Computer Vision and Image Understanding 154 (2017) 137–151

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

3D-2D face recognition with pose and illumination normalization

Ioannis A. Kakadiaris[a,1,*], George Toderici[a,2], Georgios Evangelopoulos[a,3], Georgios Passalis[a,b,4], Dat Chu[a,5], Xi Zhao[a,6], Shishir K. Shah[a,1], Theoharis Theoharis[a,b,7]

[a] Computational Biomedicine Laboratory (CBL), Department of Computer Science, Univ. of Houston, 4800 Calhoun, Houston, TX 77204, USA
[b] Department of Informatics, Univ. of Athens, TYPA Buildings, Panepistimiopolis, Ilisia, 15784, Athens, Greece

ARTICLE INFO

Article history:
Received 28 April 2015
Revised 24 February 2016
Accepted 28 April 2016
Available online 21 May 2016

Keywords:
Face and gesture recognition
Biometrics
Physically-based modeling
3D-2D face recognition
Illumination normalization
Model-based face recognition
3D-2D model fitting
Object recognition
Computer vision

ABSTRACT

In this paper, we propose a 3D-2D framework for face recognition that is more practical than 3D-3D, yet more accurate than 2D-2D. For 3D-2D face recognition, the gallery data comprises of 3D shape and 2D texture data and the probes are arbitrary 2D images. A 3D-2D system (UR2D) is presented that is based on a 3D deformable face model that allows registration of 3D and 2D data, face alignment, and normalization of pose and illumination. During enrollment, subject-specific 3D models are constructed using 3D+2D data. For recognition, 2D images are represented in a normalized image space using the gallery 3D models and landmark-based 3D-2D projection estimation. A method for bidirectional relighting is applied for non-linear, local illumination normalization between probe and gallery textures, and a global orientation-based correlation metric is used for pairwise similarity scoring. The generated, personalized, pose- and light- normalized signatures can be used for one-to-one verification or one-to-many identification. Results for 3D-2D face recognition on the UHDB11 3D-2D database with 2D images under large illumination and pose variations support our hypothesis that, in challenging datasets, 3D-2D outperforms 2D-2D and decreases the performance gap against 3D-3D face recognition. Evaluations on FRGC v2.0 3D-2D data with frontal facial images, demonstrate that the method can generalize to databases with different and diverse illumination conditions.

1. Introduction

Face recognition (FR) has been a key topic in computer vision, pattern recognition, and machine learning research, with extensions to perceptual, behavioral, and social principles. In parallel, FR technology has been advancing in terms of sensors, algorithms, databases, and evaluation frameworks. This increasing interest is driven partly by the difficulty and challenges of the task (i.e., a complex, intra-class object recognition problem) and partly by a wide variety of applications involving identity management. Research challenges include (i) separating intrinsic from extrinsic appearance variations; (ii) developing discriminative representations and similarity metrics; and (iii) discovering performance invariants across heterogeneous data and conditions. Application-wise, face is emerging as a powerful biometric, a high-level semantic for content-based indexing and retrieval, and a natural and rich communication modality for human-computer interaction. The existing frameworks for face recognition vary across approaches (e.g., data-driven, model-based, perceptual) or facial data domains (e.g., images, point clouds, depth maps).

Methods for image-based FR have been pushing performance boundaries on nearly-frontal-view faces and constrained illumination conditions (Abate et al., 2007). However, the appearance of 2D images, under real-life and realistic acquisition conditions, is affected by extrinsic and identity-independent factors, such as variations in pose/viewpoint, illumination, facial expressions, time-lag, and occlusions (partial-data). In challenging, image "in the wild" benchmarks (Wolf et al., 2011), state-of-the-art performance depends on face pre-alignment, combining representations, or large-scale training. To alleviate extrinsic variability, increase the

* Corresponding author.
E-mail addresses: ikakadia@central.uh.edu, ioannisk@uh.edu (I.A. Kakadiaris), george.toderici@gmail.com (G. Toderici), gevang@mit.edu (G. Evangelopoulos), passalis@di.uoa.gr (G. Passalis), dattanchu@gmail.com (D. Chu), zhaoxi1@gmail.com (X. Zhao), shah@cs.uh.edu (S.K. Shah), theotheo@idi.ntnu.no (T. Theoharis).
[1] CBL, University of Houston.
[2] Google Inc., was with CBL, University of Houston when this work was performed.
[3] the Laboratory for Computational and Statistical Learning, MIT and Istituto Italiano di Tecnologia was with CBL, University of Houston when this work was performed.
[4] Accenture, was with CBL, University of Houston and University of Athens, Greece when this work was performed.
[5] CBL, University of Houston when this work was performed.
[6] School of Management, Xian Jiaotong University, 710049, Xian, P.R. China, was with CBL, University of Houston when this work was performed.
[7] IDI, NTNU, Norway, University of Athens, Greece, and CBL.

http://dx.doi.org/10.1016/j.cviu.2016.04.012
1077-3142/

discriminative ability, and boost the performance of conventional, image-based methods, alternative facial modalities, and sensing devices have been considered.

Three-dimensional recognition, from depth images, depth point clouds, or 3D meshes, has emerged as a distinct principle in biometrics and face recognition research (Abate et al., 2007; Bowyer et al., 2006), driven by improved 3D sensors, publicly available databases, and systematic evaluation benchmarks like the Face Recognition Grand Challenge (FRGC) (Phillips et al., 2005) and Face Recognition Vendor Test (FRVT) (Phillips et al., 2010). In these frameworks, which explored the feasibility of using 3D data both for enrollment and recognition, the 3D-based algorithms demonstrated a potential for very high recognition rates. For example, on FRGC v2.0, the 3D-3D face recognition system by Kakadiaris et al. (2007) reported a 97.5% rank-1 recognition and an average verification rate of 97.1% at 0.001 false acceptance rate (Ocegueda et al., 2011b), and the system of (Wang et al., 2010) 98.3% and 98.13%, respectively. Ocegueda et al. (2011a) achieved state-of-the-art performance both in FRGC v2.0, with 99% identification and 98% verification rate, and in the challenging 3D Twins Expression database (3DTEC) (Vijayan et al., 2011).

As an alternative, an asymmetric recognition system may involve 3D data for enrollment and 2D for verification or identification (3D-2D) or, the converse, 2D data for gallery and 3D data for probes (2D-3D). In the former, the need for 3D acquisition hardware is restricted to enrollment only and can facilitate the acquisition, storage, and distribution of high-quality databases of 3D models. In the latter, the abundance of existing face databases, composed primarily of 2D images, can provide reference enrollment sets for matching new 3D data. Independently, 3D model-based facial signatures are more discriminative and robust to condition variations. In this work, we propose a 3D-2D recognition framework which makes use of 3D data for enrollment, while requiring only 2D data for recognition, and which can be readily applied to the 2D-3D case also.

From the 3D gallery data, we build subject-specific, non-parametric 3D facial models by fitting a deformable Annotated Face Model (AFM) (Kakadiaris et al., 2007). The model surface parametrization defines a canonical 2D representation, the geometry image, that enables texture values assignment to corresponding 3D model points (Theoharis et al., 2008). A probe 2D image is mapped onto a subject-specific gallery model by explicitly accounting for relative pose and camera parameters using point-landmark correspondences (*pose estimation*). The estimated 3D-2D projection transformation is employed to generate pose-normalized texture images from the 2D image data and 3D model points (*texture lifting*). For matching, probe and gallery textures are lifted using the same 3D model. Their lighting conditions are further normalized using an illumination transfer method based on an analytical reflectance model (*texture relighting*). The final matching score between relit gallery and probe textures is a global similarity value obtained from low-level local orientation features.

Compared to asymmetric or heterogeneous recognition methods that map features across different modalities, the developed 3D-2D framework (termed UR2D) relies on a modality synergy, in which a 3D model is used for registration, alignment, and pose-light normalization of 2D image and texture data. Compared to previous approaches for 3D-2D registration and fitting (Gu and Kanade, 2006), UR2D employs the 3D shape information for relighting (using surface normal information) and score computation (extracting signatures in the geometry image space). Compared to existing multimodal 2D+3D methods (Jahanbin et al., 2011; Mian et al., 2007), UR2D integrates facial data across modalities and across enrollment/recognition phases in a subject-specific manner. In addition, unlike existing 3D-aided 2D recognition methods that use a 2D image to infer a 3D gallery model (Romdhani et al., 2006), UR2D is based on personalized gallery models constructed by fitting a model on the actual 3D facial data.

Our contributions can be summarized as follows: (i) we describe a conceptual framework for 3D-2D (or 2D-3D) face recognition; (ii) we propose a novel 3D-2D system for face image verification and identification from 3D datasets; (iii) we advocate the use of 2D+3D data to build subject-specific 3D gallery models that allow for personalized, texture-based similarity scores; (iv) we propose a method for model-based texture representation and a relighting algorithm for illumination normalization that improves recognition under lighting variations; and (v) we demonstrate empirically that 3D-2D recognition surpasses 2D-2D on challenging 2D+3D data with pose and illumination variations, and can approximate 3D-3D, shape-based similarity methods.

2. Related work

3D-2D and 3D-aided 2D face recognition: Recognition with 3D data spans an extensive body of work in 3D, 2D+3D (Bowyer et al., 2006), and 3D-aided 2D (Abate et al., 2007) FR. Rama et al. (2006) presented a method for simultaneous pose estimation and 2D face recognition that uses 3D data for training, though as a cylindrical texture image representation and not a full shape model. Riccio and Dugelay (2007) proposed using geometric invariants on the face to establish a correspondence between the 3D gallery face and a 2D probe image, disregarding the texture registered with the 3D data. Yin and Yourst (2003) used frontal and profile 2D images to construct models for 3D-based recognition. Blanz and Vetter (2003) introduced the 3D morphable model that captures face geometry and texture from 2D images. Using a statistical 3D face model and point correspondences, a gallery model is built from a 2D image. This is in principle different from our work, which uses real 2D and 3D data to build a 3D subject-specific gallery model. The morphable model framework was adopted for 3D-model-based 2D face recognition under illumination and pose variations (Romdhani et al., 2006), in fitting, synthesis, or normalization approaches (Zhang et al., 2014). Wang *et al.* extended it for a spherical harmonic representation (Wang et al., 2009). In contrast, methods for asymmetric 3D-2D FR learn a mapping between 3D and 2D data. Huang et al. (2010) map features extracted from gallery range images (2.5D) to 2D for 2D-matching of texture probe images. An extension, based on pose-light normalization and a mid-level representation, is reported to attain 95.4% recognition rate on FRGC v2.0 (Zhang et al., 2012).

3D face models from 2D images: In 2D-based methods, facial surface reconstruction from single or multiple images has been approached through stereo, structure-from-motion (Bregler et al., 2000), photometric-stereo (Georghiades et al., 2001), and shape-from-shading methods (Atick et al., 1996; Kemelmacher-Shlizerman and Basri, 2011) by estimating depth values from geometric, photometric, and gradient properties. Given a 3D prototype, reconstruction from images is obtained by fitting the model to 2D (i.e., estimating model parameters from image and geometric constraints) (Lee and Ranganath, 2003; Levine and Yua, 2009; Park et al., 2005). A model can be constructed from facial sample collections or prior knowledge on facial properties and physiology, and may be sparse to the number of points (i.e., point distribution models) or dense (i.e., vertex points with surface parametrization). Examples include shape-subspace projections, such as active appearance models (Matthews et al., 2007) and 3D morphable models (Blanz and Vetter, 2003), statistical deformable models (Kakadiaris et al., 2007; Mpiperis et al., 2008), elastic models (Prabhu et al., 2011), or reference samples (Kemelmacher-Shlizerman and Basri, 2011). Gu and Kanade (2006) fit a sparse set of surface points and the associated texture patches and simultaneously estimate deformation parameters and pose, requiring a

relatively large set of manual 3D-2D correspondences and training on synthetic 3D faces. Methods based on shape-subspace projections, like the morphable model (Blanz and Vetter, 2003; Levine and Yua, 2009; Romdhani et al., 2006), achieve accurate reconstructions from single images, though requiring large-sample training with point-wise aligned data and image-to-model point correspondences.

Illumination normalization: Smith and Hancock (2005) formulated an albedo estimation approach from 2D images based on the 3D morphable model. Wang et al. (2009) introduced neighboring coherence constraints on the model for albedo estimation and relighting using a subdivision-based random field minimization. Image co-registered 3D point clouds are used by Al-Osaimi et al. (2011) to estimate albedo by explicitly estimating the light parameters. The 2D-based albedo estimation by Biswas et al. (2009) does not handle specular light and assumes images without shadows. Self-shadowing problem is also a problem for the relighting approach by Lee et al. (2005). A nearest-subspace patch matching was used by Zhou et al. (2008) to warp a near-frontal to a frontal face and project it into a low-dimensional illumination subspace. The method requires training from patches in multiple illumination conditions. In this work, the proposed relighting method (Section 4.4) has significantly fewer constraints and limitations than many of the existing approaches and contributes towards the wider applicability of generic illumination normalization.

