Recent Advances on Singlemodal and Multimodal Face Recognition: A Survey

Hailing Zhou, Ajmal Mian, Lei Wei, Doug Creighton, Mo Hossny, and Saeid Nahavandi, *Senior Member, IEEE*

***Abstract*—High performance for face recognition systems occurs in controlled environments and degrades with variations in illumination, facial expression, and pose. Efforts have been made to explore alternate face modalities such as infrared (IR) and 3-D for face recognition. Studies also demonstrate that fusion of multiple face modalities improve performance as compared with singlemodal face recognition. This paper categorizes these algorithms into singlemodal and multimodal face recognition and evaluates methods within each category via detailed descriptions of representative work and summarizations in tables. Advantages and disadvantages of each modality for face recognition are analyzed. In addition, face databases and system evaluations are also covered.**

***Index Terms*—Face recognition, infrared, multimodal, singlemodal, survey, visual, 3-D.**

I. Introduction

FACE recognition aims to identify or verify a person's identity by matching input face biometrics against known faces in a database. The face set, with known identities, is referred to as the *gallery* and the input face requiring identity determination is the *probe*. One problem in face recognition is *verification* (or *authentication*), and the other is *identification* (or *recognition*). Verification is the process of determining a person's claimed identity. A query face is compared with only the template of the claimed identity. It is used in access control situations, where users are typically cooperative. Face identification is more challenging as it crosschecks a complete gallery for verification matching, and user collaboration is not assumed.

Visual face recognition systems have achieved high performance under controlled environments, but they are still challenged by background clutter, image quality, illumination, pose, expression, time delay, disguise, and occlusion. Work on alternative imaging modalities attracts much interest recently [1]. Three-dimensional face recognition may outperform visual face recognition [2] with greater recognition accuracy and robustness with respect to pose and illumination variations [3], [4], as a face is a 3-D object. Unlike visual face recognition, which involves 3-D to 2-D projections of 2-D camera systems and causes data loss, 3-D face recognition directly works on 3-D face models. The 3-D model contains complete facial geometry information and is relatively invariant to lighting conditions. Three-dimensional face recognition has the following issues [5], [6].

1) *Capture devices:* Relative to 2-D cameras, existing 3-D sensor technologies are still immature for practical applications. Existing problems include noise and artifacts, small depth of field, long acquisition time, and issues related to eye safety, points sampling, and accuracy.
2) *Expression variations:* A face is not a rigid surface and can change with different expressions.
3) *Disguise and time delay:* Faces can be disguised and element size can change over time.

Three-dimensional capture devices are not completely independent of lighting. Laser or structured light scanners can still be affected by strong light sources [3].

Infrared (IR) face recognition is promising [7], as it is nearly invariant to illumination changes including the total darkness [8] and may identify individuals with time delay, makeup, or simple disguises. The IR spectrum consists of the reflected and thermal IR bands. Most commonly used IR bands are near-IR (NIR) and thermal-IR. NIR images record solar NIR radiation reflected from objects. NIR cameras can view unlit areas. Thermal-IR images record thermal radiation emitted from objects. The human face emits thermal radiation through superficial blood vessels. Both sources of IR images are relatively independent of lighting conditions. As IR images are acquired by 2-D imaging sensors, IR face recognition also suffers from sensibility to pose or viewpoint variations. It introduces challenges.

1) *Capture devices:* IR sensors have lower image resolution and higher image noise than visual cameras [9].
2) *Ambient or body temperature:* Cold or hot air and physical activities can change face temperature, thus influencing capture results of a face image.
3) *Glasses:* IR sensors are opaque to glass, which results in a loss of information near the eyes.

Face recognition improvements are expected with the increase of available information. Studies [10]–[14] investigated fusion techniques using visual, IR, and 3-D data. Under the same conditions, the performance using multiple modalities for face recognition is frequently better than using a single one. How and which of these modalities should be combined for optimal results is still unknown. If face modalities are mixed inappropriately, the unreliability of face recognition can be exacerbated. For example, when faces are not frontally oriented, the use of

Manuscript received February 18, 2014; revised July 8, 2014; accepted July 11, 2014. Date of publication August 12, 2014; date of current version November 12, 2014. This paper was recommended by Associate Editor H. Liu.

H. Zhou, L. Wei, D. Creighton, M. Hossny, and S. Nahavandi are with the Centre for Intelligent Systems Research, Deakin University, Geelong, Vic. 00113B, Australia (e-mail: hailing.zhou@hotmail.com; lei.wei@deakin.edu.au; douglas.creighton@deakin.edu.au; mohammed.hossny@deakin.edu.au; saeid.nahavandi@deakin.edu.au).

A. Mian is with the Department of Computer Science and Software Engineering, The University of Western Australia, Crawley, W.A. 6009, Australia (e-mail: ajmal.mian@uwa.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2014.2340578

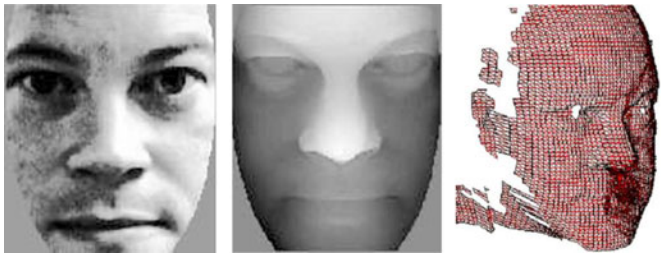

Fig. 1. Face data from [2]. (a) Two-dimensional intensity image, (b) a 2.5-D range image, and (c) a 3-D mesh. (Courtesy of [2]).

facial texture information may have an adverse effect on 3-D face recognition because the pose causes illumination changes.

This paper categorizes and characterizes the singlemodal and multimodal techniques and presents advantages and disadvantages of each modality for face recognition. Singlemodal face recognition is addressed in Section II and multimodal in Section III. Section IV discusses available benchmarking face databases and summarizes performance of the approaches reviewed. A discussion is given in Section V.

II. Singlemodal Face Recognition

A. Visual Face Recognition

As there exist many surveys related to visual face recognition [1], [3], [7], [15]–[17], in this paper, we give a brief discussion on visual face recognition. High performance is achieved in visual face recognition [18]–[21] with recognition rates over 90%. Challenges include illumination, pose, and expression variations. To alleviate the illumination problem, multiple samples of a face in different lighting conditions are used in the training stage. Local features are developed to mitigate the challenge of expressions. Pose is the major issue in visual face recognition as its effects are nonlinear and global.

B. Three-Dimensional Face Recognition

Three-dimensional face recognition works on 3-D facial data. The main advantage of 3-D facial data is the preservation of face shape (geometry) information. Many facial variations of extrapersonal and intrapersonal can be captured.

To acquire 3-D face data, one needs a stereo camera system, a structured light sensor, or a laser range scanner. The output is either range images or 3-D polygonal meshes (or clouds). A *range* image, which is also referred to as a *depth* map, a 2.5-D image, or a *scan* image, records the depth value z for each pixel in the XY plane. It provides a single viewpoint of the face. The major difference between 2.5-D and 3-D data is that the former contains at most one depth value for every point in the XY plane [22]. Multiple 2.5-D scans for full views of a face can produce one 3-D facial model. Fig. 1 illustrates 2-D, 2.5-D, and 3-D face representations.

Three-dimensional face recognition techniques can be classified into feature-based and holistic methods (see Table I). Some techniques use 3-D information for face recognition but are actually based on visual images. These approaches do not involve the sensing/acquisition of real 3-D face data. Blanz and Vetter [23] introduce a 3-D morphable face model to fit each face image and, then, to compute shape and texture coefficients. These coefficients capture intrinsic structure information of the face and, thus, are independent of extrinsic parameters such as poses or illumination variations. Face recognition is performed by comparing these coefficients. Experiments on the CMU-PIE database [24] and the FERET database [25] report 95.0% and 95.9% identification rates, respectively. Blanz *et al.* [26] generate a frontal facial image from a nonfront view after model fitting, which improves the performance of recognizing faces with large variations in viewpoint. Instead of using a generic 3-D model, Lu *et al.* [27] and Hu *et al.* [28] extract a specific face model for each face. A 3-D face model is constructed from a single frontal face image and used to synthesize various face images with different poses, illuminations, and expressions. Face subspaces are characterized with these synthesis images and compared for face recognition. A recognition rate of 85% is reported in [27] on a dataset of ten subjects, and 95% in [28] on a database of 68 subjects. The underlying problems of these methods are the precision of the 3-D reconstruction from a single image [29] and the precision of a synthesized face Image [3].

1) Feature-Based Methods: Feature-based approaches extract features to represent a face and then identify a face by comparing these features. Two problems with these approaches are face representation and face comparison in terms of a face representation.

In [60], Chua *et al.* propose point signatures as a representation of a 3-D surface instead of 3-D coordinates. Point structure describes complete structural neighborhood of points and is invariant to rotation and translation. The method is limited to recognize 3-D rigid objects. Chua *et al.* [30] extends the study by using point signatures for nonrigid objects such as faces. A 3-D face model is observed consisting of rigid and nonrigid parts that correspond to areas with less and more deformation among different expressions. The rigid parts are identified as the regions with low match errors by comparing point signatures among face models from the same person with different expressions. An indexed library of face models is built by collecting these rigid parts. Face recognition searches the library, selects and then verifies the most likely model. A recognition rate of 100% is achieved in a small dataset including six subjects with four expressions, but the computational cost of extracting point signatures is high. Mpiperis *et al.* [31] introduce a selection procedure of suitable points, where a small subset of points are used for calculating signatures.

Normal vectors of face surfaces contain more discriminatory information than coordinate vectors of 3-D points [32], [33]. A normal map of a 3-D surface is extracted by projection for face representation in [32]. A difference map between two normal maps is calculated for face recognition. A recognition rate of 98.9% is achieved on a database including 31 subjects with ten expressions.

TABLE I
DETAILED TAXONOMY FOR 3-D FACE RECOGNITION TECHNIQUES

Approaches	Representative work
Feature-based approach	
-Point signature	Point signatures for rigid and nonrigid facial parts [30] and simplified point signature [31]
-Normal vector	Normal map with a difference map [32] and ICNP [33]
-Curvature	Curvature-based facial regions [34] and Gaussian curvatures with a Hausdorff distance [35]
-Tensor	Third-order tensor [36] and the rank zero tensors [37]
-Facial curve	Level curves with a NN classifier [38] and isocontours [31], radial curves under a Riemannian framework [39], and isogeodesic stripes with graph matching [40]
-Subregions of face	Fusion of face patches around nose [41], [42] and regions of nose and eyes-forehead [43]
-Summation invariant	Affine and Euclidean invariant features for object recognition [44], [45]
Holistic approach	
-PCA	Principal components and optimal components of facial scan images [46]–[49] and 2-D-PCA [50]
-ICP	Coarse to fine alignments [51] and the ICNP algorithm [33]
-Deformable face model	A generic and person-specific deformable models [5], [52]–[54]
-Face expression-invariant model	Isometric face surface [55], [56] and an expression model [57]
-Classification	Classifiers of PSDMs [6], SSDM [58], and spherical harmonic features [59]

Gordon [61] shows that curvature information is suitable to describe faces, which is invariant to viewpoints. Regions of the nose, eyes, and the surrounding areas are extracted and defined based on their curvature values. A vector formed by these curvature values is used to represent and determine a face. A recognition rate is reported as high as 100% on a database of eight subjects with three different views for each. Moreno *et al.* [34] use curvature information to segment a set of 86 regions of a face and create a curvature feature vector based on these regions. Experiments on 420 3-D-facial meshes of 60 subjects report a recognition rate of 78%. In [35], principal and Gaussian curvatures are used to describe faces with a depth-weighted Hausdorff distance to compare them. The motivation of using the Hausdorff distance is its partial invariance to incomplete data such as holes in 3-D data. A best recognition rate of 98% is stated for a database of 42 subjects.

