IWSSIP, Vienna, Austria, 2012, pp. 433–436.
[252] Y. Ning, T. Sim, Smile, you're on identity camera, in: International Conference on Pattern Recognition, Tampa, FL, USA, 2008.
[253] K.A. Nixon, V. Aimale, R.K. Rowe, Spoof detection schemes, in: Handbook of Biometrics, Springer, 2008, pp. 403–423.
[254] E.M. Nowara, A. Sabharwal, A. Veeraraghavan, Ppgsecure: biometric presentation attack detection using photopletysmograms, in: 2017 12th IEEE International Conference on Automatic Face Gesture Recognition, FG 2017, 2017, pp. 56–62.
[255] T. Ojala, M. Pietikáinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognit. 29 (1996) 51–59, https://doi.org/10.1016/0031-3203(95)00067-4, http://www.sciencedirect.com/science/article/pii/0031320395000674.
[256] A.J. O'Toole, J. Harms, S.L. Snow, D.R. Hurst, M.R. Pappas, J.H. Ayyad, H. Abdi, A video database of moving faces and people, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 812–816, https://doi.org/10.1109/TPAMI.2005.90, http://www.utdallas.edu/dept/bbs/FACULTY_PAGES/otoole/database.htm.
[257] A.J. O'Toole, D.A. Roark, H. Abdi, Recognizing moving faces: a psychological and neural synthesis, Trends Cogn. Sci. 6 (2002) 261–266, https://doi.org/10.1016/S1364-6613(02)01908-3.
[258] G. Pala, C.E. Erdem, Performance comparison of deep learning based face identification methods for video under adverse conditions, in: The 15th Int. Conf. on Signal-Image Technology and Internet Based Systems, SITIS, 2019.
[259] M. Paleari, C. Velardo, B. Huet, J. Dugelay, S. Antipolis, Face dynamics for biometric people recognition, in: IEEE International Workshop on Multimedia Signal Processing, Rio de Janeiro, Brazil, 2009.
[260] G. Pan, Z. Wu, L. Sun, Liveness detection for face recognition, in: Recent Advances in Face Recognition, IntechOpen, 2008.
[261] U. Park, A.K. Jain, 3d model-based face recognition in video, in: International Conference on Biometrics, Seoul, Korea, 2007, pp. 1085–1094.
[262] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: British Machine Vision Conference, Swansea, UK, 2015.
[263] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: BMVC, 2015, https://www.robots.ox.ac.uk/~vgg/data/vgg_face/.
[264] C. Peng, X. Gao, N. Wang, J. Li, Graphical representation for heterogeneous face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2016) 301–312.
[265] P.J. Phillips, Human identification technical challenges, in: Proceedings. International Conference on Image Processing, 2002, http://www.nd.edu/cvrl/HID-data.html.
[266] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, K. Hoffman, J. Marques, W. Worek, Overview of the face recognition grand challenge, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR'05, vol. 1, 2005, pp. 947–954, https://cvrl.nd.edu/projects/data/#face-recognition-grand-challenge-frgc-v20-data-collection.
[267] P.J. Phillips, P. Grother, R. Micheals, D.M. Blackburn, E. Tabassi, M. Bone, Face recognition vendor test 2002, in: 2003 IEEE International SOI Conference. Proceedings (Cat. No.03CH37443), 2003, p. 44.
[268] P.J. Phillips, H. Wechsler, J. Huang, P. Rauss, The feret database and evaluation procedure for face-recognition algorithms, Image Vis. Comput. 16 (1998) 295–306, https://nist.gov/itl/products-and-services/color-feret-database.
[269] P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, M. Sharpe, Frvt 2006 and ice 2006 large scale experimental results, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 831–846, https://doi.org/10.1109/TPAMI.2009.59.
[270] N. Poh, S. Bengio, Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication, Pattern Recognit. 39 (2006) 223–233, https://doi.org/10.1016/j.patcog.2005.06.011.

[271] N. Poh, C. Chan, J. Kittler, J.F. UAM, J.G. UAM, Description of metrics for the evaluation of biometric performance, Technical Report, (BEAT) Biometrics Evaluation and Testing https://www.beat-eu.org/project/deliverables-public/d3.3-description-of-metrics-for-the-evaluation-of-biometric-performance, 2011.

[272] X. Qi, L. Zhang, Face recognition via centralized coordinate learning, arXiv preprint arXiv:1801.05678, 2018.