Face similarity scores: Scores for pairs of face images are measured based on either holistic, global representations, or local, part- or patch-based methods. Huang et al. (2011) extracted histograms of Local Binary Patterns features from 2D gallery, 2D probe, and 3D gallery range data. Scoring is done by fusing the 2D-2D histogram distance and the 3D-2D LBP mapping through canonical correlation analysis. Jahanbin et al. (2011) used local Gabor features on co-registered range and profile images, assuming highly accurate landmark localization. Tzimiropoulos et al. (2010) introduced a normalized gradient correlation coefficient for affine-illumination invariant image similarity, which is used for face texture matching in Toderici et al. (2010) (Section 4.5). An equivalent, cosine-based similarity measure formulation on the image gradient orientations (local edges) was employed for 2D-2D FR using subspace learning (Tzimiropoulos et al., 2012).

3. 3D-2D face recognition

The proposed framework for 3D-2D face recognition assigns a similarity score for one-to-one comparisons of 3D face data (i.e., surface points with triangulation and associated texture) to 2D images, containing a non-aligned, in general, face. In a practical scenario, the 3D+2D data form a gallery set $\mathcal{G} = \{(\mathcal{M}, I_i)\}_{i=1}^{n_g}$ of 3D shape (point cloud or polygonal mesh) $\mathcal{M}_i$ and possibly image (or mesh texture) I_i data for $n_g = |\mathcal{G}|$ subjects (also termed the enrollment or target set); the 2D data are samples from a probe set $\mathcal{P} = \{I_i\}_{i=1}^{n_p}$ of some size $n_p = |\mathcal{P}|$ (also termed the query or recognition set). Cross-modal similarity between 3D+2D and 2D data can be measured in the 3D domain, using an estimated or reconstructed shape model for each unknown probe image. Instead, we adopt here a 2D domain approach that constructs an appearance-based facial signature through gallery model-based registration and alignment of 2D image and 3D texture data (Procedure 1) .

The registration of 3D texture and 2D probe image under a given 3D shape is the main principle of our approach, by which 3D-2D face recognition can be viewed under two different perspectives: (i) using a 3D model to extract facial signatures from 2D images (irrespective of the model or image), and (ii) matching a 2D facial image to a specific subject face from 3D data (using the subject's model). More specifically, a 3D model of one gallery instance is used to map image values to 3D surface points and vice-versa. Through an implicit coordinate-system alignment, imposed by the surface representation of the model and an explicit, estimated 3D-2D projection, we achieve three types of data registration: local feature alignment, pose normalization, and illumination normalization. A similarity score is computed using the texture images for the 3D data (gallery texture) and the 3D-registered 2D data (probe texture). The resulting probe-gallery similarity values can be used for $1-1$, accept/reject-type verification decisions, or $1-N$ identification ranking. Given that the same model is used to generate both textures, the framework is individual-model based, in the sense that the signature for a face image is conditioned on the 3D data of an assumed identity (personalized model selection).

Procedure 1: Enrollment with 3D+2D data.

Data: 3D mesh, 2D image, 3D deformable model (AFM)
Result: Gallery metadata: fitted AFM, alignment transform, 3D landmarks, lifted texture, visibility map

1. Pre-process 3D mesh (smoothing, hole-filling) (Kakadiaris et al.,2007).
2. Annotate reference landmarks on 3D mesh.
3. Register and fit AFM to 3D mesh (Kakadiaris et al.,2007).
4. Lift texture from 2D image (or mesh texture) using the fitted AFM and estimated 3D-2D projection matrix.
5. Compute a 2D map of visible face regions.

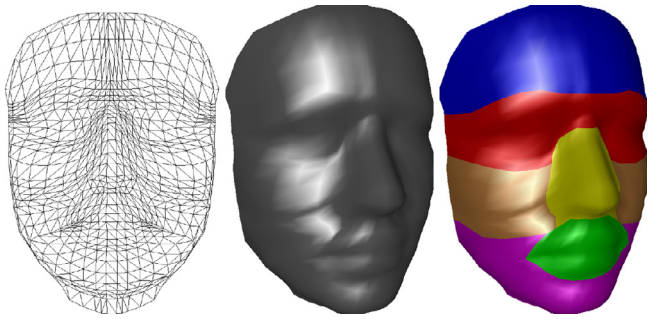

Fig. 1. Depiction of the 3D deformable face model as a wire-frame (left), surface (middle) and annotated surface (right).

3.1. 3D+2D enrollment

The enrollment process (i.e., the extraction of gallery metadata for each known subject) is depicted in Procedure 1 and Fig. 2. It is based on a 3D deformable AFM, originally proposed for 3D-3D face recognition (Kakadiaris et al., 2007) and employed also for identifying discriminative facial areas (Ocegueda et al., 2011b) and facial expression analysis (Fang et al., 2012). After pre-processing (smoothing, hole filling, and landmark annotation), the AFM is fitted to the 3D data through a subdivision-based deformation framework (Kakadiaris et al., 2007) and represented as a geometry image, i.e., a regularly-sampled, multi-channel image, using the model surface parametrization (Theoharis et al., 2008) (Figs. 2 (a) and (d)). The representation channels can include shape (vertex coordinates), normals, texture (the texel of closest 3D point), non-visible 3D points to the 2D sensor, and an annotation layer. On the 2D facial image, the corresponding facial landmarks are annotated and used to estimate a 3D-2D projection matrix from the fitted AFM to the image plane (Figs. 2(b) and (c)). Using the projection transformation, the facial texture is obtained from 2D image values and the non-visible facial regions are indicated by a binary map

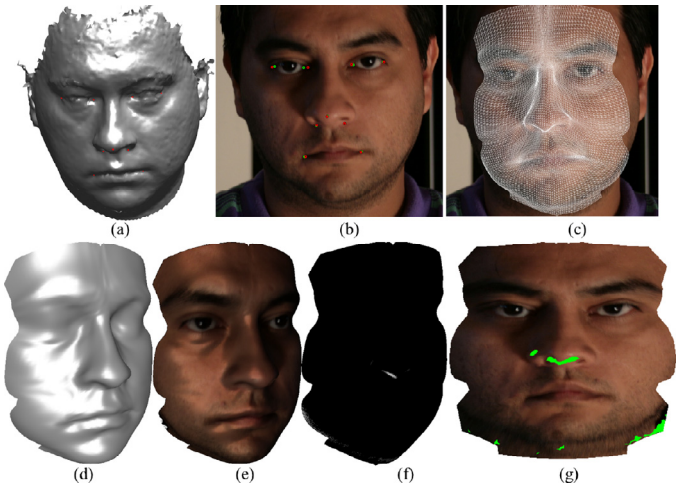

Fig. 2. Enrollment data and processing for 3D-2D face recognition: The subject's 3D and 2D data, (a) and (b), are fitted using a 3D deformable face model. The fitted mesh (d), texture (e), and visibility maps (f) are saved as a geometry image. Correspondences of landmark points, shown in both (a) and (b), are used for 3D-2D projection (c) and texture lifting (g) to form the subject's 2D facial signature (facial texture is overlaid with visibility map denoted in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

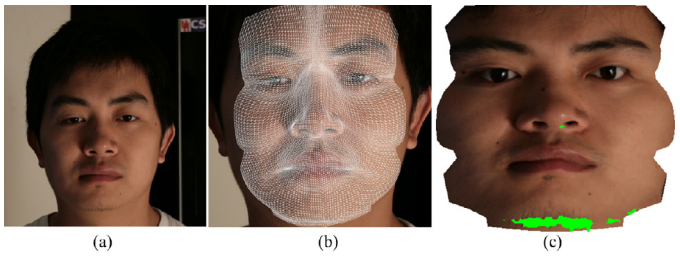

Fig. 3. Conversion of 2D images to textures in geometry image space (texture lifting): (a) raw 2D image data, (b) fitted AFM of same subject registered and superimposed on the image, (c) texture image with visibility map superimposed (invisible area denoted in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

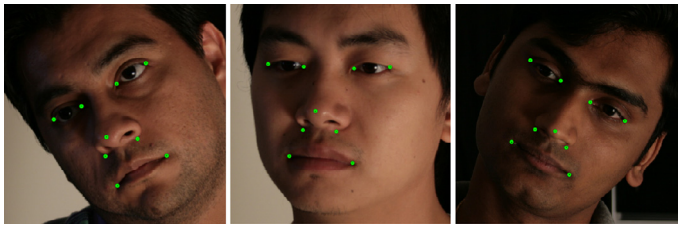

Fig. 4. Images of faces in different head poses with the nine reference points (facial landmarks) superimposed. The instances are probe images (2D data) from dataset UHDB11, used in the example depicted in Fig. 6.

(Figs. 2(e)–(g)). The fitted AFM, 3D and 2D annotations, facial texture, and visibility map are stored as gallery metadata.

3.2. 2D recognition

For verification or identification using an input 2D image, a small set of reference landmarks are localized on the facial region (some examples are shown in Fig. 4). The set is the same as the one the 3D landmark set of the gallery data. In this work, we use manual landmark labeling, which is equivalent to assuming highly accurate estimates of an automatic 2D landmark detector (Efraty et al., 2011; Sagonas et al., 2013), and chose nine facial points (mouth corners, four inner and outer eye corners, nose tip and nose inner corners). Using the 3D-2D point correspondences, a projection transformation is estimated using a full perspective projection model (Section 4.2). The projection parameters define a mapping between mesh and image points, which is used to assign texture values (Section 4.3) on the geometry image of a gallery model (Fig. 3). To match the illumination of the lifted probe texture (Fig. 3(c)), an analytical skin reflectance model is employed to relight the gallery texture using the fitted face model (Section 4.4). The pose- and light-normalized textures are assigned a similarity score using a global, correlation-based metric of local gradient orientations (Section 4.5) (Procedure 2 summarizes our 2D recognition pipeline).

Procedure 2: Recognition from 2D images.

Data: 2D image (probe), candidate model (gallery ID)
Result: Similarity score
1. Retrieve fitted AFM data.
2. Annotate reference landmarks on 2D image.
3. Register AFM to 2D image using 3D-2D landmark correspondences (pose estimation).
4. Lift texture from 2D image using 3D-2D projection of fitted AFM.
5. Compute a 2D map of visible face regions.
6. Relight the model texture to match the lifted 2D texture.
7. Compute gradient orientation correlation between relit gallery and probe texture.

4. Methods and modules for recognition

In this section, we describe in detail the methods for normalizing 2D images for pose and illumination and the modules of the proposed 3D-2D recognition system. **Notation**: We denote vectors or scalars by lower case x, matrices or columnwise arrangement of vectors by upper case X, vectorized matrices/images by bold upper case $\mathbf{X}$; a tilde accent $\tilde{x}$ denotes $\mathbb{R}^3$ vectors, $\tilde{x} := (x, y, z)^T$, or matrices, $\tilde{X}$, and no accent denotes $\mathbb{R}^2$ vectors, $x := (x, y)^T$; subscripts and superscripts denote instances x_i and dimensions x^i respectively. We use $I, T, \mathcal{M}$ for images, UV textures and 3D models or surfaces respectively; as argument in image functions, we use x for image space, e.g. $I(x)$, and $u := (u, v)^T$ in UV space, e.g., $T(u)$.

4.1. Texture images from 3D models

A framework for facial data representation and registration is provided by the use of models built from the AFM (depicted in Fig. 1) (Theoharis et al., 2008). The surface parametrization (or UV parametrization) of the model is an injective function

$$h: \tilde{x} \in \mathcal{M} \subset \mathbb{R}^3 \mapsto u(\tilde{x}) \in \mathcal{U} \subset \mathbb{R}^2 \quad (1)$$

that maps the original polygonal representation or surface $\mathcal{M}$ to a regularly-sampled, fixed-size 2D grid $\mathcal{U}$, termed the geometry image. As a result, any model fitted to 3D data inherits this predefined parametrization and the same geometry image grid. By associating texture values with model points, texture images are constructed on the UV coordinate space, which provides a common reference frame for local facial features. Texture images in the UV space are by construction aligned (due to being geometry images) and registered (due to the regularity of the grid) and can be compared using local similarity metrics.

For a 3D model with registered 2D texture data, image values at each UV coordinate are obtained from the texel corresponding to the closest 3D point. However, in practice, 2D and 3D data can be mis-aligned, the texture and surface acquisition sensors can involve time-lags, data might be collected across separate 3D and

2D sessions and the camera parameters may be unknown. In this work, we use the same method for automatic 3D-2D registration on both gallery and probe textures, or in other words on model-image pairs that have the same or different identities. In addition, our empirical observations showed that applying the same texture mapping method improves recognition, possibly due to introducing any estimation errors *symmetrically* on both textures.