Curvature-based features are sensitive to noise. Tensors extracted from a large neighborhood are more robust to noise than curvatures. Moreover, the tensor can provide a facial representation invariant to rigid transformations. Mian *et al.* [36] propose to use rank-3 tensors, which are calculated locally in defined 3-D coordinate bases. They use a 4-D hash table to index the tensors of gallery faces and guide the search for matching. A recognition rate of 86.4% is achieved on a database with 935 range images of 277 subjects. The algorithm is pose invariant without manual guidance for face registration and any preprocessions to remove noise or fill holes of the 3-D face models. Al-Osaimi *et al.* [37] use rank-0 tensors as discriminating features from 2.5-D pointclouds of a face, where both local and global rank-0 tensors at every point are calculated and, then, integrated in a 2-D histogram. Principal component analysis (PCA) [62] is applied to each histogram. A single facial feature vector is formed by concatenating PCA coefficients and used for face recognition. Experiments on the FRGC v2.0 [63] using 2410 records from 466 subjects yield 93.78% identification and 95.37% verification rate at 0.1% FAR, while using the fusion of local and global geometrical cues.

In [38], novel features are proposed and are derived from geometric shape information of face surfaces. The features are level curves (called facial curves) of surface depth values (see Fig. 2). They are independent of planar motion and global scale but varying with nonplanar rotations. They preprocess face models by applying surface alignment over all directions. Face recognition is, then, performed by comparing the corresponding facial curves and classifying them using a nearest neighbor classifier. Experiments with 300 facial surfaces of 50 persons with six facial expressions from the Florida State University (FSU) database and 720 scans belonging to 162 subjects from the University of Notre Dame (UND) database achieved recognition rates of 92% and 90.4%, respectively. Mpiperis *et al.* [31] extract level contours/isocontours of facial surface for representation and adopts the contour matching algorithms for recognition. They compare the performance of isocontours with point signatures, indicating that isocontour features give a better recognition rate. However, both point signature and isocontour features are not robust enough for facial expressions. To solve this problem, isogeodesic strips with spatial displacement between points of stripes are proposed to exhibit variations caused by facial expressions [40]. Drira *et al.* [39] observe that the changes in facial expressions affect local regions of a facial surface differently, and facial curves chosen appropriately have

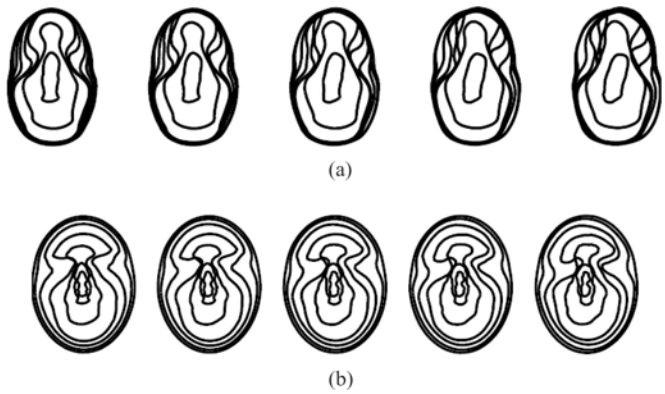

Fig. 2. Facial curves in [38]. (a) Same person with six expressions, and (b) the same expression on six different individuals. (Courtesy of [38]).

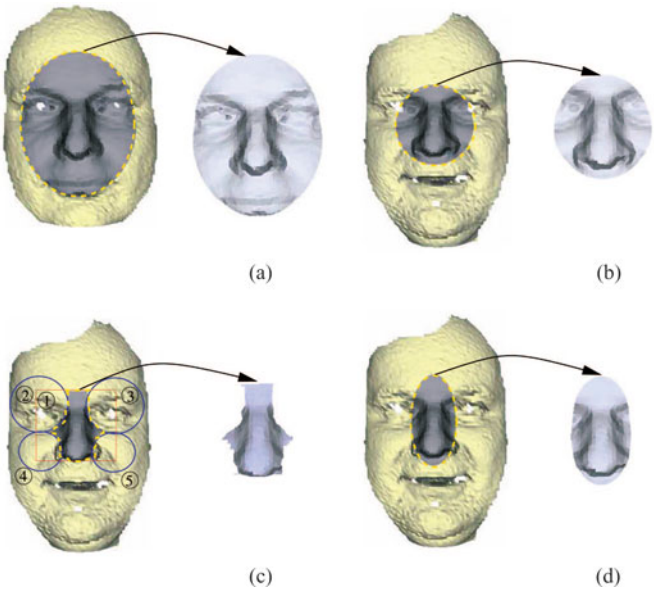

Fig. 3. Surface patches in [41]. (a) Gallery patch; (b), (c), and (d) circular, rectangular, and elliptical probe region around the nose tip, respectively. (Courtesy of [41]).

potential to capture the difference of region shapes. They propose radial curves emanating from tip of the nose for face representation and develop an elastic Riemannian metric for face matching. Their method can measure facial deformations of expressions and complete partially obscured curves due to occlusions, which is robust to face expressions and occlusions. High performance is achieved on three databases: FRGC v2.0, GavaDB, and Bosphorus with recognition rates of 97.7%, 96.99%, and 89.25%, respectively.

Regions around the nose area were observed to have low shape variation with expressions, which have the potential to be the discriminating facial features for accurate and robust face recognition. Chang *et al.* [41] use multiple overlapping surface patches as shown in Fig. 3 [i.e., (b), (c), and (d)] from the nose area for facial recognition, where these patches are detected using a curvature-based segmentation technique. PCA and iterative closest point (ICP, proposed in [64]) based recognition engines are applied in these regions. The maximum mean recognition rate of 87.1% is achieved on a database consisting of 4000 scans of 449 subjects with expression variation. Faltemier *et al.* [42] observe that 28 different subregions of a face are invariant to expression. They apply ICP as the recognition engine to each of the regions and use the Borda count technique as the score-based fusion scheme to produce the final recognition result. Experiments on FRGC v2.0 report 97.2% recognition rate at rank 1 and 93.2% verification rate at a 0.1% FAR. Lei *et al.* [43] take the regions of nose and eyes-forehead into account. Local geometrical features like angles, distances, and radii are extracted from the regions and are fused to be one feature vector as the input of a support vector machine (SVM). Experiments on FRGC v2.0 show a verification rate of 97.6% at 0.1% FAR and an identification rate of 95.6%.

Lin *et al.* [44] propose a novel invariant of affine transformation, which is a weighted summation of discrete data. Since the summation invariants do not change with affine transformation, they can serve as representative features of objects. In [45], summation invariants are applied in 3-D face recognition, where they develop invariants for Euclidean groups over both curves and surface patches. Geometrically invariant features are extracted, and then, PCA is performed to reduce the dimension. The similarity metrics produced by these features are fused by the sum rule to determine the final recognition results. A 97.2% verification rate is achieved on FRGC v1.0. Their results support the observation that features surrounding the nose region can result in better performance than other facial regions.

2) Holistic Approaches: Holistic approaches use global representation to describe and recognize a face. The descriptions in holistic approaches are based on entire range images or 3-D models rather than high-level features of the data.

A typical holistic approach is PCA extended for 3-D face recognition. Hesher *et al.* [47] apply PCA in range images. The images are reduced to 10-D vectors. A nearest neighbor algorithm using the Euclidian distance for measurement is performed on the vectors to identify a face. Experiments on a dataset of 37 subjects with six expressions for each yield a recognition rate of 100%. In [46], they test different surface representations such as depth and curvature representations and distance metrics such as Euclidean and cosine measurements, achieving a recognition rate of 87.3% on a dataset of 330 3-D face models from 100 subjects. However, face size greatly affects the performance of PCA-based face recognition. Russ *et al.* [48] propose a technique to preserve the size information by introducing a generic 3-D reference face model that is scaled to fit a probe face model. PCA-based approaches need to transform an image into 1-D vectors, where it is difficult to evaluate the covariance matrix accurately. 2-D-PCA for 3-D face recognition overcomes the problem by working on 2-D matrices rather than 1-D vectors [50], [65].

In [49], instead of using PCA, the optimal linear representation of faces is searched by choosing a projection that maximizes the face recognition performance on the training image set. The performance is evaluated on a dataset of 67 people with six facial expressions achieving 99%. Since the optimization is related to the training set, the size and variety of training set affect the performance directly, as well as the computational complexity. In [66], the authors discuss converting a 3-D face model into a 2-D domain; therefore, methods on 2-D images can be applied to 3-D models. A flatten technique is introduced for the conversion, and the intrinsic geometric properties of a face are preserved with the technique. An experiment using the eigenface method on the constructed depth images for 3-D face recognition yields a recognition rate of 95% on the FRGC v2.0 database.

However, 3-D face recognition by converting 3-D surfaces into 2-D domains has some common problems of 2-D face recognition such as degrading performance with expression and viewpoint variations. In [51], fullview 3-D face models are built in a gallery, each of which is constructed based on multiview 2.5-D facial images (i.e., scan or range images). With testing 2.5-D face model such as a profile or frontal scan image, they coarsely transform it to align with a 3-D model in the database. Then, a fine alignment transformation is performed on the

testing model using the ICP registration process to find a matching 3-D model with the minimum distance between them. A recognition rate of 96.5% is achieved on a database of 18 3-D face models with 113 2.5-D face scans. In [67], a watershed segmentation technique is applied to partition a face model into regions, and the regions are assigned with weights for distinguishing regions significance while calculating the global recognition score. Russ *et al.* [48] use ICP to obtain standardized facial models with pose and viewpoint variations before performing PCA for recognition. A 97% identification rate is achieved on FRGC v1.0 and v2.0 with neutral expressions, and a 82.6% identification rate with nonneutral expressions. A modified ICP algorithm is proposed in [33], which is called iterative closest normal point (ICNP), where the alignment is based on normal vectors of models instead of their points. By using ICNP, closest normal points between each face and a generic reference face are found as input to a discriminant analysis method [68] for face recognition. They report a 99.6% verification rate at a FAR 0.1% on the FRGC v2.0 database.

The shortcoming of ICP is that it only works well for rigid transformations. As human faces contain nonrigid transformation, methods for handling nonrigid transformations have been developed. An annotated face model (AFM) [5], [52], [53] is a deformed facial model with annotated facial areas. Given a 3-D face model, it is parameterized using the AFM to capture geometric characteristics, where alignment and deformation operations are required. Kakadiaris *et al.* [5] transform the characteristics to a wavelet domain, where a small portion of the wavelet data is required for face matching. Efficiency and accuracy are demonstrated as well as robustness to pose and expression changes. A recognition rate of 97.3% is achieved on the FRGC v2.0 database. However, a generic face model may suppress characteristic differences between faces. To avoid it, the person-specific deformable model is developed for each subject in the gallery [54]. With the person-specific deformable model, the deformation on a test face can be synthesized. ICP is used to match a deformable model with a test scan image for face recognition. Experiments on a dataset of ten subjects with three poses and seven expressions yield 92.1% accuracy. In [57], an expression deformation model is built by calculating shape residues between nonneutral and neutral pairs and projecting them into a PCA subspace. Their face recognition is performed on models morphing out the expression deformation.