[273] Y. Qian, W. Deng, J. Hu, Task specific networks for identity and face variation, in: 2018 13th IEEE International Conference on Automatic Face Gesture Recognition, FG 2018, 2018, pp. 271–277.

[274] Q. Qiu, R. Chellappa, Compositional dictionaries for domain adaptive face recognition, IEEE Trans. Image Process. 24 (2015) 5152–5165, https://doi.org/10.1109/TIP.2015.2479456.

[275] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE 77 (1989) 257–286.

[276] R. Ramachandra, C. Busch, Presentation attack detection methods for face recognition systems, ACM Comput. Surv. 50 (2017) 1–37.

[277] R. Ranjan, C.D. Castillo, R. Chellappa, L2-constrained softmax loss for discriminative face verification, arXiv preprint arXiv:1703.09507, 2017.

[278] R. Ranjan, V.M. Patel, R. Chellappa, Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2019) 121–135, https://doi.org/10.1109/TPAMI.2017.2781233.

[279] R. Ranjan, S. Sankaranarayanan, A. Bansal, N. Bodla, J.C. Chen, V.M. Patel, C.D. Castillo, R. Chellappa, Deep learning for understanding faces: machines may be just as good, or better, than humans, IEEE Signal Process. Mag. 35 (2018) 66–83.

[280] R. Ranjan, S. Sankaranarayanan, C.D. Castillo, R. Chellappa, An all-in-one convolutional neural network for face analysis, in: 2017 12th IEEE International Conference on Automatic Face Gesture Recognition, FG 2017, 2017, pp. 17–24.

[281] Y. Rao, J. Lin, J. Lu, J. Zhou, Learning discriminative aggregation network for video-based face recognition, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 3801–3810.

[282] Y. Rao, J. Lu, J. Zhou, Attention-aware deep reinforcement learning for video face recognition, in: 2017 IEEE International Conference on Computer Vision, ICCV, 2017, pp. 3951–3960.

[283] D. Rathod, A. Vinay, S. Shylaja, S. Natarajan, Facial landmark localization-a literature survey, Int. J. Curr. Eng. Technol. 4 (2014) 1901–1907.

[284] S. Ren, X. Cao, Y. Wei, J. Sun, Face alignment at 3000 FPS via regressing local binary features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1685–1692.

[285] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, in: Proc. Advances Neural Information Processing Systems Conf., 2015, pp. 91–99.

[286] K. Ricanek, T. Tesafaye, Morph: a longitudinal image database of normal adult age-progression, in: 7th International Conference on Automatic Face and Gesture Recognition, FGR06, 2006, pp. 341–345, https://ebill.uncw.edu/C20231_ustores/web/product_detail.jsp?PRODUCTID=8.

[287] S.A. Rizvi, P.J. Phillips, H. Moon, The feret verification testing protocol for face recognition algorithms, in: Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, IEEE, 1998, pp. 48–53.

[288] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.

[289] R. Rubinstein, A.M. Bruckstein, M. Elad, Dictionaries for sparse representation modeling, Proc. IEEE 98 (2010) 1045–1057, https://doi.org/10.1109/JPROC.2010.2040551.

[290] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.

[291] U. Saeed, J. Dugelay, Person recognition form video using facial mimics, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Honolulu, HI USA, 2007, pp. 493–496.

[292] U. Saeed, F. Matta, J. Dugelay, Person recognition based on head and mouth dynamics, in: IEEE Workshop on Multimedia Signal Processing, Victoria, BC, Canada, 2006, pp. 29–32.

[293] A. Samal, P. Iyengar, Automatic recognition and analysis of human faces and facial expressions: a survey, Pattern Recognit. 25 (1992) 65–77.

[294] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: IEEE Workshop on Applications of Computer Vision, 1994, pp. 138–142, http://cam-orl.co.uk/facedatabase.html.

[295] E.N. Sandikci, C.E. Erdem, S. Ulukaya, A comparison of facial landmark detection methods, in: IEEE Signal Processing and Applications Conference, SIU, 2018.

[296] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus database for 3d face analysis, in: B. Schouten, N.C. Juul, A. Drygajlo, M. Tistarelli (Eds.), Biometrics and Identity Management, Springer, Berlin, Heidelberg, 2008, pp. 47–56, http://bosphorus.ee.boun.edu.tr/Home.aspx.