4.2. 3D-2D registration and relative pose

Registration of a fitted 3D model to an image face is achieved through estimation of a perspective projection transformation, that involves a 3D rigid transformation (rotation and translation) and a 2D projection. In the most general case, both modalities may exhibit off-pose faces. During enrollment, the transformation can be explicitly obtained using known camera parameters and simultaneous acquisition of the 3D and 2D data. Recognition however, when a 2D probe needs to be registered to a 3D gallery model, can involve uncontrolled face rotations and different identities. We rectify the relative pose difference across modalities via explicit estimation of the projection for a given 3D-2D pair.

The 3D pose is implicitly accounted for through the rigid alignment of the raw data to the AFM, before model fitting (Kakadiaris et al., 2007). A similarity transformation A in $\mathbb{R}^3$ is estimated via the Iterative Closest Point algorithm so that the model points are aligned to the closest 3D surface points. Robustness to extreme poses can be gained through initialization of the algorithm via correspondence-based pre-alignment (Fang et al., 2012). Overall the 3D data are mapped to the canonical model pose as $A\tilde{x}$, $\tilde{x} \in \mathbb{R}^3$. After fitting, the deformed model vertices can be mapped to the original space via

$$\tilde{X} = A^{-1}\tilde{X}_{\mathrm{M}},\ \tilde{X}_{\mathrm{M}} = (\tilde{x}_1, \ldots, \tilde{x}_m),\ \tilde{x}_i \in \mathcal{M} \tag{2}$$

where $\tilde{X}_{\mathrm{M}}$ is the column arrangement matrix of points in homogeneous coordinates and A is a 4×4 matrix. This bijective mapping between the raw 3D data and the model pose allows the use of points in the 3D surface $\{\tilde{x}_i\}_{i=1}^{l} \in \mathbb{R}^3$ for establishing correspondences with 2D image points $\{x_i\}_{i=1}^{l} \in \mathcal{X} \subset \mathbb{R}^2$. The relative pose of the 3D to the 2D face is then obtained by estimating a perspective projection from the set of point correspondences $\{(x_i, \tilde{x}_i)\}_{i=1}^{l}$.

4.2.1. Perspective projection estimation

The underlying assumption is that 2D facial images were generated from viewing some facial mesh. Points from the 3D surface are mapped linearly on the 2D image plane, through a perspective projection compose of an extrinsic (or pose) component and an intrinsic (or camera) component, ignoring non-linear camera lens distortions. The linear map $P \in \mathbb{R}^{3\times4}$ of 3D to 2D, under a full *perspective projection* model is given by

$$x_i = P\tilde{x}_i,\ P = sKE,\ E = [\ \ R\ \ |\ \ \tilde{t}\ \ ] \tag{3}$$

for points $\tilde{x}_i \in \mathbb{R}^3$ that are mapped to points $x_i \in \mathcal{X} \subset \mathbb{R}^2$, both in homogeneous coordinates. K the 3×3 matrix of intrinsic parameters and E the 3×4 matrix of extrinsic parameters which can be further written with respect to a translation vector $\tilde{t}$, rotation matrix R and scale s, which is ambiguous for the perspective model.

Solving for P amounts to estimating, up to scale, the entries of the 3×4 projection matrix that maps a set of 3D points $\tilde{X}$ to the 2D image positions X through the linear transformation

$$X = sP\tilde{X},\ s \in \mathbb{R},\ P \in \mathbb{R}^{3\times4}, \tag{4}$$

where $X = (x_1, \ldots, x_l)$ and $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_l)$ are the column arrangement matrices of points in homogeneous coordinates. The system involves 11 degrees of freedom due to the ambiguity that any matrix sP, for an arbitrary scalar s, will result in the same set of projections on the image.

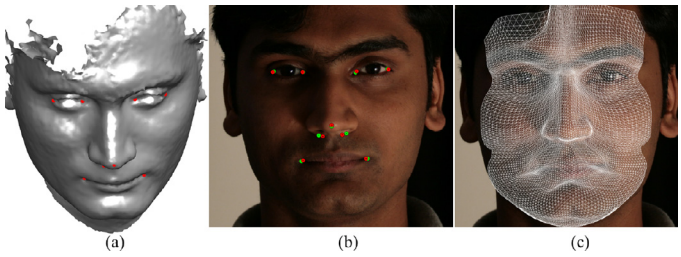

Fig. 5. Pose estimation and 3D-2D fitting of a 3D model: A set of 3D landmarks (in red), (a), are re-projected on the image plane, (b), so that the approximation error from the corresponding 2D landmarks (in green) is minimized. The estimated transformation is used for registering a 3D model to the image, depicted (c) as re-projection of the mesh. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In this paper, a set of 3D-2D point correspondences $\{(x_i, \tilde{x}_i)\}_{i=1}^{l}$ is established semantically through the localization of labeled facial landmarks in 2D and 3D. We use a reference set of $l = 9$ points (outer and inner eye corners, nose tip, nose inner corners, mouth corners), which results in an overdetermined system of approximate, within some small error, equality in (4), due to the uncertainty in landmark localization and different face identities. Example instances of the 2D, 3D, and 2D-reprojected reference points, using an estimated P, are shown in Figs. 2(a, b), 4 and 5. For evaluations in this paper, landmarks are manually annotated on both gallery and probe images. In practice, gallery landmarks can be accurately annotated in advance, while probe landmarks can be obtained during testing by a state-of-the-art facial point detector.

4.2.2. Landmark reprojection error minimization

The estimation of the projection transformation (4) is formulated as the minimization of the 3D-2D landmark reprojection error which is the error of approximating the 2D points by the projections of the 3D points through P. Using a square loss, P is estimated by solving a least-squares approximation problem over all reference points:

$$\min_{P} \sum_{i=1}^{l} \|x_i - P\tilde{x}_i\|_2^2 = \min_{P} \|X - P\tilde{X}\|_F^2. \tag{5}$$

The error in the approximation comes from a possibly overdetermined system and noise in the landmark locations. In our formulation, the objective function is parametrized with respect to the entries of the projection matrix P (i.e., the set of variables $\{(P)_j\}, j = 1, \ldots, 12$), and not individual camera and pose parameters. The minimization problem is solved using an iterative optimization procedure, in this case Levenberg–Marquadt (LM) algorithm (Lourakis, 2005), initialized using the Direct Linear Transformation algorithm (Hartley and Zisserman, 2004), that provides an efficient and close approximation to P given accurate point correspondences. For invariance to similarity transformations the point sets are normalized with respect to the location (origin) and average distance (scale) from their centroid.

The minimizer P is a coupled estimation of pose and camera parameters for the unknown, underlying setting corresponding to matching arbitrary 3D and 2D data, e.g., originating from different subjects or acquisition sessions. The estimate can be further decomposed (e.g., by a triangular-orthogonal RQ decomposition) to a matrix of intrinsic parameters K and the extrinsic matrix E that expresses the relative orientation to the camera frame (Eq. 3). However, for the purpose of obtaining the texture image using a 3D model, the projection matrix P is sufficient, as the decomposition in individual parameter matrices is in general non-unique.

The accuracy of the projection depends on the number, localization, and correspondence accuracy of 3D and 2D landmarks. For

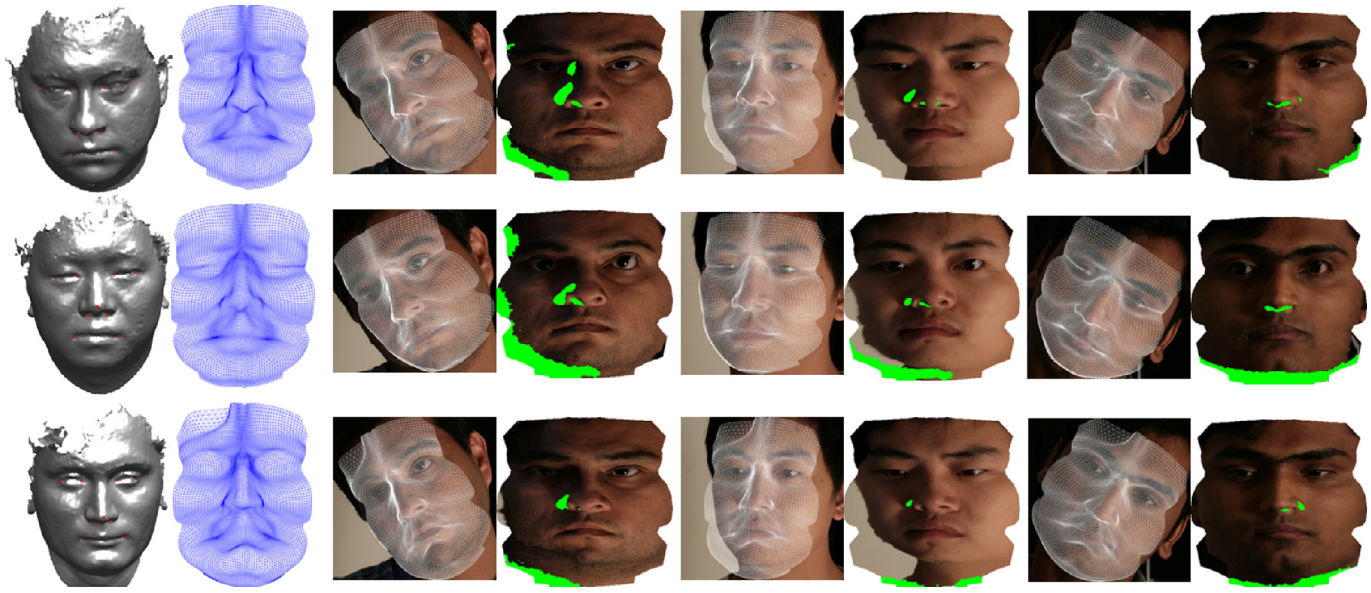

Fig. 6. Cross-subject texture lifting using personalized 3D gallery models and off-pose probe 2D images. This figure illustrates a 3 × 3 (*gallery vs. probe*) schematic comparison. The first two column pair depicts the 3D gallery data and models; the remaining column pairs depict the 3D model projected on the 2D image and the associated texture with the non-visible areas overlaid (denoted in green). Notice the artifacts and 3D-2D fitting errors for the cases where the gallery and the probe are of different identities (*off-diagonal instances*). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

registering a 3D model to images of the same subject, under unknown imaging conditions, the algorithm can handle a wide range of head pose variations. The iterative refinement of the initial approximate solution provides robustness to estimates from fewer correspondences; we noted that visually consistent textures were generated for nearly frontal poses when using five or four landmark points. This is particularly useful for handling self-occluded landmarks and large localization errors by discarding inconsistent points. Examples of the estimated projections are depicted in Fig. 5 and Fig. 6 by visualizing the 3D model points and mesh projected on the 2D image.

4.3. Texture lifting and visibility maps

Given a 3D model and 2D image pair $(\mathcal{M}, I)$ and the estimated projection transformation P, the texture image $T(u), u \in \mathcal{U}$ is generated or *lifted* from image values $I(x), x \in \mathcal{X}$ and the registered model vertices $\{\tilde{x}_i\}_{i=1}^{m} \in \mathcal{M}$. The process is analogous to extracting the UV map from a texture image co-registered to the model: through the UV parametrization, texture values in the geometry image are assigned from the image values on the locations of the reprojected model vertices.

The projections of model points are obtained from the cascade of two transformations; the 3D transformation A in Eq. (2) that maps the deformed model to the 3D data coordinate system and the perspective projection P from minimizing Eq. (5):

$$X = PA^{-1}\tilde{X}_{\mathrm{M}},\ X = (x_1, \ldots, x_m),\ \tilde{X}_{\mathrm{M}} = (\tilde{x}_1, \ldots, \tilde{x}_m) \tag{6}$$

with X and $\tilde{X}_{\mathrm{M}}$ the matrices of the model vertices in the image and 3D space respectively. At the UV coordinates corresponding to some model point $\tilde{x} \in \mathcal{M}$, the value for T is obtained using image I values in $x \in \mathcal{X}$:

$$T(u) = I(x), \forall (u, x) \in \mathcal{U} \times \mathcal{X} : u = h(\tilde{x}),\ x = PA^{-1}\tilde{x} \tag{7}$$

with $h: \mathcal{M} \to \mathcal{U}$ the model parametrization in Eq. (1). Values of T for u not corresponding to some $\tilde{x} \in \mathcal{M}$, are assigned through value interpolation from the projected model triangulation.