Bronstein *et al.* [55] develop an expression-invariant representation for a 3-D face. They observe that the deformations of a face resulting from expressions can be modeled as isometrics of the facial surface. An expression-invariant representation of a face is constructed using an isometric embedding method, which is an isometry-invariant representation of the face surface. Face recognition is a comparison of the similarities between isometric surfaces. Experiments on a dataset of 220 faces from 30 subjects with different facial expressions report a recognition rate of 100%. However, all expressions are limited to be a closed mouth. An extension [56] considers expressions with both open and closed mouth. An isometric embedding approach is employed to construct the resulting representation of canonical form, where spherical embedding is used instead of an Euclidean one leading to smaller metric distortions and better recognition rates. Al-Osaimi *et al.* [57] observe that differentiating deformations caused by facial expression from deformations caused by different subjects improves the accuracy of 3-D face recognition. They morph out the expression deformation from each 3-D face by introducing an expression deformation model. Face recognition is to measure similarities of the morphed faces. 98.35% and 97.73% verification rates at 0.1% FAR for neutral and nonneutral faces are reported on FRGC v2.0.

Converting 3-D face recognition into a classification problem is developed. In [6], pure shape difference maps (PSDMs) depict the shape difference/similarity between faces, and weak classifiers based on the PSDMs are constructed to assemble a strong collective classifier. The strong classifier is used to match a probe face in a gallery. Experiments on the FRGC v2.0 and BU-3DFE databases report a recognition rate of 98%. An extension [58] builds a signed shape difference map (SSDM) between two aligned 3-D faces using haar-like features, Gabor features, and local binary pattern (LBP) to capture shape characteristics on different positions of face. They boost and train these features to select the most discriminant features, which assembles strong classifiers. The verification rate of 97.9% at 0.1% and the rank-1 recognition rate of 98% are achieved on FRGC v2.0. Liu *et al.* [59] propose to use spherical harmonic features to capture both global and local facial information. Feature selection is followed to select discriminative features as inputs to a classifier. Experiments on FRGC 2.0 yield a recognition of 96.94% with facial expressions, poses, and occlusions.

3) Discussion: We have reviewed and classified 3-D face recognition approaches into two categories: feature-based and holistic algorithms. Some methods overlap the classification boundary by using both features and holistic techniques to recognize a face. For example, the combination of features and PCA, ICP, or classification techniques are in [6], [33], [37], [41]–[43], [45], [58], and [59].

Pose, viewpoint, and expression variations are challenges of face recognition. To address these problems, feature-based approaches extract features of faces invariant to poses, viewpoints, and expressions. Point signature is invariant to rotations and translations. Normal vector and curvature features are invariant to viewpoints. They are defined locally and are sensitive to noises. Tensor features combine local and global geometrical cues. Invariance to rigid transformations is not enough, as facial expressions are not-rigid transformations. Facial curves and region-based methods are developed to capture variations caused by expressions. Holistic approaches such as PCA and classification statistically learn face patterns for recognition, which rely on the variety of training sets. ICP compares pairs of faces via rigid transformations, while variations of model sizes and expressions often reduce performance. Deformed face models identify faces through alignment and deformation transformations. Isometric embedding method and an expression deformation model are proposed to address varying expression problems. While great performances is achieved, the method suffers from intensive computation.

TABLE II
DETAILED TAXONOMY FOR INFRARED FACE RECOGNITION TECHNIQUES

Approaches	Representative work
Statistical	
-Linear representation for IR faces	PCA, LDA, and ICA [69], [70]
-Classification	SVM, RBF, and Bayesian classifiers [71]–[74]
-Spectral components model	Bessel forms and the Bayesian classification [75], [76]
Feature based	
-Thermal features	Vascular features and faceprint [77]–[79]
-NIR features	LBP [80]–[82] and spectral reflectance vectors [83]

C. Infrared Face Recognition

IR-based face recognition is a promising alternative due to its nearly invariance to makeup and illumination changes even in total darkness. In this section, we review recent techniques of face recognition based on IR images, classified into two categories: statistical approaches and feature-based methods (see Table II).

1) Statistical Approaches: Most visual face recognition techniques are also applicable to IR face recognition, as they both work on 2-D facial data. The most widely used techniques include PCA, linear discriminant analysis (LDA [84]) and independent component analysis (ICA [85]). These techniques serve as standard face recognition engines for thermal IR images [9], [69], [86], [87]. Experiments on the Equniox database indicate that the performance on long-wave infrared (LWIR) images outperform one on visual images by more than 30%. Kang *et al.* [70] apply both PCA and LDA in face recognition on short-wave infrared (SWIR) images for improving performance.

Classifiers are widely used in IR face recognition. Trujillo *et al.* [71] use PCA to represent face regions including left eye, right eye, and mouth areas as eigenimages. Surprise, happy, and angry expressions can be recognized, where an SVM classifier for each of eigenimage is trained. Good recognition performance is claimed on a dataset including 30 subjects with three expressions of three poses obtained from the IRIS dataset. Zou *et al.* [72] use LDA to build a subspace for face recognition on NIR images, where four classification techniques, i.e., radial basis function (RBF) neural network, Adaboost, SVM, and nearest neighbour (NN), are used. Results show greater performance using NIR images than visual images. Akhloufi and Bendada [73] propose a probabilistic Bayesian framework for face recognition, which combines a Bayesian maximum likelihood with PCA, kernel-PCA, kernel-LDA, local linear embedding, or locality preserving projection. In [74], the authors transform thermal IR images to blood perfusion data that is less sensitive to ambient temperature. PCA and RBF are used for face recognition. A highest recognition rate of 90.1% is achieved on a database of 410 images from 41 subjects. Improvements of using blood perfusion data are demonstrated over using thermal IR images.

Srivastava *et al.* [75] propose a statistical model for IR images. They apply Gabor filters to decompose an IR image into spectral components and introduce an analytical probability model (Bessel forms) to model each component. Faces are identified by comparing their Bessel forms. Experiments on a dataset of 180 frontal IR face images of nine subjects with or without glasses under changes of expressions and poses yield a recognition rate of 97.14%. Buddharaju *et al.* [76] improve the work through automatically segmenting regions of interest from the background, using Srivastava's method to get a short list of probable matches, and applying the Bayesian classification to find an exact match. Compared with [75], the method can output a unique solution for face recognition rather than a short list of candidates. Experiments on the Equinox database demonstrate the improvements on hypothesis pruning and classification and report a highest recognition rate of 89.6%.

2) Feature-Based Methods: Face temperature is transported by underlying blood vessels of the skin. Based on thermal IR images capturing the temperature differences under facial skin, the pattern of superficial blood vessels [88] can be derived. The pattern of blood vessels exhibits the anatomical structures of faces. As each person has a unique anatomical structure of the face, this information can be used for identification.

In [77], the authors describe vascular features, their uniqueness, and the time-invariant property. They build a database of faces by collecting vascular features from each subject. Given a query face image, they extract the features and match them against the database to establish the identity of an individual. Different poses of a face are also considered by recording vascular features of each pose in the database building step. Experiments on a multiposes database of 7590 images belonging to 138 subjects with varying poses and expressions yield a recognition rate of 86%. In order to compare with the PCA algorithm, experiments are conducted on a UND database that consists of 2294 images from 63 subjects during different sessions with specific lighting and expressions. A recognition rate near 80% is achieved as compared with PCA with 75%. In [78], they propose a faceprint as the unique feature for each individual. The faceprint consists of a medial axis of isotherm regions. A 100% recognition rate is stated for the small database. To reduce the variations of vasculature over time, Guzman *et al.* in [79] and [89] propose using four images at various instants to extract a combined signature pattern for describing a face. Experiments on 13 subjects under indoor conditions show an average accuracy of 90.39%.

However, thermal IR images have a drawback of being sensitive to ambient temperature. Nevertheless, NIR images are less dependent on skin temperature compared with thermal images. Li *et al.* [80]–[82] investigate face recognition using NIR images, where they design an NIR imaging system to produce face images with high quality regardless of lighting conditions. LBP features are introduced as face descriptors. These features are weighted by using the Adaboost algorithm to build a classifier for face recognition. Robustness on face recognition is evaluated with respect to illumination, glasses, time lapse and, ethnic groups. Experiments on 10 000 face images of 1000 subjects yield a verification rate of 91.8% at FAR $= 0.1\%$. Pan *et al.* [83] propose spectral reflectance vectors to represent a face on

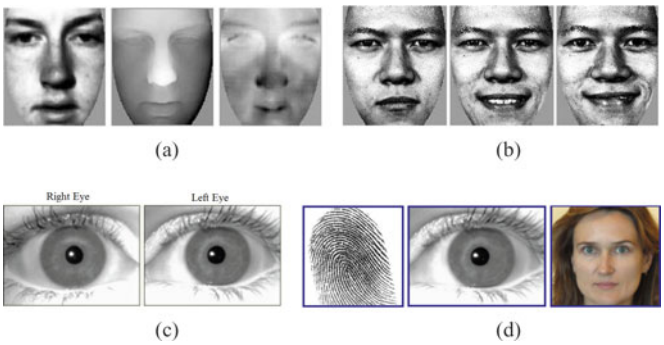

Fig. 4. Multiple meanings of "multimodal" [14], [90]. (a) Multi-sensors. (b) Multi-samples. (c) Multi-units. (d) Multi-biometrics.

TABLE III
DETAILED TAXONOMY FOR MULTIMODAL FACE RECOGNITION TECHNIQUES

Approaches	Representative work
visual+3D	
-Decision fusion	PCA [11], [12], [91], [92], stepwise-LDA [93], and [94]
-Data fusion	Pixel level fusion [95]–[97] and feature level fusion [98]–[100]
-Hybrid	Pattern classification [101]
visual+IR	
-Decision fusion	PCA [102], [103] and fuzzy integral fusion [104]
-Data fusion	Data fusions in spatial, wavelet, and eigenspace domains [105]–[109]
-Hybrid method	Glass removal and image modularization [110], [111]
visual+IR+3-D	
-Decision fusion	Performance studies using PCA [13], [14], [112]

NIR images. Experiments demonstrate the effectiveness on a database of 200 subjects with different poses and expressions with a highest identification rate over 90%.

3) Discussion: IR face recognition techniques have been classified into statistical and feature-based approaches. There are some methods that combine both statistical and feature-based techniques [75], [82].

IR face recognition is almost invariant to varying illuminations. Moreover, for altering facial appearances using makeup, artificial materials, or plastic surgery, IR spectrum is able to detect and then recognize them due to thermal signatures of faces. The most challenging cases for IR face recognition are glasses, ambient temperatures, as well as body physical conditions. NIR images are more robust to these factors, but lose the advantages of thermal features.

III. MULTIMODAL FACE RECOGNITION

There are multiple meanings of "multimodal" (see Fig. 4). *Multisensors:* Multiple sensors such as CCD cameras, IR sensors, or 3-D scanners can be used to collect data from the same biometric source. Face recognition using visual and IR imageries belong to this category. *Multisamples*: Multiple samples (such as different facial expressions or illuminations) of the same biometric are sensed in both gallery and probe datasets and fused to determine human identity. *Multialgorithms*: Different algorithms (such as PCA + LDA) are performed on the same sensing image with a single sample. *Multiunits*: Different but similar biometric sources (such as ten fingers or two irises) are collected, which are multiple instances of similar body parts. *Multibiometrics*: Multiple biometric sources on a person (such as gate, speech, palm-dorsa, fingerprint, iris, and face) are collected with different sensors.