[297] A. Scheenstra, A. Ruifrok, R.C. Veltkamp, A survey of 3d face recognition methods, in: International Conference on Audio and Video Based Biometric Person Authentication, 2005, pp. 891–899.

[298] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, C. Busch, Face recognition systems under morphing attacks: a survey, IEEE Access 7 (2019) 23012–23026, https://doi.org/10.1109/ACCESS.2019.2899367.

[299] F. Schroff, D. Kalenichenko, J. Philbin, Facenet: a unified embedding for face recognition and clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Boston, MA, USA, 2015, pp. 815–823.

[300] S. Sengupta, J. Chen, C. Castillo, V.M. Patel, R. Chellappa, D.W. Jacobs, Frontal to profile face verification in the wild, in: IEEE Winter Conference on Applications of Computer Vision, WACV, 2016, pp. 1–9, http://www.cfpw.io/.

[301] R. Shao, X. Lan, P.C. Yuen, Joint discriminative learning of deep dynamic textures for 3d mask face anti-spoofing, IEEE Trans. Inf. Forensics Secur. 14 (2019) 923–938, https://doi.org/10.1109/TIFS.2018.2868230.

[302] Y. Shen, C. Chiu, Local binary pattern orientation based face recognition, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Brisbane, QLD, Australia, 2015, pp. 1091–1095.

[303] Y. Shen, P. Luo, P. Luo, J. Yan, X. Wang, X. Tang, FaceID-GAN: learning a symmetry three-player GAN for identity-preserving face synthesis, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 821–830.

[304] J.W. Shepherd, G.M. Davies, H.D. Ellis, Cue saliency in faces as assessed by the "photofit" technique, in: G.M. Davies, H.D. Ellis, J.W. Shepherd (Eds.), Studies of Cue Saliency, Academic Press, London, U.K., 1981.

[305] M. Shreve, E.A. Bernal, Q. Li, J. Kumar, R. Bala, A study on the discriminability of facs from spontaneous facial expressions, in: IEEE International Conference on Image Processing, ICIP, Phoenix, AZ, USA, 2016, pp. 1674–1678.

[306] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression database, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 1615–1618, https://doi.org/10.1109/TPAMI.2003.1251154, http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html.

[307] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Comput. Vis. Pattern Recognit. (2015) 1–14, https://arxiv.org/pdf/1409.1556.pdf.

[308] M. Singh, A. Arora, A novel face liveness detection algorithm with multiple liveness indicators, Wirel. Pers. Commun. 100 (2018) 1677–1687.

[309] D.A. Socolinsky, L.B. Wolff, J.D. Neuheisel, C.K. Eveland, Illumination invariant face recognition using thermal infrared imagery, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2001, pp. I–I.

[310] F. Solina, P. Peer, B. Batagelj, S. Juvan, J. Kovač, Color-based face detection in the "15 seconds of fame" art installation, in: Proceedings of Mirage 2003, INRIA Rocquencourt, France, 2003, pp. 37–47, http://www.lrv.fri.uni-lj.si/facedb.html.

[311] H. Song, U. Yang, S. Lee, K. Sohn, 3d face recognition based on facial shape indexes with dynamic programming, in: D. Zhang, A.K. Jain (Eds.), Advances in Biometrics, Springer, Berlin, Heidelberg, 2005, pp. 99–105.

[312] M. Soriano, E. Marszalec, M. Pietikainen, Physics-based face database for color research, J. Electron. Imaging 9 (2000) 32–38, http://www.cse.oulu.fi/CMV/Downloads/Pbfd.

[313] J. Stallkamp, H.K. Ekenel, R. Stiefelhagen, Video-based face recognition on real-world data, in: IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007.

[314] H. Sun, X. Zhen, Y. Zheng, G. Yang, Y. Yin, S. Li, Learning deep match kernels for image-set classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3307–3316.

[315] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Proc. Advances in Neural Information Processing Systems, NIPS, 2014, pp. 1988–1996, http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html.

[316] Y. Sun, D. Liang, X. Wang, X. Tang, DeepID3: face recognition with very deep neural networks, arXiv preprint arXiv:1502.00873, 2015.

[317] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 1891–1898.

[318] C. Szegedy, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 1–9.

[319] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Web-scale training for face identification, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Boston, MA, USA, 2015, pp. 2746–2754.

[320] Y. Taigman, M. Yanga, M. Ranzato, L. Wolf, DeepFace: closing the gap to human-level performance in face verification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.