For a model-image pair, two types of self-occlusions can affect the generated texture, due to 3D pose and 2D pose: model surface occlusions along the camera viewpoint direction (non-visible 3D triangles) and region occlusions on the 2D image plane (non-visible face regions). In occluded facial areas, the reprojected mesh is not area injective and the mapping of the surface triangulation to image regions will result in overlapping 2D triangles. As a result, the same image values will be assigned to texture points corresponding to the visible and occluded triangle areas.

Pseudo-value points due to occlusions are excluded from subsequent processing by computing *visibility maps*, i.e., indicator functions of the hidden UV regions under the estimated projection transformation. To determine 2D visibility, a depth-buffering method is applied by keeping track of the depth values for projected points on the image plane. The value from the target image at x is uniquely assigned, by Eq. (7), to the point u that corresponds to the 3D point $\tilde{x}$ with the smallest depth value (closest point to the camera). The visibility map is then the indicator function:

$$V(u(\tilde{x}_i)) = 1\ \forall \tilde{x}_i, \tilde{x}_j \in \mathcal{M} : x_i = x_j \text{ and } (\tilde{x}_i - \tilde{x}_j) \cdot \tilde{e} > 0, \tag{8}$$

where x_i, x_j are the reprojections given by $x = PA^{-1}\tilde{x}$ in Eq. (6) and $\tilde{e} = (0, 0, 1)^T$ is the depth coordinate unit vector in $\mathbb{R}^3$. Conversely, the non-visible map is the indicator function of the points u that compete for the same image values with others of smaller depth. An additional 3D visibility map can be estimated by excluding in UV the 3D points with transformed surface normals $\tilde{n} = \tilde{n}(\tilde{x}_i) \in \mathbb{R}^3$ of direction opposite to the viewpoint: $V(u(\tilde{x}_i)) = 1$ if $E\tilde{n}(\tilde{x}_i) \cdot \tilde{e} = \tilde{e}^T E\tilde{n} \geq 0$, where E is the extrinsic matrix that needs to be computed from the 3D-2D projection in Eq. (3).

4.4. Bidirectional relighting

To normalize the illumination conditions between a pair of textures we apply illumination transfer via optimization without explicit albedo estimation. The proposed relighting algorithm operates on textures represented in the AFM's UV space, minimizing their element-wise illumination difference. We use an analytical skin reflectance model (ASRM) in the form of a hybrid bidirectional reflectance distribution function (BRDF).

Overall, the method has significantly fewer constraints and limitations than existing approaches: (i) no assumptions are made on the source facial image (i.e., viewpoint, registration), since the UV

Fig. 7. Texture images for the gallery set of UHDB11 database, extracted using same-subject 3D-2D projection estimation. Non-visible facial regions are superimposed (in green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

representation is inherited through 3D-2D registration; (ii) no assumptions are made on the light sources (number, distance, or direction) (Smith and Hancock, 2005), the presence of shadows (Biswas et al., 2009) or specularities; (iii) relies on minimal input and no light calibration information (i.e., the pair of probe-gallery texture images and the 3D fitted model); (iv) no training for learning global (Biswas et al., 2009; Georghiades et al., 2001), or local (Zhou et al., 2008) light variations from different examples is required; (v) it involves a single optimization step, minimizing a joint error instead of solving two separate inverse problems to independently estimate the albedo of each texture; (vi) it is bidirectional (i.e., the roles of source and target textures can be interchangeable); and (vii) it employs the same 3D model, used for pose normalization and texture lifting, for obtaining estimates of the surface normals.

4.4.1. Texture lighting model

A texture image $T(u)$, $u \in \mathcal{U} \subset \mathbb{R}^2$ in the UV space is the result of lighting applied on the unknown face albedo $B(u)$:

$$T(u) = L_s(u) + (L_d(u) + L_a(u))B(u), \tag{9}$$

where $L_d(u)$, $L_s(u)$ are the diffuse and specular reflectance components (assuming white specular highlights) and $L_a(u)$ is the ambient illumination component. The illumination conditions between a pair of texture images can be normalized following two different objectives: estimating the albedo components (unlighting) (Zhao et al., 2014; 2012; 2013) or transferring illumination parameters (relighting) (Toderici et al., 2010). Solving Eq. (9) for unlighting $B(u) = (T(u) - L_s(u))/(L_d(u) + L_a(u))$ requires estimation of the lighting components. For relighting, and many practical applications, the albedo itself is not required, and is used only as an intermediate estimate. In this work, we advocate the use of texture relighting without a-priori estimating the albedo.

4.4.2. Analytical skin reflectance model

We use an analytical BRDF for the diffuse and specular reflectance components, ignoring subsurface scattering. Diffuse reflectance L_d is modeled using the basic Lambertian BRDF, which assumes an equally bright surface across all directions. For a single light source with intensity L, the intensity at a surface point $\tilde{x}(u)$ is proportional to the angle $\theta(u)$ between surface normal and incident light directions: $L_d(u) = L\cos(\theta(u))$. Specular reflections are accounted for through the Phong BRDF, which models the intensity of the specular reflection at a surface point by $L_s(u) = L\cos^{\eta}(\phi(u))$, where $\phi(u)$ is the angle between the view vector and the reflected light and η is a parameter that controls the size of the highlight. The coarse variation of specular properties across different facial areas is incorporated through a specular map based on the annotation of the AFM.

The model in Eq. (9) can be written with respect to the parameters of the analytic models. Textures are modeled in independent channels in the RGB or Hue-Saturation-Intensity (HSI) space and light color is assumed to be perfectly white, a reasonable approximation for facial images acquired in indoor or controlled conditions. In addition, multiple point light sources are aggregated by summing their individual BRFD functions:

$$\begin{aligned} T(u) &= L_a(u)B(u) + L_d(u)B(u) + L_s(u) \\ &= L_aB(u) + \sum_{i=1}^{\ell} (L_i\cos(\theta_i(u))B(u) + L_i\cos^{\eta}(\phi_i(u))) \end{aligned} \tag{10}$$

4.4.3. Optimization for illumination transfer

The parameters of the lighting model are the locations of multiple light sources $\{\tilde{c}_i\}$ and the parameters for the reflectance components, for example $\{L_a, \{L_i\}_{i=1}^{\ell}\}$, per light and color channel. To minimize the disparity of light conditions between two textures, we formulate an optimization problem for two sets of lights (located on a sphere around the model centroid); one that *removes* the illumination from a source texture and one that *adds* the illumination of a target texture. The scheme is based on minimizing the approximation error between textures of the same albedo but different light conditions:

$$R_c(u) = \left(T'(u) - L'_s(u) - (T(u) - L_s(u))\frac{L'_d(u) + L'_a(u)}{L_d(u) + L_a(u)}\right)^2, \tag{11}$$

$$\min_{\lambda} \sum_{u\in\mathcal{U}} R_c(u;\lambda), \quad \lambda = \{\{\tilde{c}_i, L_i\}_1^{\ell}, \{\tilde{c}'_i, L'_i\}_1^{\ell}\} \tag{12}$$

where L_a, L_d, and L_s are the reflectance components of the source texture T, given in Eq. (10), and L'_a, L'_d and L'_s the reflectance components of the target texture T'; the minimization is defined with respect to the composite vector λ of light parameters and positions for both T and T'. The error function can also be interpreted as the average intensity variation in the face albedo $B(u)$ in the two textures. In practice, the error function is the Euclidean norm of the color-vector valued Eq. (11) over RGB channels, and after applying the union of the visibility maps $V(u)$, $V'(u)$ of Eq. (8):

$$R(u) = \sum_{u\in\mathcal{U}} V(u)V'(u)\left(\sum_{c\in\{\mathrm{R,G,B}\}} R_c^2(u)\right)^{1/2} \tag{13}$$

A minimizer of Eq. (11) can be sought through global optimization methods. We use simulated annealing with an adaptive exponential annealing scheme. We consider multiple light points, three for simulating the illuminance on the source and three for the illuminance on the target texture. For improved performance under low lighting conditions, color textures are converted to HSI color space, and the RGB cost in Eq. (13) is replaced by a weighted average, with intensity weighted twice as much as hue and saturation. We observed empirically that such a scheme improves the synthesized appearance on dark facial regions and increases the similarity score. In this paper, the gallery is used as the source and the probe as the target texture, both in their UV representations under the same face model. An example is illustrated in Fig. 8, where

Fig. 8. Optimization for relighting gallery T of Fig. 3(c) using different probe textures (rows). Left-to-right: probe texture T', texture difference between probe-gallery (before optimization), gallery with illumination of probe, texture difference (after optimization). Textures are depicted in the geometry image space.

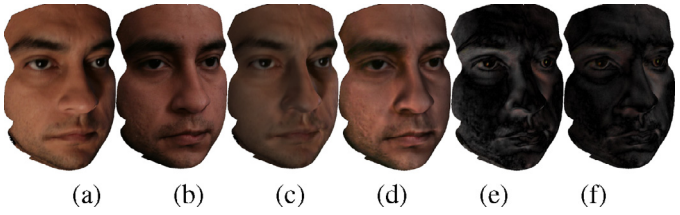

Fig. 9. Bidirectional relighting (left-to-right): (a) real texture 1 (RT1); (b) real texture 2 (RT2); (c) synthetic texture 1 (ST1): RT1 with RT2's illumination; (d) synthetic texture 2 (ST2): RT2 with RT1's illumination; (e) RT1-ST2 residual; (f) RT2-ST1 residual. Textures are visualized superimposed on the fitted 3D model.

the gallery texture of Fig. 3(c) is changed to reflect the illumination conditions of different probe textures of the same subject.

An additional property of the proposed scheme is that relighting is bidirectional, in the sense that illumination transfer can take place in any direction between source-target, and the roles of probe and gallery textures can be interchangeable in the cost function. This is illustrated in Fig. 9 for same-subject textures, with (a) having a corresponding 3D mesh and (b) being lifted from an image using the fitted model of (a). The textures with illumination conditions transfered by relighting in both directions are shown in (c) and (d), and the differences of the synthetic textures from the respective targets are depicted in (e) and (f). Closer visual inspection shows that the relit textures look natural, facial features are preserved and no significant artifacts are introduced by the relighting process.

4.5. Face signature and similarity score

The pairwise comparison of probe and gallery lifted textures is obtained through a global similarity score, based on a correlation coefficient of image gradient orientations, which is largely insensitive to serious mismatches in parts of the two images (Tzimiropoulos et al., 2010; 2012). It is particularly suitable for measuring the similarity of face data that may vary substantially, not only due to different acquisition conditions, but also due to significant variability in individual appearance. As a feature, gradient orientations can be relatively robust to non-uniform illumination variations (Osadchy et al., 2007). Note that textures in the geometry image space, are already normalized with respect to pose variability through pose estimation and registration using the same 3D model and with respect to lighting conditions variability through relighting.

For a texture image $T(u), u \in \mathcal{U} \subset \mathbb{R}^2$, we compute the complex gradient image $G(u) = \nabla_x T(u) + j\nabla_y T(u)$, where $\nabla_x T$, $\nabla_y T$ the gradients along the two spatial dimensions. After suppressing small gradient values and normalizing, we obtain the orientation maps $O(u) = G(u)/\|G(u)\|_2 = \exp(j\Theta(u))$, where $\Theta(u)$ is the orientation of the gradient vector. The similarity score between two textures T_p, T_q is based on the orientation correlation, or more explicitly, the real part of the correlation of the orientation images:

$$s(T_p, T_q) = \mathrm{Re}\left(\int_{\mathcal{U}} O_p(u) O_q^*(u) du\right) = \int_{\mathcal{U}} \cos(\Delta\Theta(u)) du, \tag{14}$$

which is equivalent to integrating the cosine of the gradient phase difference $\Delta\Theta(u) = \Theta_p(u) - \Theta_q(u)$ (Tzimiropoulos et al., 2010). As argued in Tzimiropoulos et al. (2012), $\Delta\Theta(u)$ can be assumed uniformly distributed in $[0, 2\pi)$ for regions where the two images do not match, leading to zero contribution to the integral in Eq. (14). As a result, the value of s will be specified solely from the texture regions that match. One can see that the corresponding texture distance can be derived through the L_2 norm of the orientation image difference

$$d(T_p, T_q) = \|\exp(j\Theta_p(u)) - \exp(j\Theta_q(u))\|_2^2 = \int_{\mathcal{U}} du - s(T_p, T_q). \tag{15}$$

The similarity $s : \mathbb{R}^n \times \mathbb{R}^n \to [-1, 1]$ and distance $d : \mathbb{R}^n \times \mathbb{R}^n \to [0, 2]$ functions for finite-dimension textures, written as n-dimensional vectors $\mathbf{T}_i, \mathbf{\Theta}_i \in \mathbb{R}^n, i = p, q$ are then:

$$s(T_p, T_q) = \frac{1}{n}\sum_{j=1}^{n} \cos(\Delta\mathbf{\Theta}^j), \quad d(T_p, T_q) = 1 - s(T_p, T_q) \tag{16}$$

Interestingly, Eq. (16) can be viewed as a (positive-definite) kernel function $k(\cdot,\cdot)$ on the orientation vectors $\mathbf{\Theta}$, defined via the $2n$-dimensional feature map $\Phi(\mathbf{\Theta}) = n^{-1/2}(\cos(\mathbf{\Theta}^1), \ldots \cos(\mathbf{\Theta}^n), \sin(\mathbf{\Theta}^1), \ldots, \sin(\mathbf{\Theta}^n)^T \in \mathbb{R}^{2n}$. The similarity function can then be written as $s(T_p, T_q) = k(\mathbf{\Theta}_p, \mathbf{\Theta}_q) = \langle\Phi(\mathbf{\Theta}_p), \Phi(\mathbf{\Theta}_q)\rangle$. The distance $d(T_p, T_q)$ defines a proper dissimilarity measure for T_p, T_q, satisfying *reflectivity* (i.e., $d(T_p, T_p) = 0$), *positivity* (i.e., $d(T_p, T_q) > 0$ if $T_p \neq T_q$) and *symmetry* (i.e., $d(T_p, T_q) = d(T_q, T_p)$).