Here, we review multisensor techniques for recognizing faces (referred to as multiple face modalities), in particular, visual+3-D, visual+IR, and visual+IR+3D fusion techniques (see Table III). As visual images contain texture information capturing many face features, fusion with texture data is the most popular way to improve 3-D and IR face recognition. Fusion techniques include data fusion and decision fusion. Data fusion techniques often combine different facial modalities to form an integrated representation as the input to a face recognition algorithm. Decision fusion techniques separately perform face recognition on individual facial modalities and sum their performance scores as the final output.

A. Visual+3-D

Studies [11], [12], [91] show that face recognition based on the combination of color and depth data outperforms those approaches using a single modality. Tsalakanidou *et al.* [91] evaluates face recognition using color, depth, and the combination of color and depth based on PCA. Depth-based face recognition appears to be robust to noise and poses, and the combination of color and depth leads to a high recognition rate. In [12], both a single probe and multiple probes are tested on a dataset of 275 subjects captured with different facial expressions under different lighting conditions. See the results in Fig. 5, where the vertical coordinate SCORE indicates the recognition rate. For a single probe, a recognition rate is 89% for visual images, 94.5% for 3-D images, and 98.5% for the combination. For multiple probes, using visual images can achieve a score of 89.5%, using 3-D images 93%, and the combination 99%. PCA is used for multimodal face recognition [11] on a single-image sample such as a visual image or a 3-D facial model, multisample, and multisensor of texture or/and depth images. Multimodal biometrics outperform a single modality, and the improvement of face recognition can start from the combination of 1) texture and depth images as well as 2) two or more texture (or depth) images taken under varied illumination or expression.

In [92], a face region is offline segmented into eyes-forehead, nose, and cheeks, where the cheeks are discarded to avoid facial expression variations in the recognition phase. A modified ICP algorithm is applied to the nose and forehead regions for feature-based matching, and then the PCA algorithm is performed on visual and depth images for holistic matching. The final recognition result is to fuse individual matching scores. The method achieves 100% verification rate at 0.0006 FAR on the FRGC v1.0 dataset, and shows that the eyes-forehead (upper) region of a face contains much more discriminating features than the nose (lower) region.

However, PCA finds orthogonal components of data without the consideration of class affiliations. The "small sample size" problem often arises when the number of training samples is less than the number of feature dimensions. Jahanbin *et al.* [93]

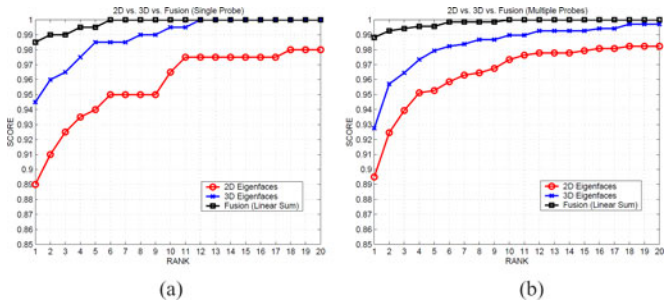

Fig. 5. Comparisons on 2-D (red), 3-D (blue), and fusion (black) face recognition performances in [11] using (a) a single probe and (b) multiple probes. (Courtesy of [11]).

employ stepwise-LDA [113] for face recognition. They apply stepwise-LDA to 2-D and 3-D Gabor features as well as features of geodesic and Euclidean distances, and then fuse their resultant scores with weights. High performance is reported on a dataset with 1149 pairs of range and texture images (Texas 3DFRD).

In most methods of using 3-D data, high recognition performance relies on the quality of 3-D faces. Tsalakanidou *et al.* [94] propose a method based on noisy 3-D data from a low-cost scanner. It starts localizing a face by using statistical modeling of the head-torso points and face bilateral symmetry property, and then applies the embedded hidden Markov model (HMM) technique in both depth and color images for face recognition. Experiments yield a recognition rate of 90.7% on visual images, 80% on 3-D images, and 91.7% on the combination of visual and 3-D facial data.

Al-Osaimi *et al.* [95] propose spatially optimized fusion of 3-D and texture data to model variations of expressions and illuminations discriminating from interpersonal information. In [96], 4-D face data are created for face recognition, which is a facial mesh of vertices with texture and geometry attributes. They develop a 4-D ICP algorithm to align faces for measuring similarities for face recognition. 100% accuracy is achieved at a frontal view on a dataset of 62 subjects with five head postures and three expressions. Lu *et al.* [97] integrate texture and depth information into an augmented scan image and build a facial database with full-view face models that are constructed based on several augmented scans. Surface matching between a probe and a face model in a gallery is performed by coarse-to-fine alignment and the LDA-based matching. Experiments show a recognition rate of 90% on 200 3-D face models with 598 2.5-D test scans with different poses, lightings, and expressions.

Wang *et al.* [98] propose a feature-based fusion method, where features of point signatures and Gabor coefficients are, respectively, extracted from range and gray images. These features are projected into subspaces using PCA and are integrated as an augmented feature. The augmented feature passes through an SVM classifier to accomplish facial identification tasks. In [99], they use Gabor filters to extract features from both intensity and depth images. These features are selected and reduced by using a hierarchical scheme embedded with LDA and AdaBoost learning algorithms. Based on these selected depth and intensity features, a strong classifier is constructed in the AdaBoost learning procedure for face recognition. The method is

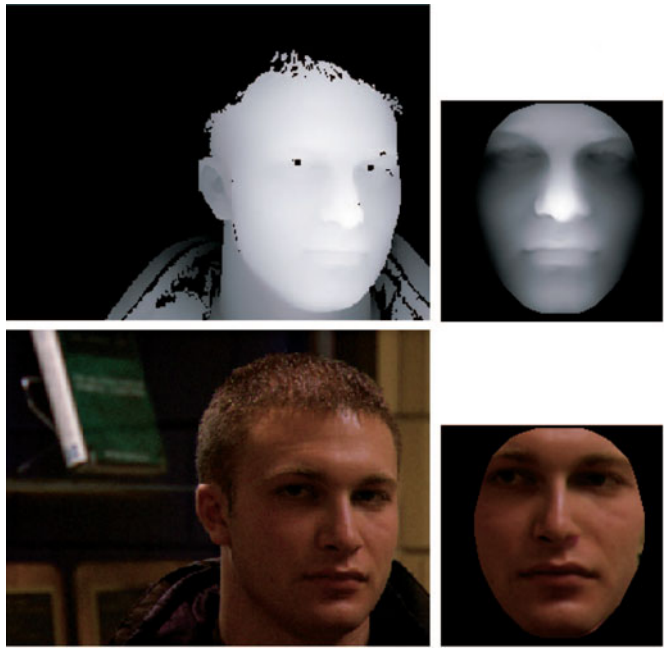

Fig. 6. Pose correction and normalization in [100]. Input images are shown on the left, and the results are shown on the right. (Courtesy of [100]).

automatic and efficient, achieving a verification rate of 95.3% at 0.1% FAR on the FRGC v2.0 database.

In [100], they combine a 3-D spherical face representation with the 2-D scale invariant feature transform descriptor to form a classifier for fast facial rejection in a large gallery. Candidate gallery faces are further verified using an ICP-based matching method. A pose correction technique is proposed to transform all faces (in both visual and depth images) into a standard view (see Fig. 6), which improves robustness to facial poses. As the forehead, the region around eyes and the nose are stable for various expressions; they segment these areas from a face and separately match them using a modified ICP algorithm. The separate matching makes their method robust to expressions. High performance is reported on FRGC v2.0 with an identification rate of 95.37% for probes with a nonneutral expression.

A hybrid approach is proposed in [101], where both data and decision fusions are applied in face recognition. A pattern classifier is developed for three different inputs: texture, depth, and their fusion, where local feature analysis is used to match facial regions: nose, eyes, and mouth. Recognition performance is affected by recognition techniques used. With the FaceIt technology (developed by [114]), 100% accuracy is achieved in a face dataset of 185 subjects, improving the performance of using a single modality. In their method, seven fiducial points of nose tips, eye centers, and mouth corners are required from users for face registration.

B. *Visual+IR*

A study on 385 subjects is conducted in [102] for face recognition using visual and thermal modalities under indoor and outdoor conditions. The conclusions are: 1) outdoor

recognition has worse performance for both modalities than indoor recognition; 2) fusion of both modalities improves face recognition performance under indoor and outdoor conditions and different algorithms; and 3) face recognition with thermal images is more stable under imaging conditions than the visual case. Recognition scores from visual and thermal images sum with equal weights as the final result, as in [115] and [116]. In [103], the physiology-based method of [77] and PCA are separately performed on thermal IR and visual images. Chen *et al.* [104] propose a decision-based fuzzy integral fusion scheme on extracted visual and IR vectors from eigenface components. Experiments on a dataset of 120 training images and 1350 test images from 30 subjects yield a 100% recognition rate. Pop *et al.* [117] use PCA for each modality and fuses the scores for the final recognition result. Different weights of the score from visual images are evaluated, and a maximal recognition rate 97.5% is achieved at the weight of 0.82.

In [105], face recognition is based on fusion of visual and thermal data, where a Gabor filter is used to extract facial features on fused images. The facial features are selected as the points with maximum Gabor responses in a particular window. Experiments on the Equinox database yield a recognition rate of 97.4%. Bhowmik *et al.* [106] study the weights of fusing the two modalities. Four different fusion weights are compared: 0.7 and 0.3, 0.6 and 0.4, 0.5 and 0.5, and 0.4 and 0.6 for visual and thermal data. Experiments on the IRIS database indicate that the optimum fusion weights are 0.4 for visual and 0.6 for thermal with a recognition rate of 93%.

Instead of fusing data in a spatial domain, fusion in the wavelet and eigenspace domains are investigated in [107] and [108]. A genetic algorithm is employed to find the optimal fusion strategy. Experiments are conducted for glasses and facial expression cases. Performance improvements are reported by fusing both modalities as compared with using a single modality. The fusion in the wavelet domain demonstrates better performance in face recognition than the fusion in the eigenspace domain. In [109], the authors compare different data fusion techniques in spatial and wavelet domains for face recognition. An optimized wavelet domain fusion is proposed, which has the best performance with an accuracy of 95.8%.

In [110], data fusion and decision fusion schemes are implemented based on commercial face recognition software, FaceIt, and a method is introduced to detect glasses regions in a thermal image and to replace them with template eye patterns. Experiments conclude: 1) decision fusion with average matching score has a superior recognition accuracy over data fusion, and 2) glasses removal gives great improvements on thermal and data fusion face recognition. In [111], a hybrid technique of using data and decision fusions is developed. They propose phase congruency features, extracted from visual and thermal images. These features are not sensitive to illumination changes. A modular kernel eigenspace for these features is introduced to capture local facial variations caused by expressions, makeup, or accessories. A highest recognition rate of 99.24% can be achieved using decision level fusion, 95.5% using data level fusion, 98.1% using visual images, and 95.5% using thermal images.