[321] M. Taskiran, M. Killioglu, N. Kahraman, C.E. Erdem, Face recognition using dynamic features extracted from smile videos, in: IEEE Int. Symp. on Innovations in Intelligent Systems and Applications, INISTA, Sofia, Bulgaria, 2019.

[322] Y. Tayal, R. Lamba, S. Padhee, Automatic face detection using color based segmentation, Int. J. Sci. Res. Publ. 2 (2012) 1–7.

[323] J. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.

[324] D. Thomas, K.W. Bowyer, P.J. Flynn, Multi-factor approach to improving recognition performance in surveillance-quality video, in: IEEE Second International Conference on Biometrics: Theory, Applications and Systems, Arlington, VA, USA, 2008.
[325] C.E. Thomaz, G.A. Giraldi, A new ranking method for principal components analysis and its application to face image analysis, https://fei.edu.br/~cet/facedatabase.html, 2010.
[326] M. Tistarelli, M. Bicego, E. Grosso, Dynamic face recognition: from human to machine vision, Image Vis. Comput. 27 (2009) 222–232, https://doi.org/10.1016/j.imavis.2007.05.006.
[327] I. Tosic, P. Frossard, Dictionary learning, IEEE Signal Process. Mag. 28 (2011) 27–38, https://doi.org/10.1109/MSP.2010.939537.
[328] L. Tran, X. Yin, X. Liu, Disentangled representation learning gan for pose-invariant face recognition, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2017, pp. 1283–1292.
[329] P. Tsai, L. Cao, T. Hintz, T. Jan, A bi-modal face recognition framework integrating facial expression with facial appearance, Pattern Recognit. Lett. 30 (2009) 1096–1109, https://doi.org/10.1016/j.patrec.2009.05.008.
[330] M.W. Tsigie, R. Thakare, R. Joshi, Face recognition techniques based on 2d local binary pattern, histogram of oriented gradient and multiclass support vector machines for secure document authentication, in: Second International Conference on Inventive Communication and Computational Technologies, ICICCT, Coimbatore, India, 2018, pp. 1671–1676.
[331] S. Tulyakov, T. Slowe, Z. Zhang, V. Govindaraju, Facial expression biometrics using tracker displacement features, in: IEEE Int. Conf. Computer Vision and Pattern Recognition, CVPR, Minneapolis, MN, USA, 2007.
[332] M. Turk, A. Pentland, Eigenfaces for face recognition, J. Cogn. Neurosci. 3 (1991) 71–86.
[333] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: IEEE Conference on Computer Vision and Pattern Recognition, 1991, pp. 586–591.
[334] L. Vaina, J. Solomon, S. Chowdhury, P. Sinha, J. Belliveau, Functional neuroanatomy of biological motion perception in humans, Proc. Natl. Acad. Sci. USA 98 (2001).
[335] E. Vezzetti, F. Marcolin, 3D human face description: landmarks measures and geometrical features, Image Vis. Comput. 30 (2012) 698–712, https://doi.org/10.1016/j.imavis.2012.02.007, http://www.sciencedirect.com/science/article/pii/S0262885612000224, 3D Facial Behaviour Analysis and Understanding.
[336] E. Vezzetti, F. Marcolin, Geometrical descriptors for human face morphological analysis and recognition, Robot. Auton. Syst. 60 (2012) 928–939.
[337] C. Vinette, F. Gosselin, P.G. Schyns, Spatio-temporal dynamics of face recognition in a flash: it's in the eyes, Cogn. Sci. 28 (2004) 289–301, https://doi.org/10.1207/s15516709cog2802-8.
[338] P. Viola, M.J. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57 (2004) 137–154.
[339] D. Wang, S. Kong, A classification-oriented dictionary learning model: explicitly learning the particularity and commonality across categories, Pattern Recognit. 47 (2014) 885–898, https://doi.org/10.1016/j.patcog.2013.08.004, http://www.sciencedirect.com/science/article/pii/S0031320313003245.
[340] F. Wang, J. Cheng, W. Liu, H. Liu, Additive margin softmax for face verification, IEEE Signal Process. Lett. 25 (2018) 926–930.
[341] F. Wang, X. Xiang, J. Cheng, A.L. Yuille, Normface: L2 hypersphere embedding for face verification, in: Proceedings of the 25th ACM International Conference on Multimedia, ACM, 2017, pp. 1041–1049.
[342] H. Wang, Y. Wang, Y. Cao, Video-based face recognition: a survey, Int. J. Comput. Inf. Eng. 3 (2009) 293–302.
[343] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, W. Liu, Cosface: large margin cosine loss for deep face recognition, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5265–5274.
[344] H. Wang, D. Zhang, Z. Miao, Fusion of ldb and hog for face recognition, in: 37th Chinese Control Conference, CCC, Wuhan, China, 2018, pp. 9192–9196.
[345] L. Wang, Y. Li, S. Wang, Feature learning for one-shot face recognition, in: 2018 25th IEEE International Conference on Image Processing, ICIP, IEEE, 2018, pp. 2386–2390.
[346] M. Wang, W. Deng, Deep face recognition: a survey, arXiv:1804.06655, 2019.
[347] M.J. Wang, Face feature dynamic recognition method based on intelligent image, in: International Conference on Virtual Reality and Intelligent Systems, ICVRIS, Changsha, China, 2018.
[348] N. Wang, X. Gao, D. Tao, H. Yang, X. Li, Facial feature point detection, Neurocomputing 275 (2018) 50–65, https://doi.org/10.1016/j.neucom.2017.05.013.
[349] R. Wang, S. Shan, X. Chen, W. Gao, Manifold-manifold distance with application to face recognition based on image set, in: IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008.