The gradients are computed using two-point, forward-backward finite differences from the grayscale, luminance images. To exclude large-pose, self-occlusion artifacts from the score, we use the union of visibility maps obtained during texture lifting (Section 4.3) to indicate the valid texture points. The score can then be computed either from the distance of unit-norm gradient images (15) or the kernel function on the gradient orientations (16).

4.6. Score normalization

Different probes can have different ranges of score values when mapped on the gallery textures, resulting in non-comparable similarity and an underestimate of the true identification or verification performance of a multi-probe similarity matrix. Typically, similarity scores per gallery are normalized with respect to scale, range or distribution characteristics prior to predicting system performance. For our evaluations, results are reported with scores normalized using standard Z-score (Jain et al., 2005) for 1-N normalization and a metric multidimensional scaling method for N-N

Fig. 10. Sample images (2D data) from probe set of 3D-2D database UHDB11 (Toderici et al., 2013) depicting variations in illumination conditions and pose.

normalization using distances extracted from the gallery data (termed *E-normalization* in Toderici et al. (2010)).

Given a set of textures $\mathcal{T}$, the dissimilarity matrix is a matrix $D \in \mathbb{R}^{n\times n}$ such that $D_{ij} = d(T_i, T_j)$ for $T_1, \ldots, T_n \in \mathcal{T}$ and the distance function defined in (16). Let $\mathcal{G} = \{T_i(\mathbf{x})\}$ and $\mathcal{P} = \{T_j(\mathbf{x})\}$ be the sets of gallery and probe textures respectively. Then the gallery-gallery and gallery-probe matrices are:

$$D_G \in [0,2]^{n_g\times n_g} : (D_G)_{ij} = d(T_i, T_j), \quad \forall T_i, T_j \in \mathcal{G}$$
$$D_P \in [0,2]^{n_g\times n_p} : (D_P)_{ij} = d(T_i, T_j), \quad \forall T_i \in \mathcal{G}, T_j \in \mathcal{P}$$

The normalized scores are obtained through linear projection of the gallery-probe dissimilarity matrix D_P on a Euclidean (or pseudo-Euclidean) space obtained via an isometric embedding of D_G (Pekalska et al., 2002). The computation involves the Eigen-decomposition of D_G and the selection of the size of the embedding space (number of eigenvectors). Reducing the dimensionality ($\leq |\mathcal{G}|$) can result to increased robustness to noisy distances.

5. Results and performance evaluation

5.1. 3D-2D database with pose and illumination

Database UHDB11 is a challenging set of 3D and 2D data, designed for asymmetric 3D-2D face recognition benchmarks under pose and illumination variations (Toderici et al., 2013). It consists of high-resolution, systematically acquired 3D gallery and 2D probe image data from 23 subjects, of both genders and multiple ethnicities, with 72 pose/illumination variations per subject; 12 pose (3 roll × 4 yaw angles) and 6 diffuse-light illumination conditions. The UHDB11 3D-2D experimental configuration is designed to be a closed-set scenario, with 23 frontal pose, same illumination gallery datasets (3D mesh + 2D texture, 2D image) and a disjoint set of 1,602 probe 2D images. A sample from the gallery is shown in Fig. 2(a and b), lifted textures for the gallery set in Fig. 7, and sample images from a single subject of the probe set across different illumination conditions in Fig. 10.

We choose to rely primarily on UHDB11 for evaluations over other public 3D face databases because it provides greater and controlled face pose variations, jointly with illumination changes; existing 3D databases are limiting due to the availability of either frontal only face images (FRGC v2.0 (Phillips et al., 2005)) or uniform illumination conditions (Bosphorus Savran et al., 2008). Though UHDB11 dataset covers fewer subjects, compared to these databases, it provides more instances of 2D faces with varying pose and illumination for each subject. The proposed method is dataset-independent, since it is not trained or developed specifically for UHDB11. To demonstrate this, we present face recognition results on a 3D-2D Cohort from 250 subjects of FRGC v2.0 forming a 250 × 470 set (Al-Osaimi et al., 2011) (Section 5.9). Samples from the gallery and probe images are shown in Fig. 11.

Fig. 11. Sample images (3D co-registered textures) from the FRGC v2.0 3D-2D cohort used in Al-Osaimi et al. (2011); Zhao et al. (2014) for the gallery (top) and probe (bottom) sets.

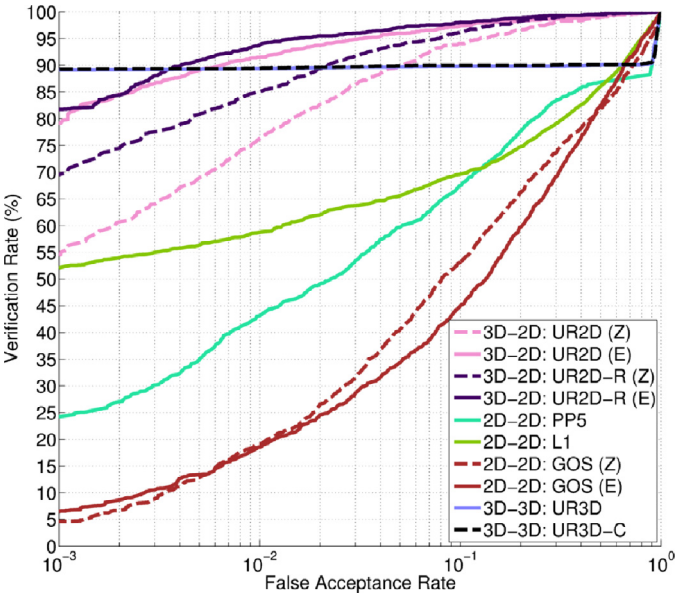

Fig. 12. ROC curves for face verification on UHDB11 using 2D-2D, 3D-3D and the proposed 3D-2D framework.

5.2. Recognition performance metrics

For performance evaluation, we consider a realistic 3D-2D recognition scenario, in which 3D+2D data are acquired during enrollment, forming a gallery set, and 2D data during a verification (one-to-one matching) or identification (one-to-many matching) phase. Following standard protocols for biometric systems, performance is evaluated using receiver operating characteristics (ROC) curves that report verification rate (VR) at varying false acceptance rate (FAR) (Fig. 12), for specified decision distance thresholds, and cumulative match characteristics (CMC) curves that report rank-k recognition rate (RRk) (Fig. 13). For quantitative comparisons among different baselines we use curve-extracted measures; for verification, VR at FAR = 0.001 and FAR = 0.01, equal error rate (EER), i.e., point where false alarm equals false reject rate and area under the ROC curve (AUC); for identification, rank-1 recognition performance. A comparative overview of 3D-2D face recognition results on UHDB11 is provided in Table 1.

5.3. Baseline algorithms

Comparisons were carried out with two goals: (i) to verify the effectiveness of the pose and illumination normalization of UR2D for face recognition and (ii) to support a proof-of-concept for 3D-2D face recognition. We postulate that 3D-2D recognition surpasses 2D-2D performance on challenging sets with 2D+3D data, and can approximate 3D-3D, shape-based similarity methods. Further, we advocate that the texture representation and normalization through the use of 3D models can compensate for dissimilarities due to non-frontal poses or different lighting conditions.

Table 1
Face Recognition on UHDB11 using 3D-3D, 2D-2D and 3D-2D systems (UR2D: raw texture comparison, UR2D-R: with light normalization, (Z): Z-score normalized, (E): E-normalized with multidimensional scaling).

Method	Type	Verification				RR1(%)
		VR@10^{-3}	VR@10^{-2}	EER	AUC	
PP5 (Pittsburgh Pattern Recognition, 2011)	2D-2D	0.242	0.431	0.211	0.826	68.2
L1 (L1 Identity Solutions)	2D-2D	0.519	0.588	0.230	0.845	62.6
GOS(Z)	2D-2D	0.046	0.191	0.279	0.778	32.9
GOS(E)	2D-2D	0.064	0.186	0.305	0.764	29.1
UR2D(Z)	**3D-2D**	0.544	0.763	0.072	0.975	77.0
UR2D-R(Z)	**3D-2D**	0.695	0.851	0.056	0.986	85.6
UR2D(E)	**3D-2D**	0.793	0.915	0.043	0.988	89.9
UR2D-R(E)	**3D-2D**	0.817	**0.939**	**0.037**	**0.991**	**90.8**
UR3D (Kakadiaris et al., 2007)	3D-3D	0.890	0.894	0.108	0.905	90.2
UR3D-C (Ocegueda et al., 2011a)	3D-3D	**0.892**	0.895	0.101	0.906	90.1

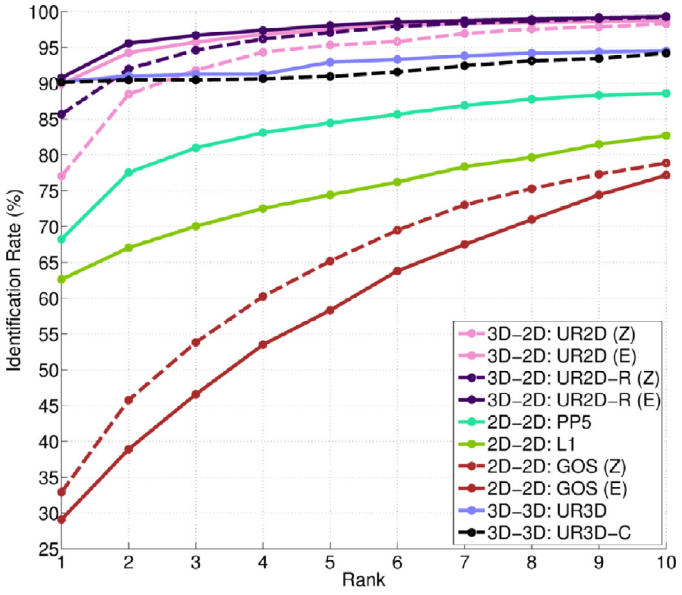

Fig. 13. CMC curves for face identification on UHDB11 using 2D-2D, 3D-3D and the proposed 3D-2D framework.

3D-2D: We present results from two variations of the proposed method, one with the bidirectional relighting module (UR2D-R) and one without (UR2D). The former corresponds to computing the orientation correlation similarity metric between the relighted-to-probe gallery texture, using the relighting algorithm of Section 4.4, and the raw probe texture. The latter amounts to comparing the raw lifted textures without light normalization and will disassociate the effect in performance due to the light normalization module. The resulting similarity matrices from each system are Z-score and E-normalized (Section 4.6), retaining all embedding dimensions ($d = 23$), denoted by (Z) and (E) respectively. For computing scores, only the depth-buffering visibility was used for area masking, since the used 3D gallery poses are frontal.

2D-2D: We consider three baselines for comparing the 2D images for probe and gallery; L1 IdentityToolsSDK L1 Identity Solutions system for FR denoted by (L1), PittPattSDK v5.2 (Pittsburgh Pattern Recognition, 2011) for FR, denoted by (PP5), and computing the gradient orientation similarity of Eq. (14), denoted by (GOS). Since the first two are closed-source, commercial systems, details on the recognition pipelines and similarity methods are not available. For GOS, facial area detection and geometric image alignment were performed using the available 2D landmarks. A facial region-of-interest was defined by the area of a rectangle circumscribing the set of points with an additional fixed margin. The resulting faces were aligned using the point correspondences and warped by the estimated similarity transform.