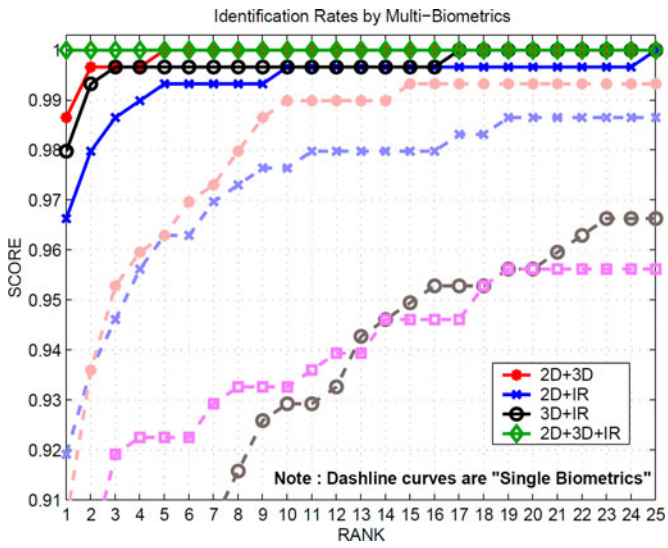

Fig. 7. Singlemodal (dash curves) and multimodal (solid curves) face recognition performances in [112]. Single face modalities include visual data using PCA (red disk), visual data using FaceIT (pink square), 3-D data using PCA (blue fork), and IR data using PCA (brown circle). (Courtesy of [112]).

C. *Visual+IR+3-D*

Studies to examine performances of a single face modality and multiple face modalities on face recognition are presented in [13] and [14]. PCA is used as the face recognition engine. Experiments show that higher accuracy and robustness are achieved by using the combination of multiple face modalities, and the multimodal technique with visual+IR+3D preforms best for the fusions of modalities (see Fig. 7).

In [112], the characteristics of a face (texture, geometry, and physiological information) are utilized for face recognition. The authors construct a deformation image and a vasculature image by fitting AFM to a 3-D face model and a 2-D thermal face map. They use these images for face matching. Recognition performances on different databases and different combinations of modalities are investigated. The method has a recognition rate of 99.3% outperforming FRGC baseline algorithms [63]. Robustness is also demonstrated in the presence of noise or facial expressions.

D. *Discussion*

The increase in available information is expected to improve face recognition performance as the combination of different facial modalities can overcome some recognition challenges. However, each facial modality has its drawbacks like being sensitive to illuminations, glasses, ambient temperatures, or expressions. Combining them can enhance performance and reduce reliability as well.

How to combine to maximize the performance is challenging. In [96], [104], [106], and [117], rules for fusing data or decisions are investigated. They use weights to control the influences from different modalities. Appropriate weights chosen for fusion can improve the performance, but it is also possible to decrease the performance. For example, in [96], when user faces are not

frontal, the use of facial texture information has an adverse effect on their algorithm because pose causes illumination changes on faces.

IV. Face Databases and Performance Evaluation

To train and test faces for recognition, most methods need gallery and probe datasets. Most face recognition methods are tested on datasets with a small number of images with few variations in the different probe datesets. Benchmark databases have been created, where the size of the gallery and probe set, the inclusion of face variations, as well as multiple face modalities are important parameters to evaluate how accurate and robust face recognition methods are. Next, we characterize these datasets, and performance evaluation of different methods is also discussed.

A. Face Databases

A 3-D face model and IR face image databases are presented in Tables IV and V. For *3-D face databases*, UND has contributed ND-2006, 3-D TEC, Collection D, and Collection NDOff-2007 and has supported FRGC v2.0. The ND-2006 database consists of 13 450 scans of 888 subjects with six different types of expressions. The 3-D TEC dataset includes 428 scans from 107 pairs of twins with two expressions. The Collection D dataset has 953 frontal face images from 277 subjects captured in different sessions. Collection NDOff-2007 has 6940 face images from 387 subjects with a variety of head orientations. Each image includes the information of facial colours and 3-D locations. FRGC v2.0 [63] is widely used as a benchmark database. Fifty thousand scan images are contained from 466 individuals, where frontal views of faces with expression variations are captured under different illuminations. Challenging nonneutral facial expressions (like blown cheeks and open mouths) are also contained. However, none have major pose variations or glasses.

The University of York (UOY) provides 5000 face range images with lighting, pose, and expression variations. BJUT-3-D [118] includes 500 Chinese people (250 females). Each subject has varying expressions and is captured under controlled illumination. A 3-D_RMA contains 120 subjects captured twice in two sessions with three poses. GavabDB contains 427 triangle meshes from 61 subjects (16 females). Each individual is asked to rotate the head with neutral expressions and view straight forward with different expressions. XM2VTS is a multimodal face database of 295 subjects recording color images, sound files, video sequences, and 3-D models. The MPI face database collects seven views for each head without hair for constructing a full 3-D head model. Five 3-D head models with full views of each face are available in this database. FRAV3D [119] provides multiple representations of faces from 105 subjects, including visual color images, 2.5-D range images, and 3-D meshes. The data are acquired with different expressions, and variations of poses and illuminations. The basel face model (BFM) database [120] collects registered 3-D scans with textures from 100 females and 100 males, where each subject is captured with neutral expression under controlled illumination, and more synthesized face models can be constructed by 3-D morphable face models. CAESAR provides the largest collection of 5000 subjects' body scans, where both texture and 3-D points are included. The CASIA 3-D database consists of 4624 scans and color images from 123 persons, which covers both the single variation and combined variations of poses, expressions, and illuminations. For the subjects with glasses, one additional scan with glasses is available in the database. SHREC2007 collects 640 face scans from 64 subjects. All subjects have neutral expressions without any pose variations and occlusions on faces.

Some datasets provide landmarks of faces. Bosphorus consists of 4666 face scans from 105 subjects, which contains a wide range of expression, pose, and occlusion variations. Three-dimensional coordinates of 24 labelled landmarks for each face model are also available. Texas 3DFRD consists of 1149 pairs of color and range images from 105 subjects. Each face is acquired with a frontal pose in a standardized position. Twenty-five labeled landmarks for a face are available in the database. UMB-DB provides 1473 records from 143 subjects with expression variations and occlusions. Seven landmarks for a face are available as well as the mask of visible parts.

For *IR face databases*, UND's collection C consists of 2492 LWIR images for frontal faces from 241 subjects. UND's collection X1 dataset contains 2292 pairs of IR and visual frontal face images from 82 subjects. Both collections consider the variations of expressions and illuminations. SCface [121] contains 4160 images in both visual and IR spectra from 130 subjects with different poses, and the images are acquired using different quality and resolution cameras under uncontrolled lighting conditions. It also provides the coordinates of eyes, nose, and mouth. USTC-NVIE [122] covers visual and LWIR data from 215 subjects. Each subject has six expressions with or without glasses captured under the illuminations from three different directions. NIST Equinox contains more than 600 subjects in visible, LWIR, MWIR, and SWIR spectra and covers various conditions such as expression and illumination changes with or without glasses. In IRIS, 4228 pairs of thermal and visual images from 30 subjects are collected with lighting condition and expression variations.

B. Performance Evaluation

Tables VI, VII, and VIII summarize, respectively, 3-D, IR, and multimodal face recognition methods discussed herein. The numbers of subjects and images used in the databases are listed to compare different algorithms and corresponding performance rates of recognition or verification at 0.1% FAR. The text describes performance of methods under variations such as wearing glasses, different expressions, poses, illuminations, sessions, and ambient temperatures. CMU-PIE [24], FERET [25], and FRVT 2002 [123] in Table VI are visual face databases.

To evaluate performances across different methods, a benchmark database is required with sufficient challenging variations. For 3-D and 3-D related multimodal face recognition, FRGC v2.0 is the most challenging [124], widely used for performance evaluation [5], [33], [37], [42], [43], [48], [57]–[59], [66], [95], [99], [100]. Comparing these methods, Mohammadzade and

TABLE IV
3-D FACE MODEL DATABASES

Data Set	Web Address	Tex. Img	Description
UND dataset	http://www.nd.edu/~cvrl/CVRL/Data_Sets.html	Yes	Several sets of facial collections in different sessions with varying poses and expressions
FRGC v2.0	http://www.nist.gov/itl/iad/ig/frgc.cfm	Yes	50000 records from 466 subjects with neutral and non-neutral expressions under controlled and uncontrolled illuminations
UOY dataset	http://www-users.cs.york.ac.uk/~nep/research/3Dface/tomh/3DFaceDatabase.html	No	5000 3D face models from 350 subjects with lighting, facial orientation and expression changes
BJUT-3D	http://www.bjut.edu.cn/sci/multimedia/mul-lab/3dface/face_database.htm	Yes	500 subjects with neutral, happy, surprise and anger expressions under controlled illumination
3-D RMA	http://www.sic.rma.ac.be/~beumier/DB/3d_rma.html	No	120 subjects with three poses for two sessions
GavabDB	http://gavab.escet.urjc.es/articulos/GavabDB.pdf	No	427 triangle meshes of 61 subjects with pose and expression variations
XM2VTS	http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/	No	Commercial database with VRML 3-D data from 295 subjects
MPI	http://faces.kyb.tuebingen.mpg.de/	No	200×7 scan images from 200 subjects with five sets of full 3-D head models
FRAV3D	http://www.frav.es/pdf/2006/icip2006.pdf	Yes	106 subjects covering different poses, lighting conditions, and gestures
BFM	http://faces.cs.unibas.ch/bfm/main.php?nav=1-0&id=basel_face_model	Yes	200 subjects with neutral expressions under controlled illumination
BU-3DFE	http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html	Yes	100 subjects with seven expressions from two views
CAESAR	http://store.sae.org/caesar/	Yes	Texture and scan images from 5000 subjects (whole bodies)
CASIA 3D	http://biometrics.idealtest.org/	Yes	4624 records from 123 subjects with expression, pose, and illumination variations
SHREC2007	http://give-lab.cs.uu.nl/SHREC/shrec2007/	No	640 face scans from 64 subjects with neutral expression
Bosphorus	http://bosphorus.ee.boun.edu.tr/default.aspx	No	4666 scans from 105 subjects with different expressions, poses, and occlusions as well as available facial landmarks
Texas 3DFRD	http://live.ece.utexas.edu/research/texas3dfr/	Yes	1149 pairs of color and range images from 105 subjects with facial fiducial points
UMB-DB	http://www.ivl.disco.unimib.it/umbdb/description.html	Yes	1473 acquisitions from 143 subjects with various expressions and occlusions with landmarks

TABLE V
INFRARED FACE IMAGE DATABASES

Data Set	Web Address	Vis. Img	Description
UND Co.C	http://www.nd.edu/~cvrl/CVRL/Data_Sets.html	No	2492 LWIR frontal face images from 241 subjects
UND Co.X1	http://www.nd.edu/~cvrl/CVRL/Data_Sets.html	Yes	2292 IR (and visual) frontal face images from 82 subjects
SCface	http://www.scface.org/	Yes	4160 images from 130 subjects with different poses under controlled lighting conditions
USTC-NVIE	http://nvie.ustc.edu.cn/	Yes	Six expression images from 215 subjects with or without glasses
Equinox	http://www.equinoxsensors.com/products/HID.html	Yes	18629 infrared images including expressions, illumination changes with or without glasses
IRIS	http://www.cse.ohio-state.edu/otcbvs-bench/	Yes	4228 pairs of thermal and visual images from 30 subjects with changes of poses, lighting, and expressions

Hatzinakos [33] report the best verification rate of 99.6%. The size of training datasets affects the performance, as studied in [125] where different face representations, feature extractions, and various fusion rules are investigated. If the training size is small, face representation is more important than feature extraction; otherwise, feature extraction dominates face recognition.