[350] W. Wang, R. Wang, Z. Huang, S. Shan, X. Chen, Discriminant analysis on Riemannian manifold of Gaussian distributions for face recognition with image sets, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2048–2057.
[351] W. Wang, R. Wang, Z. Huang, S. Shan, X. Chen, Discriminant analysis on Riemannian manifold of Gaussian distributions for face recognition with image sets, IEEE Trans. Image Process. 27 (2018) 151–163.
[352] X. Wang, X. Tang, Dual-space linear discriminant analysis for face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2004.
[353] X. Wang, X. Tang, Face photo-sketch synthesis and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 1955–1967, https://doi.org/10.1109/TPAMI.2008.222, http://mmlab.ie.cuhk.edu.hk/archive/facesketch.html.
[354] X. Wang, X. Tang, Face photo-sketch synthesis and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 1955–1967, https://doi.org/10.1109/TPAMI.2008.222, http://mmlab.ie.cuhk.edu.hk/archive/cufsf/.
[355] C. Wei, Y.F. Wang, Undersampled face recognition via robust auxiliary dictionary learning, IEEE Trans. Image Process. 24 (2015) 1722–1734.
[356] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. 10 (2009) 207–244, http://dl.acm.org/citation.cfm?id=1577069.1577078.
[357] Y. Weiwei, Face recognition using constrained active appearance model, in: Third International Symposium on Intelligent Information Technology Application Workshops, 2009, pp. 348–351.
[358] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: European Conference on Computer Cision, Springer, 2016, pp. 499–515.
[359] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J.C. Adams, T. Miller, N.D. Kalka, A.K. Jain, J.A. Duncan, K.E. Allen, J. Cheney, P. Grother, Iarpa Janus benchmark-b face dataset, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, 2017, pp. 592–600, https://www.nist.gov/programs-projects/face-challenges.
[360] L. Wiskott, J.M. Fellous, N. Kruger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 775–779, https://doi.org/10.1109/ICIP.1997.647401.
[361] L. Wolf, T. Hassner, I. Maoz, Face recognition in unconstrained videos with matched background similarity, in: IEEE Int. Conf. Computer Vision and Pattern Recognition, CVPR, Colorado Springs, CO, USA, 2011, pp. 529–534, https://www.cs.tau.ac.il/~wolf/ytfaces/, https://doi.org/10.1109/CVPR.2011.5995566.
[362] D.K. Wong, R. Janakiraman, Face liveness detection, US Patent App. 15/610,273, 2018.
[363] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 210–227, https://doi.org/10.1109/TPAMI.2008.79.
[364] F. Wu, X.Y. Jing, X. Dong, R. Hu, D. Yue, L. Wang, Y.M. Ji, R. Wang, G. Chen, Intraspectrum discrimination and interspectrum correlation analysis deep network for multispectral face recognition, IEEE Trans. Cybern. 50 (2020) 1009–1022.
[365] Y. Wu, T. Hassner, K. Kim, G. Medioni, P. Natarajan, Facial landmark detection with tweaked convolutional neural networks, IEEE Trans. Pattern Anal. Mach. Intell. 40 (2018) 3067–3074, https://doi.org/10.1109/TPAMI.2017.2787130.
[366] Y. Wu, Q. Ji, Facial landmark detection: a literature survey, Int. J. Comput. Vis. 127 (2019) 115–142, https://doi.org/10.1007/s11263-018-1097-z.
[367] Y. Wu, H. Liu, Y. Fu, Low-shot face recognition with hybrid classifiers, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1933–1939.
[368] S. Xie, Z. Tu, Holistically-nested edge detection, in: 2015 IEEE International Conference on Computer Vision, ICCV, 2015, pp. 1395–1403.
[369] Z. Xie, J. Zeng, G. Liu, Z. Fang, A novel infrared face recognition based on local binary pattern, in: International Conference on Wavelet Analysis and Pattern Recognition, Guilin, China, 2011, pp. 55–59.
[370] E.P. Xing, M.I. Jordan, S.J. Russell, A.Y. Ng, Distance metric learning with application to clustering with side-information, in: S. Becker, S. Thrun, K. Obermayer (Eds.), Advances in Neural Information Processing Systems, vol. 15, MIT Press, 2003, pp. 521–528, http://papers.nips.cc/paper/2164-distance-metric-learning-with-application-to-clustering-with-side-information.pdf, 2003.
[371] X. Xiong, F. De la Torre, Supervised descent method and its applications to face alignment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 532–539.
[372] Y. Xu, Z. Li, J. Yang, D. Zhang, A survey of dictionary learning algorithms for face recognition, IEEE Access 5 (2017) 8502–8514.
[373] L. Xue-fang, P. Tao, Realization of face recognition system based on Gabor wavelet and elastic bunch graph matching, in: 25th Chinese Control and Decision Conference, CCDC, 2013, pp. 3384–3386.
[374] O. Yamaguchi, K. Fukui, K. Maeda, Face recognition using temporal image sequence, in: IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998, pp. 318–323.
[375] J. Yang, Q. Liu, K. Zhang, Stacked hourglass network for robust facial landmark localisation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, 2017, pp. 2025–2033.
[376] J. Yang, S.E. Reed, M.H. Yang, H. Lee, Weakly-supervised disentangling with recurrent transformations for 3d view synthesis, in: C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett (Eds.), in: Advances in Neural Information Processing Systems, vol. 28, Curran Associates, Inc., 2015, pp. 1099–1107, http://papers.nips.cc/paper/5639-weakly-supervised-disentangling-with-recurrent-transformations-for-3d-view-synthesis.pdf, 2015.
[377] J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 131–137.