3D-3D: As a 3D-3D performance baseline, we rely on the original UR3D (Kakadiaris et al., 2007) and the UR3D-C framework (Ocegueda et al., 2011a) that uses compact signatures via dimensionality reduction. Both are based on the same 3D model-fitting front-end with the 3D-2D system, through extracting shape-based signatures (no texture information is used). The configuration involved the 3D gallery datasets and the corresponding 3D of the probe images (available from the data acquisition process), which are not formally part of the 3D-2D UHD11 database.

5.4. 3D-2D verification

The goal of the full 23 × 1,602 verification experiment on UHDB11 is to validate our hypothesis that the proposed framework and system for 3D-2D recognition will significantly improve over 2D-2D recognition and be closer to 3D-3D recognition performance on this challenging set. The ROC curves for different UR2D and baseline algorithms are shown of Fig. 12. The best performance for 3D-2D is achieved using the relighting module and E-normalized scores, reaching an 81.7% VR at 0.001 FAR and 93.9% at 0.01 FAR. This is 29.8% more, at 0.001 FAR, than the best of the 2D-2D methods on the 2D datasets (L1). Moreover, this is 7.5% less at 0.001 FAR than the 3D-3D system on the 3D datasets; however it is 4.4% above 3D-3D at 0.01 FAR. This is in accordance with our hypothesis. Table 1 summarizes the experimental results.

The explicit pose estimation and pose normalization, enforced by the 3D model fitting and texture lifting in UR2D, overcome many of the challenging off-pose instances in the database probes. Compared to both tested commercial, reference 2D-2D systems, the 3D-2D framework is found to perform superiorly. This is indicated by the high AUC values (above 0.97 in every case), and low EER (below 0.07 in all cases), reaching an optimal performance of EER equals 0.037 for UR2D-R (E).

An evaluation of the relative significance of the different modules in UR2D (e.g., similarity score, 3D-2D fitting or pose normalization, illumination normalization) is conducted by comparing the curves for the GOS, UR2D, and UR2D-R methods. Without pose and illumination normalization, VR performance drops significantly as evidenced by the low VR performance obtained using 2D similarity score on detected and rigidly aligned faces (GOS). Even with E-normalization that employs gallery-to-gallery similarity scores embedding (GOS (E)), the distance of the match/non-match distributions of scores does not improve. This highlights the bound in the performance of the gradient orientation similarity score under off-pose faces, and that feature correlation-based metrics are not sufficient in themselves in the presence of inconsistent acquisition

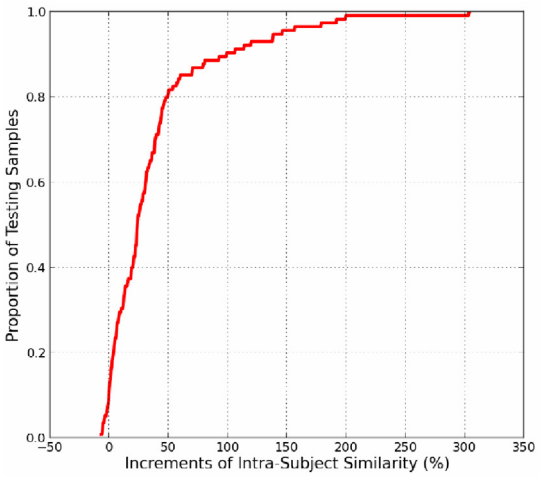

Fig. 14. Increments of intra-subject similarities after illumination normalization through bidirectional relighting.

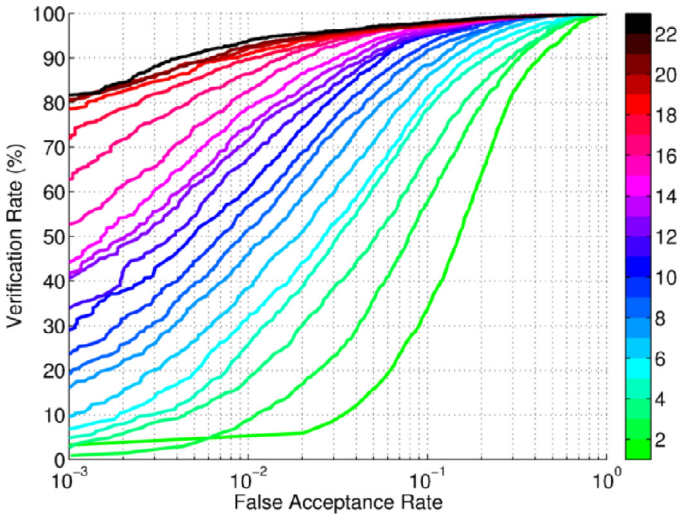

Fig. 15. Effect of changing the normalization embedding dimension ($d \in [2, 23]$, color-coded) on UR2D-R (E) verification. The UHDB11 gallery defines an embedding space of 23 dimensions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

conditions and uncontrolled environments. The increase in performance is dramatic when the same metric is applied in geometry image space to the textures lifted using the 3D fitted model and pose normalization, reaching 72.9% VR with (E) and 49.8% with (Z) at 0.001 FAR.

Another key observation is that illumination normalization is aiding 3D-2D FR, as the proposed bidirectional relighting method improves verification rate by 15.1% at 0.001 FAR. Improvement is more obvious at low FAR rates, where the limits of performance using raw texture images are reached. To evaluate the effect of relighting, we computed the similarity scores of 115 pairs of textures before and after illumination normalization. Each pair corresponds to the same, single subject and the same pose but different illumination conditions. We quantify the effectiveness of the illumination normalization by the relative increment $(s_{\text{rel}} - s)/s$ of the similarity scores $s(T_p, T_q)$ after relighting $s_{\text{rel}} = s(R_q(T_p), T_q)$. The results in Fig. 14 support that illumination normalization improves the similarity scores of over 90% of the pair-wise comparisons and half of them achieve over 25% improvement.

Notably, when using E-normalized scores, the merit of 3D-2D with relighting (UR2D-R) over raw matching (UR2D) is less evident. This cannot be attributed to representation in a subspace of reduced-dimensionality, as all dimensions of the embedding space were retained for normalization. Fig. 15 depicts the effect of varying the dimension of the normalizing subspace on the verification performance of UR2D-R.

The 3D-3D framework performs consistently above 89%, surpassing 3D-2D at low FARs (0.001). However, for FAR above 0.004 (for UR2D-R), and 0.003 (for UR2D), 3D-2D outperforms 3D-3D. This surprising result, noted also on the higher AUC and lower EER values for 3D-2D, may be due to the discriminative difference of the similarity metrics for shape or texture on the specific dataset, tailored to 3D-2D experiments and high-resolution 2D data, and that UR3D does not adapt to or compensate for posed 3D data.

5.5. 3D-2D identification

The CMC curves for the 23 × 1,602 identification experiment on the 23 identities of UHDB11 can be seen in Fig. 13. The performance of different systems is shown as rank 1–10 identification rate. Consistent with the verification task, UR2D outperforms the two commercial 2D-2D systems across all rank tests, attaining a 22.6% and 17.4% improvement in rank-1 recognition over the best of them (PP5), using (E) and (Z) normalization schemes respectively. Compared to 3D-3D, the identification rate of UR2D-R at rank-1 is similar (−4.6%, +0.6% for (Z) and (E) respectively), and consistently better for ranks higher than one.

Table 2
3D-2D recognition on UHDB11 with automatic landmark.

Landmarks	AUC	VR@10^{-3}FAR	RR1 (%)
Manual	0.991	0.817	90.8
Automatic	0.965	0.513	76.8
Perturbed (45 pix.)	0.968	0.500	76.7

5.6. Recognition with landmark detection and uncertainty

The proposed 3D-2D method relies on establishing a point-wise correspondence between 3D data and 2D images through a few facial landmarks. The localization of five to nine points is crucial for estimating the 3D-2D projection and generating the texture signatures. For the gallery, pre-labeled landmarks can be part of the stored metadata. However, for a fully automatic system, landmarks in the probe images must be detected in recognition time. This dependency is a known bottleneck for many analysis and recognition methods in-the-wild. To highlight the performance dependency on landmark localization accuracy, we evaluated the proposed system using automatic detection and uncertainty in point locations. Recognition results for UR2D-R (E) are presented in Fig. 16. The resulting similarity matrices from our system are E-normalized (Section 4.6).

For automatic detection, we use the method by Kazemi and Sullivan (2014) based on an ensemble of regression trees for fitting a landmark-based shape model from image values. We used the pre-trained model included with (King, 2009). Landmarks were automatically detected on both test and gallery images, thus affecting also the reference facial textures. The 3D (gallery) landmarks were kept the same and all scores were E-normalized using the same 23 × 23 gallery distance matrix. We also present results obtained by applying perturbations on the manual landmarks through adding random spatial noise within a radius of 45 pixels for each point independently, which accounts for approximately 10% of the inter-ocular-distance (IOD).

The variations and resolution in UHDB11 have proven to be very challenging for automatic landmark detection, yielding a low overall detection accuracy. The results are summarized in Table 2.

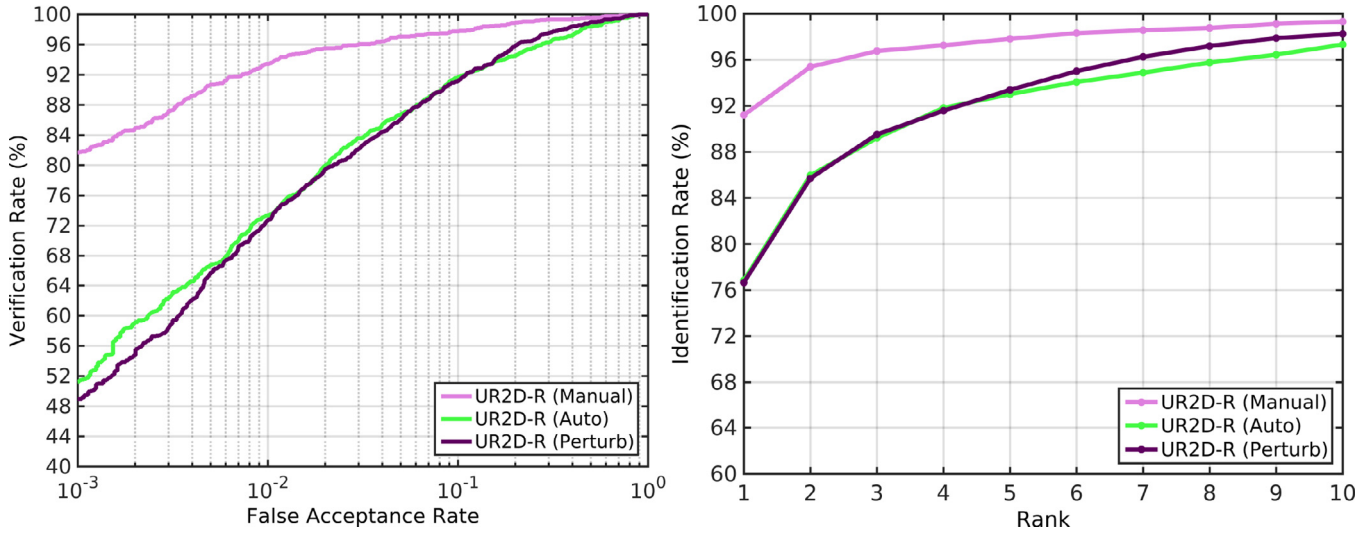

Fig. 16. Recognition with landmark uncertainty: Verification (left) and identification (right) results on UHDB11 using UR2D-R(E), with manual landmarks, automatic landmark detection (Kazemi and Sullivan, 2014) and manual landmarks perturbed by random spatial noise on each landmark within a radius of 45 pixels.

Table 3
Comparisons to state-of-the-art 3D-aided 2D recognition.

Method	Gallery /Subject	Verification				RR1(%)
		VR@10^{-3}	VR@10^{-2}	EER	AUC	
UR2D-R (E)	1	**0.817**	**0.939**	**0.037**	**0.991**	**90.8**
PP-3Da (Z)	49	0.115	0.372	0.211	0.864	81.7
3DMM-Pose (Zhang et al., 2014)*	1	–	–	–	–	88.8
3D-SDM-LBP (Moeini et al., 2015)	441	–	–	–	–	51.1
3D-SDM-HOG (Moeini et al., 2015)	441	–	–	–	–	56.0

NOTE: * indicates that the result is from the original paper.

Compared with manual landmarks, the average localization error of automatic landmark detection, normalized by IOD, is around 0.07, much larger than the reported results obtained on other databases (Kazemi and Sullivan, 2014). As a result, the recognition performance with automatic landmarks shows a 0.0265 decrement and a 14.36% drop in AUC and rank-1 identification rate, respectively. It should be noted though that the detector was not trained on UHDB11 data. Similar results are obtained using perturbed manual landmarks, with a 0.0228 decrement and a 14.55% drop in AUC and rank-1 rate.