For IR and IR related multimodal face recognition, the Equinox database is a popular one for evaluating performances [9], [103], [105], [109]–[111]. Based on the comparison of these methods, the best record is reported in [103] with a recognition rate over 95%. Guzman *et al.* [79] achieve an average accuracy of 90.39%. However, their experiment is conducted on a small dataset from 13 subjects under indoor conditions.

Performance is dependent on a few factors. First, the variety of gallery and probe images greatly changes the performance. Face images with time delay, presence of glasses, hair, scarf and disguises, and variations of image qualities, lighting conditions, expressions, and poses challenge face recognition. Neural and frontal faces under controlled conditions are easier cases for face recognition. Occlusions and time delay are more challenging. In [111], partial occlusions are considered with scarves covering the lower half of the face. In [82], the effect of time lapse is evaluated on NIR face recognition. The challenge of image quality is investigated in [94]. Race variety is another consideration in gallery and probe images. Training with Caucasian faces may not work for Asian faces. Li *et al.* [82] discuss the performance of their method with respect to unseen ethnic faces. In addition, the size of gallery and probe images directly affects the final recognition rate. The second factor is computational resources. Most work usually ignores this factor. However, computational resources are important for real applications, where accuracy may be sacrificed for speed. Finally, some results are reported based on tuning parameters, which specifically adjust face recognition systems for different

TABLE VI
3-D FACE RECOGNITION ALGORITHMS

Methods	Subjects	Images	Database	Rates(%)	Description
Blanz [23]	68 194	4488, 1940	CMU-PIE, FERET	95, 95.9	Face images varying in poses, illuminations, and expressions
Blanz [26]	87	435	FRVT 2002	<90	Five poses included in images
Hu [28]	68	41368	CMU-PIE	<95	Frontal and neutral face images used for synthesizing images with different poses, illuminations and expressions
Mo. [33]	466	4007	FRGC v2.0	99.6	Faces of varying illuminations, expressions, and views
Al. [37]	466	1597	FRGC v2.0	93.78	Faces with neutral expressions
Samir [38]	50, 162	300, 720	FSU, UND	92, 90.4	Six expressions for each subject in FSU
Drira [39]	466, 61, 105	1597, 427, 381	FRGCv2, GavaDB, Bosphorus	97.7, 96.99, 89.25	Faces with expression, occlusion, and pose variations
Fa. [42]	466	3541	FRGC v2.0	97.2	Expression variations and partial faces occluded in images
Lei [43]	466	1000	FRGC v2.0	95.6	Expression variations
Lin [45]	275	943	FRGC v1.0	97.2	Faces under varying illumination, focus, and expression
He. [46]	100	330	UOY	–	Images with different expressions and poses
Russ [48]	200	468	FRGC	<97	Faces with neutral and nonneutral expressions
Pan [66]	276	943	FRGC v1.0	95	Faces under varying illumination, focus, and expression
Ka. [5]	466	3541	FRGV v2.0	97.3	Expression and pose changes in images
Al. [57]	466	1597	FRGC v2.0	>97.73	Each subject with one neutral and six other expressions
Wang [6]	100, 466	2500, 1597	BU-3DFE, FRGC v2.0	98	Faces captured under different illuminations, expressions, and views
Wang [58]	466	1597	FRGC v2.0	>98	Each subject with one neutral and six other expressions
Liu [59]	466	4007	FRGC v2.0	96.94	Each subject with neutral, smile, surprise, and angry expressions

TABLE VII
INFRARED FACE RECOGNITION ALGORITHMS

Methods	Subjects	Images	Database	Rates(%)	Description
Trujillo [71]	30	270	IRIS	–	Three expressions of three poses in images
Wu [74]	41	2614	–	<90.1	Faces wearing glasses or not under varying poses, expressions, sessions, and ambient temperatures
Sr. [75]	9	180	–	97.14	Faces with glasses or not under varying poses and expressions
Bu. [76]	50	2750	Equinox	89.6	MWIR images with glasses or not, different expressions, and poses
Bu. [77]	138	7590	–	86	Subjects with varying poses and expressions
Guzman [79]	13	–	–	90.39	MWIR images captured with a frontal view in different days under similar temperature
Li [82]	1000	10000	–	91.8	NIR images with glasses or not, differnt illumination, time lapse, and ethnic groups
Pan [83]	200	1400	–	<95	NIR images with poses and expression variations

probe images. These factors complicate the assessment of face recognition methods.

V. DISCUSSION

Varying expression, pose, illumination, aging, and makeup remain challenges for face recognition [100]. Three-dimensional and IR face recognition may address unsolved issues in visual face recognition. Whether they improve the performance in terms of these factors is unclear. Chang *et al.* [12] show that for using a single image, the performance of 3-D face recognition is equivalent to visual face recognition, where the PCA-based recognition engine is used for both visual and 3-D. Other algorithms can better exploit 3-D information and can result in better performance than visual face recognition, or *vice versa*.

Mian *et al.* [36], [37], [57], [100] demonstrate advantages of the use of 3-D data to build expression-invariant representation and automatically correct face poses, where high performance is achieved in the presence of large expression and pose variations. The 3-D modality has inherent advantages over visual for face recognition. Huang *et al.* [126] claim that 3-D images "are theoretically reputed to be robust to variations in illumination."

As 3-D face images capture shape information and represent the geometry of faces, they are much less dependent on illuminations and viewpoint than visual data [100]. Drira *et al.* [39] claim that among different facial modalities, 3-D images are advantageous to illumination and small pose changes. As concluded in [2], "it is commonly thought that the use of 3-D sensing has the potential for greater recognition accuracy than 2-D." However, 3-D facial scans are likely to be more sensitive to expression changes [126]. Much effort addressed expression-invariant 3-D face recognition [124]. The challenge of aging is also a issue of 3-D face recognition. The 3-D aging model [127] is developed to assist face recognition to be aging-invariant. Aging-invariant 3-D face recognition has not received adequate attention. In addition, 3-D-based approaches often suffer from computation, possible imprecise alignments, and undesirable artifacts generated on the model-based virtual views [128]. Li *et al.* [82] also mention the disadvantages of 3-D face recognition: the increased cost and slowed speed.

IR face recognition is a growing area. A study of visual and IR face recognition [116] reports that for two factors (degree of lighting variation and time lapse between the gallery and probe images), visual and IR-based face recognition have

TABLE VIII
MULTIMODAL FACE RECOGNITION ALGORITHMS

Methods	Subjects	Images	Database	Rates(%)	Description
visual+3-D					
Mian [92]	275	943	FRGC v1.0	100	Faces with expressions under controlled illumination
Jahanbin [93]	105	1149	Texas 3FRD	99.76	Faces with expression and illumination variations
Al-Osaimi [95]	466	1000	FRGC v2.0	97.4	Neutral and nonneutral expressions
Xu [99]	466	3200	FRGC v2.0	95.3	2 expressions under varying illuminations
Mian [100]	466	3541	FRGC v2.0	97.4	Neutral and nonneutral expressions
visual+IR					
Bu. [103]	45	4552	Equinox	98.79	MWIR images with different illumination conditions and wearing glasses or not
Pop [117]	30	4228	IRIS	74.85	Three expressions and eight poses of each subject
Ahmad [105]	24	–	Equinox	<97.4	Images with a frontal illumination
Bhowmik [106]	16	210	IRIS	93	Different lightings, expressions, and poses
Hanif [109]	7	17	Equinox	95.84	Variable expressions and lightings
Heo [110]	90	3224	Equinox	<98.5	With/without glasses, different expressions, and lighting
Gu. [111]	40	520	Equinox	<99.24	Images under varying lighting, expression, and partial occlusion conditions
visual+IR+3-D					
Chang [13]	191	424	–	100	Images captured with a normal expression against a plain gray background

similar performance in terms of lighting variation, and visual face recognition outperforms IR face recognition in terms of time lapse. Their IR images are acquired from 7.0–14.0 μm range (i.e., thermal images). The degradation using thermal imagery for face recognition is reported in [102] and [107]. However, the conclusion is in debate. The study uses the recognition engine PCA that is appearance-based, where the temperature values are directly used as face spaces. As analyzed in [77], with the difference between gallery and probe images, the temperatures sensitive to environmental temperatures, emotions and health condition, and glasses may cause the PCA algorithm to fail.

Recent developments are to exploit more powerful information from thermal images for face recognition. For example, in [77] and [103], they find the uniqueness of superficial blood vessels of the face (i.e., facial vascular network), which is time-invariant physiological information and can be extracted from thermal images. Comparing PCA-based thermal face recognition and facial vascular network-based face recognition in time-lapse experiments shows that face recognition using physiological feature vectors can effectively address the low-performance problem of PCA-based thermal face recognition. Moreover, the physiological information is invariant to physical conditions, an advantage of thermal imagery to other face modalities. In terms of light variations, NIR shows the capability to achieve high performance. Li *et al.* [82] demonstrate that NIR images are advantageous over visible light under different lighting directions. Therefore, the IR modality is suggested as a possible biometric due to its potential of being constant for a person across the time [77] and different illuminations [82].

To rank the superiority among the three modalities for face recognition relies on the development of recognition engines and evaluations. Visual imagery is more easily acquired with high quality. Three-dimensional facial data matches the human vision system, especially combined with visual texture. The IR modality, especially thermal images, can provide a facial vascular network that is unique to each subject and difficult to change. Face recognition using single visual, 3-D or IR modalities has its own advantages and disadvantages, but multimodal face recognition yields better performance than corresponding singlemodal face recognition [12]–[14], [112], [116]. However, the current state of art makes clear that no existing technique can deal with all challenges including occlusion, pose, expression, illumination, disguise, time lapse, body conditions, background clutter, and automation, robust for diverse real-world scenarios.

VI. CONCLUSION

This paper reviews recent advances on face recognition. As 3-D data contain facial geometry information, they increase robustness to viewpoints and illumination variations compared with visual images. Thus, 3-D face recognition can achieve greater recognition accuracy than visual face recognition, especially under uncontrolled conditions. Although 3-D data are asserted to be insensitive to illumination variations, it is still difficult to recognize faces in the absence of visible light. Aging is also challenging for 3-D face recognition. IR images capture internal face anatomical structure, instead of external appearance and environment and are, thus, effective for face recognition under these conditions. In addition, studies show that fusions of multiple face modalities can achieve better performance than a single modality. There are efforts on visual+3-D and visual+IR face recognition, but the work on 3-D+IR and visual+IR+3-D is comparatively rare. That is mainly because 1) visual images carry face texture information that is particularly useful for face recognition; 2) visual images are easy to acquire and process; and 3) no devices currently exist that can capture faces with the three modalities synchronously.

Based on the techniques surveyed, we conclude the following.

1) Visual face recognition has the highest technology maturity for real applications.
2) Great progress has been achieved in 3-D face recognition to deal with the challenges of facial pose and expression variations.
3) IR face recognition can address the problems in terms of time delay and lighting conditions.

4) Multimodal face recognition has potential to overcome the limitations of different imaging modalities. Multimodal face recognition with optimal fusion are promising future directions of face recognition.

ACKNOWLEDGMENT

The authors would like to thank E. Bass and anonymous reviewers for their suggestions on earlier version of this manuscript.

REFERENCES

[1] L. Introna and H. Nissenbaum, "Facial recognition technology: A survey of policy and implementation issues," Dept. Org., Work, Technol., Working Paper, 2010.

[2] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Comput. Vis. Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[3] A. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recog. Lett.*, vol. 28, no. 14, pp. 1885–1906, 2007.