[378] J. Yang, D. Zhang, J. Yang, B. Niu, Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 650–664, https://doi.org/10.1109/TPAMI.2007.1008.

[379] M. Yang, W. Liu, W. Luo, L. Shen, Analysis-synthesis dictionary learning for universality-particularity representation based classification, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI Press, 2016, pp. 2251–2257, http://dl.acm.org/citation.cfm?id=3016100.3016213.

[380] M. Yang, L. Zhang, X. Feng, D. Zhang, Sparse representation based Fisher discrimination dictionary learning for image classification, Int. J. Comput. Vis. 109 (2014) 209–232, https://doi.org/10.1007/s11263-014-0722-8.

[381] M.H. Yang, Face recognition using extended isomap, in: IEEE International Conference on Image Processing, 2002, pp. 117–120.

[382] S. Yang, P. Luo, C.C. Loy, X. Tang, From facial parts responses to face detection: a deep learning approach, in: Proc. IEEE Int. Conf. Computer Vision, 2015, pp. 3676–3684.

[383] S. Yang, Y. Xiong, C.C. Loy, X. Tang, Face detection through scale-friendly deep convolutional networks, CoRR abs/1706.02863, http://arxiv.org/abs/1706.02863, arXiv:1706.02863, 2017.

[384] S. Yang, L. Zhang, L. He, Y. Wen, Sparse low-rank component-based representation for face recognition with low-quality images, IEEE Trans. Inf. Forensics Secur. 14 (2019) 251–261, https://doi.org/10.1109/TIFS.2018.2849883.

[385] X. Yang, K.T. Cheng, Local difference binary for ultrafast and distinctive feature description, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2014) 188–194, https://doi.org/10.1109/TPAMI.2013.150.

[386] X. Yang, W. Luo, L. Bao, Y. Gao, D. Gong, S. Zheng, Z. Li, W. Liu, Face anti-spoofing: model matters, so does data, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3507–3516.