5.7. Recognition from low resolution images

The facial images in the UHDB11 database are of high resolution, with an average IOD of approximately 450 pixels. To test robustness of the proposed 3D-2D system in a low resolution setting, we resize UHDB11 images by downsampling them to 0.25 and 0.125 of the original size, forming datasets LR1 and LR2 respectively. We evaluate our 3D-2D system (UR2D, UR2D-R) and an image-based, 2D-2D system, in this case PittPattSDK v5.2 (PP5). The parameters of our 3D-2D system are kept the same except for the texture image (Section 4.3) size, which is halved (from 512 × 512 to 256 × 256) to conform to the low-resolution setting. Results for both face verification and identification are presented in Fig. 17. The resulting similarity matrices from our system are Z-score normalized (Section 4.6).

In the 1/8 resolution set (LR2), both 3D-2D systems degrade, with the one with relighting (UR2D-R) demonstrating improved robustness. Performance of the PP5 2D-2D method on the downsampled sets was close to that obtained on the original high-resolution UHDB11. Notably, and even surprisingly, in the 1/4 resolution set (LR1), the proposed 3D-2D system slightly improves recognition, achieving rank-1 identification rate of 92.0% and verification rate 0.829 at 0.001 FAR, compared to the corresponding 90.8% and 0.817 rates on the original set. This improvement might be due to the resulting textures being formed from coarser features (image values), of a smaller range of illumination variability and higher resiliency to pose errors and uncertainty.

5.8. Comparisons to state-of-the-art 3D-aided 2D recognition

The proposed 3D-2D framework normalizes probe and gallery pose by representing both textures in a canonical 2D space, using the same 3D face model. In comparison, many other methods for 3D-aided 2D recognition will align the two using synthetic views from 3D data or personalized models. In this section, we compare UR2D to two recently published methods that use synthetic images for face recognition with pose variations (Moeini et al., 2015; Zhang et al., 2014), in terms of recognition in UHDB11. Our results, summarized in Table 3, highlight that, in this set, UR2D which is based on a single gallery and a small number of landmark correspondences outperforms state-of-the-art 3D-aided face recognition systems.

The 3D reconstruction based sparse coding approach of (Moeini et al., 2015) is denoted here by 3D-SDM-LBP/HOG depending on the feature representation used (i.e., Local Binary Patterns or Histograms of Oriented Gradients). A 3D elastic model per gallery is reconstructed and used to render synthetic feature images in different poses. During testing, a probe image is matched via the minimum reconstruction error of sparse coding using a set of pose-selected gallery dictionaries. Different from such methods that synthesize multi-view gallery images, Zhang et al. (2014), denoted here 3DMM-Pose, reconstructs a 3D model per image, using a pose initialization estimated from a strong classifier, and generates normalized facial images for frontal face recognition.

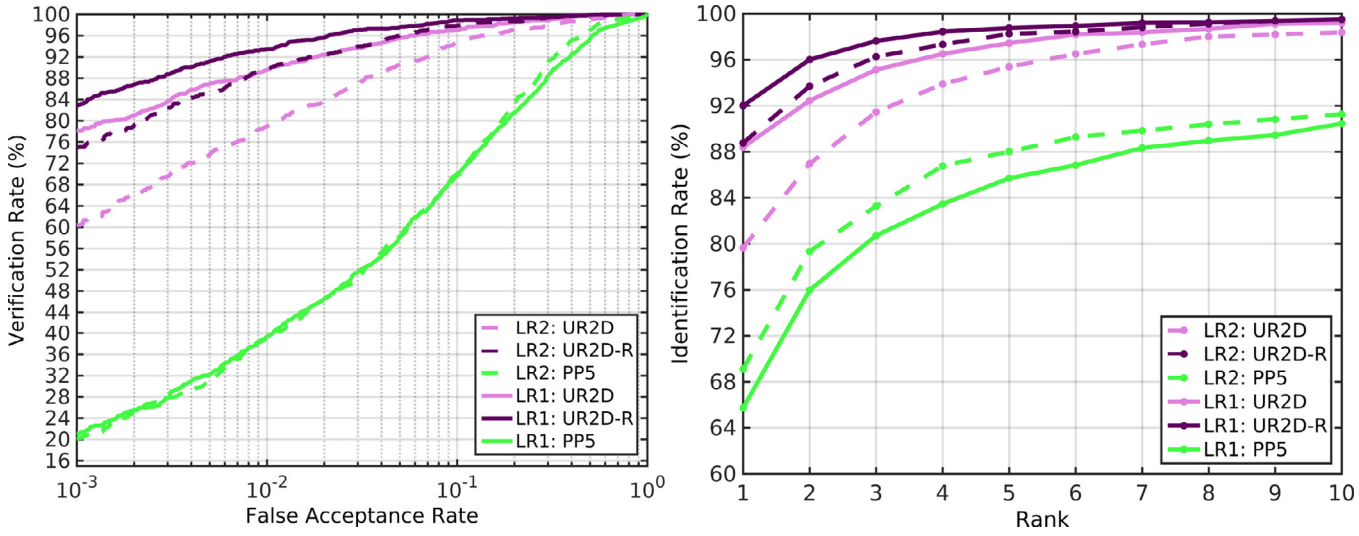

Fig. 17. Recognition with low-resolution images: Verification (left) and identification (right) results on two lower-resolution versions of UHDB11. Images in sets LR1 and LR2 are downsampled to 0.25 and 0.125 size of the original images respectively.

Fig. 18. 3D-generated galleries for 3D-aided 2D recognition, using 15° step yaw-roll angles from $(-45°, -45°)$ in upper left to $(+45°, +45°)$ in lower right.

In addition, we created another 2D-2D baseline using personalized, 3D-generated galleries in multiple poses, from the known 3D data and a regular sampling of the yaw-roll angle space. Specifically, for the original, frontal-pose gallery data, the roll and yaw angles were varied in $[-45°, 45°]$ with steps of 15°, yielding a total of 49 rotation variations; 2D views were generated through orthographic projection of the facial area (Fig. 18). The chosen angles cover the range of variations in the probe data. Note that we used the raw 3D data and not the AFM-generated 3D models. The resulting set of 49 × 23 images, reduced to 617 due to face detection failures for many of the extreme poses, formed the gallery set for verification using PP5 (Pittsburgh Pattern Recognition, 2011) (PP5-3Da in Table 3).

5.9. Dataset independent evaluation: FRGC v2.0

The proposed 3D-2D method is generic and does not depend on models pre-trained on different sets or 3D data. To demonstrate this ability for dataset independent generalization, and highlight some of the limitations of UR2D, we include evaluations on a 3D-2D cohort of FRGC v2.0 database (Phillips et al., 2010; Toderici et al., 2013), which comprises of 250 individual gallery subjects with one 3D mesh and one 2D image each and 470 probe images that was used for 3D-2D evaluations in Al-Osaimi et al. (2011); Zhao et al. (2014). As opposed to UHDB11, the FRGC v2.0 images are of lower resolution and without pose variation, but with different and more complex illumination variations. Following Section 5.7, we also halve the texture image (Section 4.3) size to conform to the low-resolution images. We used PP5 for image-based 2D-2D face recognition and depicted the results for both systems in Fig. 19. The resulting similarity matrices from our system are Z-score normalized (Section 4.6).

In the absence of pose and under a more continuous spectrum of illumination changes, we note that PP5 performance achieves a high 97.0% rank-1 identification accuracy compared to a 78.3% obtained by UR2D, with a similar gap in the verification tests. This can be interpreted from two perspectives. First, as opposed to PP5 which is trained on a large set of images with different illuminations, similar to those exhibited in FRGC v2.0, our method is unsupervised. The relighting approach does not seem to help which might be evident that it cannot compensate for more complex or uncontrolled illumination conditions (such as those coming from shadows, highlights, indoor/outdoor differences et.al.). A remedy, for example, would be to use hand-crafted or learned image representations that are more robust to large and local illumination changes. Second, the pose normalization and alignment setting of UR2D is not that crucial for recognition of roughly frontal faces. A better low-level image value representation could improve the texture comparison.

6. Conclusions

We proposed a novel framework for 3D-2D face recognition that uses 3D and 2D data for enrollment and 2D data for verification and identification. The approach, and an associated face

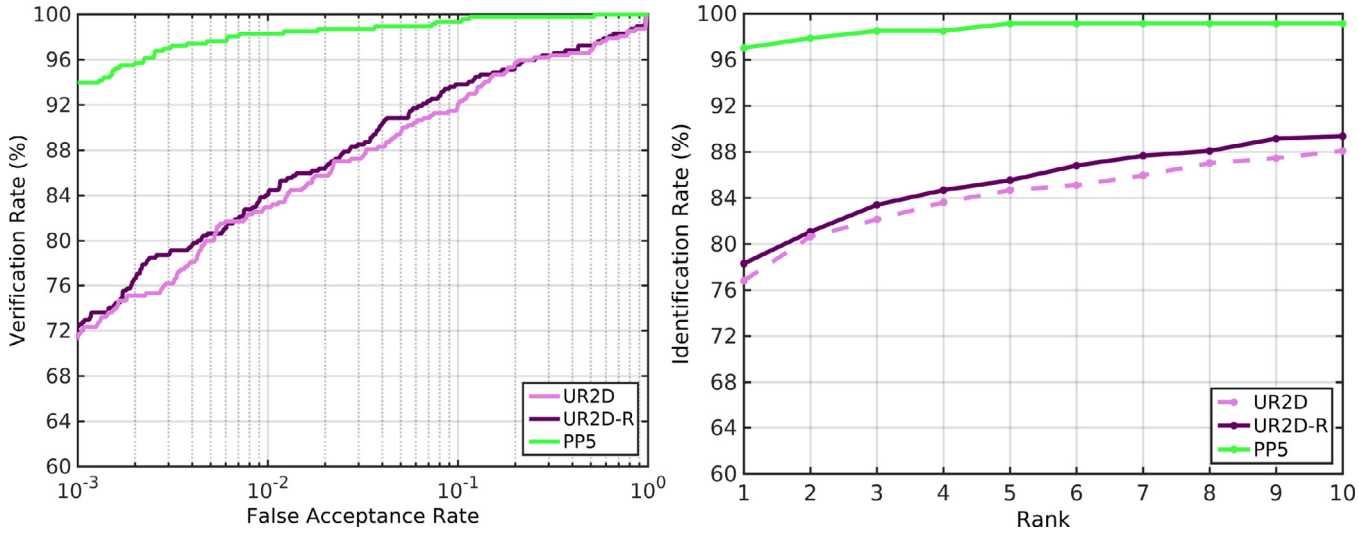

Fig. 19. Recognition on an FRGC v2.0 3D-2D cohort (250 × 470) (Al-Osaimi et al., 2011; Zhao et al., 2014): Verification (left) and identification (right) for 3D-2D and 2D-2D.

recognition system (UR2D), is based on equalizing the pose and illumination conditions between a gallery-probe pair to be matched. To that end, a deformable face model is fitted to the 3D data to register the gallery and probe face textures, obtained under the same model, in a common 2D coordinate system. The process provides a shape-driven, appearance-based representation of faces in the geometry image space and an alignment of the visible facial areas under an estimated pose. The probe light conditions are transferred locally to the gallery texture through a bidirectional relighting scheme based on an explicit skin reflectance model. Quantitative evaluations on a database designed for 3D-2D face recognition benchmarks, under large variations in pose and illumination conditions, demonstrated that 3D-2D performance outperformed 2D-2D, image-based, FR and can approximate 3D-3D, shape-based, FR. The optimality of individual modules, for example light normalization or 2D-based vs. 3D-based similarity, can be further explored by module-specific evaluations, along with comparisons to existing frameworks for 3D model-based recognition and 3D model estimation from images. The proposed framework can potentially be adapted for 3D-aided 2D face recognition on challenging face image databases (e.g., by using an external set of fitted 3D models). Extensions of this work will include the integration and adaptation of landmark detectors for automatic point correspondences under uncertainty; using model-based annotations for region-specific, robust-to-expressions 2D fitting; alternative representations for texture similarity and schemes for non-personalized 3D-model selection.

Acknowledgments

This research was funded in part by the Office of the Director of National Intelligence (ODNI) and by the Intelligence Advanced Research Projects Activity (IARPA) through the Army Research Laboratory (ARL) and by the University of Houston (UH) Eckhard Pfeiffer Endowment Fund. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representative of the official views or policies of IARPA, the ODNI, the U.S. Government, or UH. The authors would like to thank (i) S. Zafeiriou and G. Tzimiropoulos for providing the score normalization code, (ii) L. Chen and LIRIS Lab for sharing their results on UHDB11, and (iii) P. Dou for performing selected experiments in the paper.