[4] C. Xu, Y. Wang, T. Tan, and L. Quan, "Depth versus intensity: Which is more important for face recognition?" in *Proc. Int. Conf. Pattern Recog.*, 2004, vol. 4, pp. 342–345.

[5] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 640–649, Apr. 2007.

[6] Y. Wang, X. Tang, J. Liu, G. Pan, and R. Xiao, "3D face recognition by local shape difference boosting," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 603–616.

[7] S. Kong, B. A. J. Heo, J. Paik, and M. Abidi, "Recent advances in visual and infrared face recognition—A review," *Comput. Vis. Image Understanding*, vol. 97, no. 1, pp. 103–135, 2005.

[8] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, "Face recognition by fusing thermal infrared and visible imagery," *Image Vis. Comput.*, vol. 24, no. 7, pp. 727–742, 2006.

[9] D. Socolinsky, A. Selinger, and J. Neuheisel, "Face recognition with visible and thermal infrared imagery," *Comput. Vis. Image Understanding*, vol. 91, no. 1/2, pp. 72–114, 2003.

[10] D. Socolinsky and A. Selinger, "Thermal face recognition in an operational scenario," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2004, vol. 2, pp. 1012–1019.

[11] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2D and 3D facial data," in *Proc. ACM Workshop Multimodal User Authentication*, 2003, pp. 25–32.

[12] K. Chang, K. Bowyer, and P. Flynn, "An evaluation of multimodal 2D+3D face biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 619–624, Apr. 2005.

[13] K. Chang, K. Bowyer, P. Flynn, and X. Chen, "Multi-biometrics using appearance, shape and temperature," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recog.*, 2004, pp. 43–48.

[14] K. Chang, K. Bowyer, P. Flynn, and X. Chen, "Face recognition using 2-D, 3-D, and infrared: Is multimodal better than multisample?" *Proc. IEEE*, vol. 94, no. 11, pp. 2000–2012, Nov. 2006.

[15] R. Jafri and H. Arabnia, "A survey of face recognition techniques," *J. Inform. Process. Syst.*, vol. 5, no. 2, pp. 41–68, 2009.

[16] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[17] A. Patil, S. Kolhe, and P. Patil, "2D face recognition techniques: A survey," *Int. J. Mach. Intell.*, vol. 2, no. 1, pp. 74–83, 2010.

[18] H. Cevikalp, M. Neamtu, M. Wikes, and A. Barkana, "Discriminative common vectors for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 4–13, Jan. 2005.

[19] Q. Li, J. Ye, and C. Kambhamettu, "Linear projection methods in face recognition under unconstrained illuminations: A comparative study," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2004, pp. 474–481.

[20] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 572–581, May 2004.

[21] M. Er, S. Wu, J. Lu, and H. Toh, "Face recognition with radial basis function neural networks," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 697–710, May 2002.

[22] X. Lu, A. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 31–43, Jan. 2006.

[23] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.

[24] (2013). CMU-PIE database. [Online]. Available: http://vasc.ri.cmu.edu/idb/html/face/

[25] (2013). FERET database. [Online]. Available: http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

[26] V. Blanz, P. Grother, P. Phillips, and T. Vetter, "Face recognition based on frontal views generated from non-frontal images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 2, pp. 454–461.

[27] X. Lu, R. Hsu, A. Jain, B. Kamgar-Parsi, and B. Kamgar-parsi, "Face recognition with 3D model-based synthesis," in *Proc. Int. Conf. Biometric Authentication*, 2004, pp. 139–146.

[28] Y. Hu, D. Jiang, S. Yan, L. Zhang, and H. Zhang, "Automatic 3D reconstruction for face recognition," in *Proc. IEEE Int. Conf. Pattern Recog.*, 2004, pp. 843–850.

[29] M. Song, D. Tao, X. Huang, C. Chen, and J. Bu, "Three-dimensional face reconstruction from a single image by a coupled RBF network," *IEEE Tran. Image Process.*, vol. 21, no. 5, pp. 2887–2897, May 2012.

[30] C. Chua, F. Han, and Y. Ho, "3D human face recognition using point signature," in *Proc. Int. Conf. Automat. Face Gesture Recog.*, 2000, pp. 233–238.

[31] I. Mpiperis, S. Malasiotis, and M. Strintzis, "3D face recognition by point signatures and iso-contours," in *Proc. Int. Conf.: Signal Process., Pattern Recog., Appl.*, 2007, pp. 328–332.

[32] A. Abate, M. Nappi, S. Ricciardi, and G. Sabatino, "Fast 3D face recognition based on normal map," in *Proc. IEEE Int. Conf. Image Process.*, 2005, vol. 2, pp. 946–949.

[33] H. Mohammadzade and D. Hatzinakos, "Iterative closest normal point for 3D face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 381–397, Feb. 2013.

[34] A. Moreno, A. Sanchez, J. Velez, and F. Diaz, "Face recognition using 3D surface-extracted descriptors," in *Proc. Irish Mach. Vis. Image Process. Conf.*, 2003, pp. 1–8.

[35] Y. Lee and J. Shim, "Curvature-based human face recognition using depth-weighted Hausdorff distance," in *Proc. Int. Conf. Image Process.*, 2004, pp. 1429–1432.

[36] A. Mian, M. Bennamoun, and R. Owens, "Matching tensors for pose invariant automatic 3D face recognition," presented at the IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog., San Diego, CA, USA, Jun. 2005, vol. 3.

[37] F. Al-Osaimi, M. Bennamoun, and A. Mian, "Integration of local and global geometrical cues for 3D face recognition," *Pattern Recog.*, vol. 41, no. 3, pp. 1030–1040, 2008.

[38] C. Samir, A. Srivastava, and M. Daoudi, "Three-dimensional face recognition using shapes of facial curves," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1858–1863, Nov. 2006.

[39] H. Drira, B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions, and pose variations," *IEEE Tran. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2270–2283, Sep. 2013.

[40] S. Berretti, A. Bimbo, and P. Pala, "3D face recognition using isogeodesic stripes," *IEEE Tran. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2162–2176, Dec. 2010.

[41] K. Chang, K. Bowyer, and P. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1695–1700, Oct. 2006.

[42] T. Faltemier, K. Bowyer, and P. Flynn, "A region ensemble for 3D face recognition," *IEEE Trans. Inform. Forensics Security*, vol. 3, no. 1, pp. 62–73, Mar. 2008.

[43] Y. Lei, M. Bennamoun, and A. El-Sallam, "An efficient 3D face recognition approach based on the fusion of novel local low-level features," *Pattern Recog.*, vol. 46, no. 1, pp. 24–37, 2013.

[44] W. Lin, N. Boston, and Y. Hu, "Summation invariant and its applications to shape recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. 205–208.

[45] W. Lin, K. Wong, N. Boston, and Y. Hu, "Fusion of summation invariants in 3D human face recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2006, vol. 2, pp. 1369–1376.

[46] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition: An eigensurface approach," in *Proc. Int. Conf. Image Process.*, 2004, vol. 2, pp. 1421–1424.

[47] C. Hesher, A. Srivastava, and G. Erlebacher, "Principal component analysis of range images for facial recognition," in *Proc. Int. Conf. Imag. Sci., Syst. Technol.*, 2002.

[48] T. Russ, C. Boehnen, and T. Peters, "3D face recognition using 3D alignment for PCA," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2006, vol. 2, pp. 1391–1398.

[49] A. Srivastava, X. Liu, and C. Hesher, "Face recognition using optimal linear components of range images," *Image Vis. Comput.*, vol. 24, no. 3, pp. 291–299, 2006.

[50] Y. Taghizadegan, H. Ghassemian, and M. Naser-Moghaddasi, "3D face recognition method using 2DPCA Eucilidean distance classification," *ACEEE Int. J. Control Syst. Instrum.*, vol. 3, no. 1, pp. 149–155, 2012.

[51] X. Lu, D. Colbry, and A. Jain, "Three-dimensional model based face recognition," in *Proc. Int. Conf. Pattern Recog.*, 2004, vol. 1, pp. 362–366.

[52] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza, "Evaluation of 3D face recognition in the presence of facial expressions: An annotated deformable model approach," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog. Workshops*, 2005, vol. 3, pp. 1–8.

[53] G. Passalis, I. Kakadiaris, and T. Theoharis, "Intra-class retrieval of non-rigid 3D objects: Application to face recogntion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 218–229, Feb. 2007.

[54] X. Lu and A. K. Jain, "Deformation modeling for robust 3D face matching," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2006, pp. 1377–1383.

[55] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *Int. J. Comput. Vis.*, vol. 64, no. 1, pp. 5–30, 2005.

[56] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant representations of faces," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 188–197, Jan. 2007.

[57] F. Al-Osaimi, M. Bennamoun, and A. Mian, "An expression deformation approach to non-rigid 3D face recognition," *Int. J. Comput. Vis.*, vol. 81, no. 3, pp. 302–316, 2009.

[58] Y. Wang, J. Liu, and X. Tang, "Robust 3D face recognition by local shape difference boosting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 20, pp. 1858–1870, Oct. 2010.

[59] P. Liu, Y. Wang, D. Huang, Z. Zhang, and L. Chen, "Learning the spherical harmonic features for 3-D face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 914–925, Mar. 2013.

[60] C. Chua and R. Jarvis, "Point signatures: A new representation for 3D object recognition," *Int. J. Comput. Vis.*, vol. 25, no. 1, pp. 63–85, 1997.

[61] G. Gordon, "Face recognition based depth and curvature features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 1992, pp. 808–810.

[62] M. Truk and A. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 1991, pp. 586–591.

[63] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 1, pp. 947–954.

[64] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[65] J. Yang, D. Zhang, A. Frangi, and J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[66] G. Pan, S. Han, Z. Wu, and Y. Wang, "3D face recognition using mapped depth images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog. Workshops*, 2005, pp. 1–7.

[67] B. Amor, K. Ouji, M. Ardabilian, and L. Chen, "3D face recognition by ICP-based shape matching," in *Proc. Int. Conf. Mach. Intell.*, 2005, pp. 1–6.

[68] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenface versus fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[69] D. Socolinsky, L. Wolff, J. Neuheisel, and C. Eveland, "Illumination invariant face recognition using thermal infrared imagery," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2001, vol. 1, pp. 527–534.

[70] J. Kang, A. Borkar, A. Yeung, N. Nong, M. Smith, and M. Hayes, "Short wavelength infrared face recognition for personalization," in *Proc. IEEE Int. Conf. Image Process.*, 2006, pp. 2757–2760.

[71] L. Trujillo, G. Olague, R. Hammound, and B. Hernandez, "Automatic feature localization in thermal images for facial expression recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog. Workshops*, 2005, vol. 3, pp. 1–7.

[72] X. Zou, J. Kittler, and K. Messer, "Face recognition using active near-IR illumination," in *Proc. British Mach. Vis. Conf.*, 2005, pp. 153–154.

[73] M. Akhloufi and A. Bendada, "Probabilistic Bayesian framework for infrared face recognition," in *Proc. Int. Conf. Mach. vis., Image Process. Pattern Anal.*, 2009, pp. 66–70.

[74] S. Wu, W. Song, L. Jiang, S. Xie, F. Pan, and W. Yau, "Infrared face recognition by using blood perfusion data," in *Proc. Int. Conf. Audio Video-Based Biometric Person Authentication*, 2005, pp. 320–328.