[387] J. Yim, H. Jung, B. Yoo, C. Choi, D. Park, J. Kim, Rotating your face using multi-task deep neural network, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 676–684.

[388] L. Yin, X. Chen, Y. Sun, T. Worm, M. Reale, A high-resolution 3d dynamic facial expression database, in: 2008 8th IEEE International Conference on Automatic Face Gesture Recognition, 2008, pp. 1–6, http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html.

[389] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, A 3d facial expression database for facial behavior research, in: International Conference on Automatic Face and Gesture Recognition, FGR06, 2006, pp. 211–216, http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html.

[390] L. Yin, X. Wei, Y. Sun, J. Wang, M.J. Rosato, A 3d facial expression database for facial behavior research, in: International Conference on Automatic Face and Gesture Recognition, FGR, Southampton, UK, 2006, pp. 211–216.

[391] J. Yu, X. Xu, F. Gao, S. Shi, M. Wang, D. Tao, Q. Huang, Toward realistic face photo-sketch synthesis via composition-aided gans, IEEE Trans. Cybern. (2020) 1–13.

[392] S. Zafeiriou, M. Pantic, Facial behaviometrics: the case of facial deformation in spontaneous smile/laughter, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops, Colorado Springs, CO, USA, 2011, pp. 13–19.

[393] E. Zangeneh, M. Rahmati, Y. Mohsenzadeh, Low resolution face recognition using a two-branch deep convolutional neural network architecture, Expert Syst. Appl. 139 (2020) 112854.

[394] S. Zhalehpour, O. Onder, Z. Akhtar, C.E. Erdem, BAUM-1: a spontaneous audio-visual face database of affective and mental states, IEEE Trans. Affect. Comput. 8 (2017) 300–313, https://doi.org/10.1109/TAFFC.2016.2553038.

[395] S. Zhalehpour, Z. Akhtar, C.E. Erdem, Multimodal emotion recognition based on peak frame selection from video, Signal Image Video Process. 10 (2016) 827–834, https://doi.org/10.1007/s11760-015-0822-0.

[396] D. Zhang, Z.H. Zhou, S. Chen, Diagonal principal component analysis for face recognition, Pattern Recognit. 39 (2006) 140–142.

[397] J. Zhang, S. Shan, M. Kan, X. Chen, Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment, in: European Conference on Computer Vision, Springer, 2014, pp. 1–16.

[398] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, Joint face detection and alignment using multitask cascaded convolutional networks, IEEE Signal Process. Lett. 23 (2016) 1499–1503, https://doi.org/10.1109/LSP.2016.2603342.

[399] M. Zhang, N. Wang, Y. Li, X. Gao, Neural probabilistic graphical model for face sketch synthesis, IEEE Trans. Neural Netw. Learn. Syst. (2019).

[400] S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, S.Z. Li, A dataset and benchmark for large-scale multi-modal face anti-spoofing, arXiv preprint arXiv:1812.00408, 2018.

[401] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, S.Z. Li, S^3 FD: single shot scale-invariant face detector, CoRR abs/1708.05237, http://arxiv.org/abs/1708.05237, 2017.

[402] W. Zhang, X. Zhao, J. Morvan, L. Chen, Improving shadow suppression for illumination robust face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 41 (2019) 611–624, https://doi.org/10.1109/TPAMI.2018.2803179.

[403] X. Zhang, Z. Fang, Y. Wen, Z. Li, Y. Qiao, Range loss for deep face recognition with long-tailed training data, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 5409–5418.

[404] X. Zhang, Y. Gao, Face recognition across pose: a review, Pattern Recognit. 42 (2009) 2876–2896.

[405] Z. Zhang, X. Chen, B. Wang, G. Hu, W. Zuo, E.R. Hancock, Face frontalization using an appearance-flow-based convolutional neural network, IEEE Trans. Image Process. 28 (2019) 2187–2199, https://doi.org/10.1109/TIP.2018.2883554.

[406] Z. Zhang, Y. Xu, J. Yang, X. Li, D. Zhang, A survey of sparse representation: algorithms and applications, IEEE Access 3 (2015) 490–530, https://doi.org/10.1109/ACCESS.2015.2430359.

[407] H. Zhao, P.C. Yuen, J.T. Kwok, A novel incremental principal component analysis and its application for face recognition, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 36 (2006) 873–886.