References

Abate, A., Nappi, M., Riccio, D., Sabatino, G., 2007. 2D and 3D face recognition: a survey. Pattern Recognit. Lett. 28 (14), 1885–1906.

Al-Osaimi, F.R., Bennamoun, M., Mian, A.S., 2011. Illumination normalization of facial images by reversing the process of image formation. Mach. Vis. Appl. 22 (6), 899–911.

Atick, J.J., Griffin, P.A., Redlich, A.N., 1996. Statistical approach to Shape from Shading: reconstruction of three-dimensional face surfaces from single two-dimensional images. Neural Comput. 8 (6), 1321–1340. doi:10.1162/neco.1996.8.6.1321.

Biswas, S., Aggarwal, G., Chellappa, R., 2009. Robust estimation of albedo for illumination-invariant matching and shape recovery. IEEE Trans. Pattern Anal. Mach. Intell. 31, 884–899. doi:10.1109/TPAMI.2008.135.

Blanz, V., Vetter, T., 2003. Face recognition based on fitting a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell. 25 (9), 1063–1074.

Bowyer, K., Chang, K., Flynn, P., 2006. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. Comput. Vis. Image Understand. 101 (1), 1–15.

Bregler, C., Hertzmann, A., Biermann, H., 2000. Recovering non-rigid 3D shape from image streams. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head, SC, vol. 2, pp. 690–696.

Efraty, B., Huang, C., Shah, S.K., Kakadiaris, I.A., 2011. Facial landmark detection in uncontrolled conditions. In: Proceedings International Joint Conference on Biometrics. Washington, DC.

Fang, T., Zhao, X., Ocegueda, O., Shah, S.K., Kakadiaris, I.A., 2012. 3D/4D facial expression analysis: an advanced annotated face model approach. Image Vis. Comput. 30 (10), 738–749.

Georghiades, A., Belhumeur, P., Kriegman, D., 2001. From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Analys. Mach. Intell. 23 (6), 643–660. doi:10.1109/34.927464.

Gu, L., Kanade, T., 2006. 3D alignment of face in a single image. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, pp. 1305–1312. doi:10.1109/CVPR.2006.11.

Hartley, R.I., Zisserman, A., 2004. Multiple View Geometry in Computer Vision, second Cambridge University Press. ISBN: 0521540518

Huang, D., Ardabilian, M., Wang, Y., Chen, L., 2010. Automatic asymmetric 3D-2D face recognition. In: Proceedings 20th International Conference on Pattern Recognition. Istanbul, Turkey, pp. 1225–1228.

Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L., 2011. Local binary patterns and its applications on facial image: a survey. IEEE Trans. Syst. Man Cybern. Part C 41 (6), 765–781.

Jahanbin, S., Choi, H., Bovik, A., 2011. Passive multimodal 2-D+3-D face recognition using Gabor features and landmark distances. IEEE Trans. Inf. Forensics Security 6 (4), 1287–1304. doi:10.1109/TIFS.2011.2162585.

Jain, A., Nandakumar, K., Ross, A., 2005. Score normalization in multimodal biometric systems. Pattern Recognit. 38 (12), 2270–2285. doi:10.1016/j.patcog.2005.01.012.

Kakadiaris, I.A., Passalis, G., Toderici, G., Murtuza, M.N., Lu, Y., Karampatziakis, N., Theoharis, T., 2007. Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. IEEE Trans. Pattern Anal. Mach. Intell. 29 (4), 640–649. doi:10.1109/TPAMI.2007.1017.

Kazemi, V., Sullivan, J., 2014. One millisecond face alignment with an ensemble of regression trees. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, pp. 1867–1874.
Kemelmacher-Shlizerman, I., Basri, R., 2011. 3D face reconstruction from a single image using a single reference face shape. IEEE Trans. Pattern Anal. Mach. Intell. 33 (2), 394–405.
King, D.E., 2009. Dlib-ml: a machine learning toolkit. J. Mach. Learn. Res. 10, 1755–1758.
L1 Identity Solutions, L1 FaceIt SDK.
Lee, J., Moghaddam, B., Pfister, H., Machiraju, R., 2005. A bilinear illumination model for robust face recognition. In: Proceedings IEEE International Conference on Computer Vision, Beijing, China, 2, pp. 1177–1184.
Lee, M.W., Ranganath, S., 2003. Pose-invariant face recognition using a 3D deformable model. Pattern Recognit. 36 (8), 1835–1846.
Levine, M.D., Yua, Y., 2009. State-of-the-art of 3D facial reconstruction methods for face recognition based on a single 2D training image per person. Pattern Recognt. Lett. 30 (10), 908–913.
Lourakis, M., 2005. Levenberg-Marquardt nonlinear least squares algorithms in C/C++. http://www.ics.forth.gr/~lourakis/levmar/.
Matthews, I., Xiao, J., Baker, S., 2007. 2D vs. 3D deformable face models: representational power, construction, and real-time fitting. Int. J. Comput. Vis. 75 (1), 93–113.
Mian, A., Bennamoun, M., Owens, R., 2007. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 29 (11), 1927–1943.
Moeini, A., Moeini, H., Faez, K., 2015. Unrestricted pose-invariant face recognition by sparse dictionary matrix. Image Vis. Comput. 36, 9–22. doi:10.1016/j.imavis.2015.01.007.
Mpiperis, I., Malassiotis, S., Strintzis, M., 2008. Bilinear models for 3D face and facial expression recognition. IEEE Trans. Inf. Forensics Security 3 (3), 498–511.
Ocegueda, O., Passalis, G., Theoharis, T., Shah, S., Kakadiaris, I.A., 2011a. UR3D-C: linear dimensionality reduction for efficient 3D face recognition. In: Proceedings International Joint Conference on Biometrics. Washington D.C.
Ocegueda, O., Shah, S., Kakadiaris, I.A., 2011b. Which parts of the face give out your identity? In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Colorado Springs, CO, pp. 641–648.
Osadchy, M., Jacobs, D., Lindenbaum, M., 2007. Surface dependent representations for illumination insensitive image comparison. IEEE Trans. Pattern Anal. Mach. Intell. 29 (1), 98–111.
Park, U., Hong, C., Jain, A.K., 2005. 3D model-assisted face recognition in video. In: Proceedings Second Canadian Conference on Computer and Robot Vision. Victoria, Canada, pp. 322–329.
Pekalska, E., Paclik, P., Duin, R., 2002. A generalized kernel approach to dissimilarity-based classification. J. Mach. Learn. Res. 2 (2), 175–211.
Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W., 2005. Overview of the face recognition grand challenge. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, Vol. 1, pp. 947–954. doi:10.1109/CVPR.2005.268.
Phillips, P., Scruggs, W., O'Toole, A., Flynn, P., Bowyer, K., Schott, C., Sharpe, M., 2010. FRVT 2006 and ICE 2006 large-scale experimental results. IEEE Trans. Pattern Anal. Mach. Intell. 32 (5), 831–846. doi:10.1109/TPAMI.2009.59.
Pittsburgh Pattern Recognition, 2011. PittPatt face recognition software development kit (PittPatt SDK) v5.2.
Prabhu, U., Heo, J., Savvides, M., 2011. Unconstrained pose-invariant face recognition using 3D generic elastic models. IEEE Trans. Pattern Anal. Mach. Intell. 33 (10), 1952–1961. doi:10.1109/TPAMI.2011.123.
Rama, A., Tarres, F., Onofrio, D., Tubaro, S., 2006. Mixed 2D-3D information for pose estimation and face recognition. In: Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 361–364.
Riccio, D., Dugelay, J.-L., 2007. Geometric invariants for 2D/3D face recognition. Pattern Recognit. Lett. 28 (14), 1907–1914.
Romdhani, S., Ho, J., Vetter, T., Kriegman, D.J., 2006. Face recognition using 3-D models: pose and illumination. Proc. IEEE 94 (11), 1977–1999. doi:10.1109/JPROC.2006.886019.
Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., Pantic, M., 2013. 300 faces in-the-wild challenge: the first facial landmark localization challenge. In: IEEE International Conference on Computer Vision Workshops. Sydney, Australia, pp. 397–403. doi:10.1109/ICCVW.2013.59.
Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L., 2008. Bosphorus database for 3D face analysis. In: Biometrics and Identity Management. In: Lecture Notes in Computer Science, vol. 5372, pp. 47–56.
Smith, W., Hancock, E., 2005. Estimating the albedo map of the face from a single image. In: Proceedings IEEE International Conference on Image Processing, Genoa, Italy, vol. 3, pp. 780–783.
Theoharis, T., Passalis, G., Toderici, G., Kakadiaris, I.A., 2008. Unified 3D face and ear recognition using wavelets on geometry images. Pattern Recognit. 41 (3), 796–804. doi:10.1016/j.patcog.2007.06.024.
Toderici, G., Evangelopoulos, G., Fang, T., Theoharis, T., Kakadiaris, I.A., 2013. UHDB11 database for 3D-2D face recognition. In: Proceedings 6th Pacific-Rim Symposium on Image and Video Technology. Guanajuato, Mexico, pp. 73–86.
Toderici, G., Passalis, G., Zafeiriou, S., Tzimiropoulos, G., Petrou, M., Theoharis, T., Kakadiaris, I.A., 2010. Bidirectional relighting for 3D-aided 2D face recognition. In: Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, pp. 2721–2728. 978-1-4244-6983-3/10
Tzimiropoulos, G., Argyriou, V., Zafeiriou, S., Stathaki, T., 2010. Robust FFT-based scale-invariant image registration with image gradients. IEEE Trans. Pattern Anal. Mach. Intell. 32 (10), 1899–1906.
Tzimiropoulos, G., Zafeiriou, S., Pantic, M., 2012. Subspace learning from image gradient orientations. IEEE Trans. Pattern Anal. Mach. Intell. 34 (12), 2454–2466.
Vijayan, V., Bowyer, K., Flynn, P., Huang, D., Chen, L., Ocegueda, O., Shah, S., Kakadiaris, I.A., 2011. Twins 3D face recognition challenge. In: Proceedings International Joint Conference on Biometrics. Washington, DC.
Wang, Y., Liu, J., Tang, X., 2010. Robust 3D face recognition by local shape difference boosting. IEEE Trans. Pattern Anal. Mach. Intell. 32 (10), 1858–1870.
Wang, Y., Zhang, L., Liu, Z., Hua, G., Wen, Z., Zhang, Z., Samaras, D., 2009. Face relighting from a single image under arbitrary unknown lighting conditions. IEEE Trans. Pattern Anal. Mach. Intell. 31 (11), 1968–1984.
Wolf, L., Hassner, T., Taigman, Y., 2011. Effective unconstrained face recognition by combining multiple descriptors and learned background statistics. IEEE Trans. Pattern Anal. Mach. Intell. 33 (10), 1978–1990. doi:10.1109/TPAMI.2010.230.
Yin, L., Yourst, M., 2003. 3D face recognition based on high-resolution 3D face modeling from frontal and profile views. In: Proc. ACM SIGMM Workshop on Biometrics Methods and Applications. New York, NY, USA, pp. 1–8.
Zhang, W., Huang, D., Samaras, D., Morvan, J.-M., Wang, Y., Chen, L., 2014. 3D assisted face recognition via progressive pose estimation. In: Proc. IEEE International Conference on Image Processing. Paris, France, pp. 728–732.
Zhang, W., Huang, D., Wang, Y., Chen, L., 2012. 3D-aided face recognition across pose variations. In: Chinese Conference on Biometric Recognition, pp. 58–66.
Zhao, X., Evangelopoulos, G., Chu, D., Shah, S., Kakadiaris, I.A., 2014. Minimizing illumination differences for 3-D to 2-D face recognition using lighting maps. IEEE Trans. Cybern. 44 (5), 725–736. doi:10.1109/TCYB.2013.2291196.
Zhao, X., Shah, S., Kakadiaris, I.A., 2012. Illumination normalization using self-lighting ratios for 3D-2D face recognition. In: Proc. European Conference on Computer Vision Workshop. Firenze, Italy, pp. 220–229.
Zhao, X., Shah, S., Kakadiaris, I.A., 2013. Illumination alignment using lighting ratio: application to 3D-2D face recognition. In: Proc. 10th International Conference on Automatic Face and Gesture Recognition. Shanghai, China, pp. 1–6. doi:10.1109/FG.2013.6553782.
Zhou, Z., Ganesh, A., Wright, J., Tsai, S.-F., Ma, Y., 2008. Nearest-subspace patch matching for face recognition under varying pose and illumination. In: Proceedings IEEE International Conference on Automatic Face Gesture Recognition. Amsterdam, The Netherlands doi:10.1109/AFGR.2008.4813452.