[75] A. Srivastava, X. Liu, B. Thomasson, and C. Hesher, "Spectral probability models for IR images with applications to IR face recognition," in *Proc. IEEE Workshop Comput. Vis. Beyond Visible Spectrum: Methods Appl.*, 2001, pp. 66–70.

[76] P. Buddharaju, I. Pavlidis, and I. Kakadiaris, "Face recognition in the thermal infrared spectrum," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog. Workshops*, 2004, pp. 1–6.

[77] P. Buddharaju, I. Pavlidis, P. Tsiamyrtzis, and M. Bazakos, "Physiology-based face recognition in the thermal infrared spectrum," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 613–616, Apr. 2007.

[78] M. Akhloufi and A. Bendada, "Infrared face recognition using distance transform," in *Proc. World Acad. Sci., Eng. Technol.*, 2008, vol. 30, pp. 160–163.

[79] A. Guzman, M. Goryawala, J. Wang, A. Barreto, J. Andrian, N. Rishe, and M. Adjouadi, "Thermal imaging as a biometrics approach to facial signature authentication," *IEEE Trans. Inform. Technol. Biomed.*, vol. 17, no. 1, pp. 214–222, Jan. 2013.

[80] S. Li, R. Chu, M. Ao, L. Zhang, and R. He, "Highly accurate and fast face recognition using near infrared images," in *Proc. Int. Conf. Biometrics*, 2006, pp. 151–158.

[81] S. Li, L. Zhang, S. Liao, X. Zhu, R. Chu, M. Ao, and R. He, "A near-infrared image based face recognition system," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recog.*, 2006, pp. 455–460.

[82] S. Li, R. Chu, and S. L. L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 627–639, Apr. 2007.

[83] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 613–616, Dec. 2003.

[84] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Amer.*, vol. 14, pp. 1724–1733, 1997.

[85] C. Liu and H. Wechser, "Comparative assessment of independent component analysis (ICA) for face recognition," in *Proc. Int. Conf. Audio Video-Based Biometric Person Authentication*, 1999, pp. 211–216.

[86] A. Selinger and D. Socolinsky, "Face recognition in the dark," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog. Workshops*, 2004, vol. 8, pp. 1–6.

[87] A. Selinger and D. Socolinsky, "A comparative analysis of face recognition performance with visible and thermal infrared imagery," in *Proc. Int. Conf. Pattern Recog.*, 2002, vol. 4, pp. 1–6.

[88] C. Manohar, "Extraction of superficial vasculature in thermal imaging," Master Thesis, Dept. Elect. Eng., University of Houston, Houston, TX, USA, 2004.

[89] A. Guzman. (2013). Thermal imaging as a biometrics approach to facial signature authentication. Ph.D. dissertation, Dept. Elect. Eng., Florida Int. Univ., Miami, FL, USA. [Online]. Available: http://digitalcommons.fiu.edu/etd/539/

[90] J. Jasinski. (2013). Multi-modal biometric testing and evaluation. [Online]. Available: http://biometrics.nist.gov/cs_links/ibpc2012/presentations/Day1/114_jas inski.pdf

[91] F. Tsalakanidou, D. Tzovaras, and M. Strintzis, "Use of depth and colour eigenfaces for face recognition," *Pattern Recog. Lett.*, vol. 24, no. 4, pp. 1427–1435, 2003.

[92] A. Mian, M. Bennamoun, and R. Owens, "2D and 3D multimodal hybrid face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 344–355.

[93] S. Jahanbin, H. Choi, and A. Bovik, "Passive multimodal 2-D+3-D face recognition using Gabor features and landmark distances," *IEEE Trans. Inform. Forensics Security*, vol. 6, no. 4, pp. 1287–1304, Dec. 2011.

[94] F. Tsalakanidou, S. Malassiotis, and M. Strintzis, "Face localization and authentication using color and depth images," *IEEE Trans. Image Process.*, vol. 14, no. 2, pp. 152–168, Feb. 2005.

[95] F. Al-Osaimi, M. Bennamoun, and A. Mian, "Spatially optimized data-level fusion of texture and shape for face recognition," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 859–872, Feb. 2012.
[96] T. Papatheodorou and D. Rueckert, "Evaluation of automatic 4D face recognition using surface and texture registration," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recog.*, 2004, pp. 321–326.
[97] X. Lu, A. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 31–43, Jan. 2006.
[98] Y. Wang, C. Chua, and Y. Ho, "Facial feature detection and face recognition from 2D and 3D images," *Pattern Recog. Lett.*, vol. 23, no. 10, pp. 1191–1202, 2002.
[99] C. Xu, S. Li, T. Tan, and L. Quan, "Automatic 3D face recognition from depth and intensity gabor features," *Pattern Recog.*, vol. 42, no. 9, pp. 1895–1905, 2009.
[100] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1297–1943, Nov. 2007.
[101] C. Abdelkader and P. Griffin, "Comparing and combining depth and texture cues for face recognition," *Image Vis. Comput.*, vol. 23, no. 3, pp. 339–352, 2005.
[102] D. Socolinsky and A. Selinger, "Thermal face recognition in an operational scenario," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2004, vol. 2, pp. 1012–1019.
[103] P. Buddharaju and I. Pavlidis, "Multi-spectral face recognition fusion of visual imagery with physiological information," in *Face Biometrics for Personal Identification, Signals and Communication Technology*. New York, NY, USA: Springer-Verlag, 2007, pp. 91–108.
[104] X. Chen, Z. Jing, and G. Xiao, "Fuzzy fusion for face recognition," in *Proc. Int. Conf. Fuzzy Syst. Knowl. Discovery*, 2005, pp. 672–675.
[105] J. Ahmad, U. Ali, and R. Qureshi, "Fusion of thermal and visual images for efficient face recognition using Gabor filter," in *Proc. IEEE Int. Conf. Comput. Syst. Appl.*, 2006, pp. 135–139.
[106] M. Bhowmik, K. Basu, and M. Kundu, "Optimum fusion of visual and thermal face images for recognition," in *Proc. Int. Conf. Inform. Assur. Security*, 2010, pp. 311–316.
[107] S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis, "Infrared and visible image fusion for face recognition," in *Proc. SPIE*, 2004, vol. 5404, pp. 585–596.
[108] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, "Face recognition by fusing thermal infrared and visible imagery," *Image Vis. Comput.*, vol. 24, no. 7, pp. 727–742, 2006.
[109] M. Hanif and U. Ali, "Optimized visual and thermal image fusion for efficient face recognition," in *Proc. Int. Conf. Inform. Fusion*, 2006, pp. 1–6.
[110] J. Heo, S. Kong, B. Abidi, and M. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog., Workshops*, 2004, pp. 94–99.
[111] S. Gundimada and V. Asari, "Facial recognition using multisensory images based on localized kernel eigen spaces," *IEEE Trans. Image Process.*, vol. 18, no. 6, pp. 1314–1325, Jun. 2009.
[112] I. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza, "Multimodal face recognition: combination of geometry with physiological information," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 2, pp. 1022–1029.
[113] S. Sharma, *Applied Multivariate Techniques*. New York, NY, USA: Wiley, 1995.
[114] D. Blackburn, M. Bone, and P. Phillips, "Facial recognition vendor test 2000: Evaluation report," Defense Advanced Research Projects Agency, National Institute of Justice, 2001.
[115] X. Chen, P. Flynn, and K. Bowyer, "PCA-based face recognition in infrared imagery: Baseline and comparative studies," in *Proc. IEEE Int. Workshop Anal. Model. Faces Gestures*, 2003, pp. 127–134.
[116] X. Chen, P. Flynn, and K. Bowyer, "IR and visible light face recognition," *Comput. Vis. Image Understanding*, vol. 99, no. 3, pp. 332–358, 2005.
[117] F. Pop, M. Gordan, C. Florea, and A. Vlaicu, "Fusion based approach for thermal and visible face recognition under pose and expressivity variation," in *Proc. Roedunet Int. Conf.*, 2010, pp. 61–66.
[118] (2013). The BJUT-3D large-scale chinese face database. [Online]. Available: http://www.bjut.edu.cn/sci/multimedia/mul-lab/3dface/pdf/MISKL-TR-05-FM FR-001.pdf
[119] A. Serrano and E. Cabello, "Multimodal 2D, 2.5D & 3D face verification," in *Proc. IEEE Int. Conf. Image Process.*, 2006, pp. 2061–2064.
[120] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2009, pp. 296–301.
[121] M. Grgic, K. Delac, and S. Grgic, "SCface-Surveillance cameras face database," *J. Multimedia Tools Appl.*, vol. 51, no. 3, pp. 863–879, 2011.
[122] Z. Liu, S. Lv, Y. Lv, G. Wu, P. Peng, F. Chen, and X. Wang, "A natural visible and infrared facial expression database for expression recognition and emotion inference," *IEEE Trans. Multimedia*, vol. 12, no. 7, pp. 682–691, Nov. 2010.
[123] (2013). FRVT 2002 database. [Online]. Available: http://www.nist.gov/itl/iad/ig/frvt-2002.cfm
[124] D. Smeets, P. Claes, J. hermans, D. Vandermeulen, and P. Suetens, "A comparative study of 3D face recognition under expression variations," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 5, pp. 710–727, Sep. 2012.
[125] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, and B. Sankur, "Representation plurality and fusion for 3-D face recognition," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 38, no. 1, pp. 155–173, Feb. 2008.
[126] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "3-D face recognition using eLBP-based facial description and local feature hybrid matching," *IEEE Tran. Inform. Forensics Security*, vol. 7, no. 5, pp. 1551–1565, Oct. 2012.
[127] U. Park, Y. Tong, and A. Jain, "Face recognition with temporal invariance: A 3D aging model," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recog.*, 2008, pp. 1–7.
[128] P. Lee, G. Hsu, Y. Wang, and Y. Hung, "Subject-specific and pose-oriented facial features for face recognition across poses," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 5, pp. 1357–1368, Oct. 2012.

Hailing Zhou received the Ph.D. degree from Nanyang Technological University, Singapore.

She is currently a Research Fellow with the Centre for Intelligent Systems Research, Deakin University, Geelong, Australia, focusing on the areas of pattern recognition and image processing.

Ajmal Mian received the Ph.D. degree from the University of Western Australia, Australia.

He is currently a Research Professor with the University of Western Australia, Crawley, Australia. His research interests include computer vision, pattern recognition, and multimodal biometrics.

Lei Wei received the Ph.D. degree from Nanyang Technological University, Singapore.

He is currently a Research Fellow with the Centre for Intelligent Systems Research, Deakin University, Geelong, Australia, focusing in the areas of human–machine interaction and haptic rendering.

Doug Creighton received the Ph.D. degree from Deakin University, Australia.

He is currently an Associate Professor and a Deputy Director of the Centre for Intelligent Systems Research, Deakin University, Geelong, Australia. His research interests include algorithms and methodologies for complex systems.

Mo Hossny received the Ph.D. degree from Deakin University, Geelong, Australia.

He is currently a Senior Research Fellow with the Centre for Intelligent Systems Researches, Deakin University, Australia, focusing on image fusion metrics.

Saeid Nahavandi (SM'07) received the Ph.D. degree from Durham University, U.K.

He is currently an Alfred Deakin Professor, Chair of Engineering, and the Director of the Center for Intelligent Systems Research with Deakin University, Australia.

Dr. Nahavandi is a Fellow of Engineers Australia, the Institution of Engineering and Technology.