[408] J. Zhao, L. Xiong, P. Karlekar Jayashree, J. Li, F. Zhao, Z. Wang, P. Sugiri Pranata, P. Shengmei Shen, S. Yan, J. Feng, Dual-agent GANs for photorealistic and identity preserving profile face synthesis, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, Curran Associates, Inc., 2017, pp. 66–76.

[409] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (2003) 399–458.

[410] T. Zheng, W. Deng, Cross-pose LFW: a database for studying cross-pose face recognition in unconstrained environments, Technical Report 18-01, Beijing University of Posts and Telecommunications, 2018, http://www.whdeng.cn/CPLFW/index.html?reload=true.

[411] Y. Zheng, D.K. Pal, M. Savvides, Ring loss: convex feature normalization for face recognition, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5089–5097.

[412] H. Zhi, S. Liu, Face recognition based on genetic algorithm, J. Vis. Commun. Image Represent. 58 (2019) 495–502.

[413] C. Zhong, Z. Sun, T. Tan, Robust 3d face recognition using learned visual codebook, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–6, http://www.cbsr.ia.ac.cn/english/3DFace%20Databases.asp.

[414] H. Zhou, A. Mian, L. Wei, D. Creighton, M. Hossny, S. Nahavandi, Recent advances on singlemodal and multimodal face recognition: a survey, IEEE Trans. Human-Mach. Syst. 44 (2014) 701–716.

[415] F. Zhu, L. Shao, Weakly-supervised cross-domain dictionary learning for visual recognition, Int. J. Comput. Vis. 109 (2014) 42–59.

[416] M. Zhu, J. Li, N. Wang, X. Gao, A deep collaborative framework for face photo–sketch synthesis, IEEE Trans. Neural Netw. Learn. Syst. 30 (2019) 3096–3108.

[417] Z. Zhu, P. Luo, X. Wang, X. Tang, Deep learning identity-preserving face space, in: 2013 IEEE International Conference on Computer Vision, 2013, pp. 113–120.

[418] X. Zou, J. Kittler, K. Messer, Illumination invariant face recognition: a survey, in: IEEE International Conference on Biometrics: Theory, Applications, and Systems, 2007.

Murat Taskiran received B.Sc. (2013) and M.Sc. (2016) degrees in Electronics and Communication Engineering, from Yildiz Technical University (YTU), Istanbul, Turkey. He is currently a Ph.D. student at YTU. Since 2014, he has been working as a research assistant in Department of Electronics and Communication Engineering in YTU. His research interests are in image processing, neural networks and randomness analysis.

Nihan Kahraman received B.Sc. (2001), M.Sc. (2003) and Ph.D. (2008) degrees in Electronics and Communication Engineering, from Yildiz Technical University (YTU), Istanbul, Turkey. Currently, she is working as an assistant professor at YTU. Her research interests are includes VLSI design, hardware and software implementations of neural networks and neural network architectures.

Cigdem Eroglu Erdem received the B.S. and M.Sc. degrees in Electrical and Electronics Engineering from Bilkent University, Ankara, Turkey in 1995 and 1997, respectively, with high honors. She received the Ph.D. degree in Electrical and Electronics Engineering from Bogazici University, Istanbul, Turkey, in 2002. From September 2000 to June 2001, she was a visiting researcher in the Department of Electrical and Computer Engineering, University of Rochester, NY, USA. Between 2003-2004, she was a postdoctoral fellow at the Faculty of Electrical Engineering at Delft University of Technology, the Netherlands, where she was also affiliated with the video processing group at Philips Research Laboratories, Eindhoven. Between 2002-2009, she was the director of research at Momentum Digital Media Technologies Inc., a technology SME located in İstanbul. Between 2009-2016, she was a faculty member in the Department of Electrical and Electronics Engineering at Bahcesehir University, Istanbul, Turkey. Since Sep. 2016, she is a faculty member in the Department of Computer Engineering at Marmara University, Istanbul, Turkey.

Dr. Erdem's current research interests are in the areas of digital image and video processing, computer vision and pattern recognition, with applications to affective computing, motion estimation, video segmentation, object tracking, and human computer interaction. She served as a referee for numerous technical journals and conferences including IEEE Transactions on Image Processing, IEEE Transactions on Circuits and Systems for Video Technology, Pattern Recognition, Signal Processing: Image Communication, Image and Vision Computing. She also serves as an independent expert and vice chair during project evaluations for the European Commission.