

# UHDB11 Database for 3D-2D Face Recognition

George Toderici<sup>1</sup>, Georgios Evangelopoulos<sup>1</sup>, Tianhong Fang<sup>1</sup>  
Theoharis Theoharis<sup>2,1</sup>, and Ioannis A. Kakadiaris<sup>1</sup>

<sup>1</sup> Computational Biomedicine Lab, Department of Computer Science,  
University of Houston, Houston, TX 77204, USA

<sup>2</sup> IDI, Norwegian University of Science and Technology (NTNU), Norway

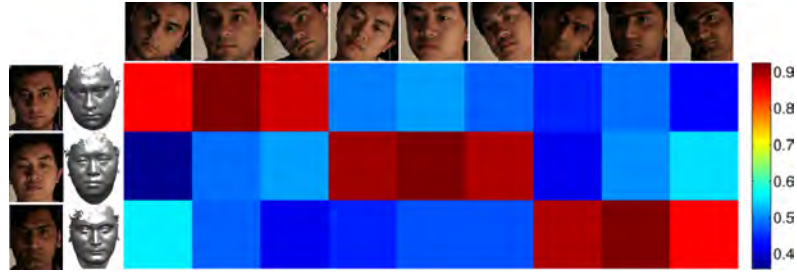
**Abstract.** Performance boosts in face recognition have been facilitated by the formation of facial databases, with collection protocols customized to address challenges such as light variability, expressions, pose, sensor/modality differences, and, more recently, uncontrolled acquisition conditions. In this paper, we present database UHDB11, to facilitate 3D-2D face recognition evaluations, where the gallery has been acquired using 3D sensors (3D mesh and texture) and the probes using 2D sensors (images). The database consists of samples from 23 individuals, in the form of 2D high-resolution images spanning six illumination conditions and 12 head-pose variations, and 3D facial mesh and texture. It addresses limitations regarding resolution, variability and type of 3D/2D data and has demonstrated to be statistically more challenging, diverse and information rich than existing cohorts of 10 times larger number of subjects. We propose a set of 3D-2D experimental configurations, with frontal 3D galleries and pose-illumination varying probes and provide baseline performance for identification and verification (available at <http://cbl.uh.edu/URxD/datasets>).

**Keywords:** face recognition, face databases, 3D-2D facial data, illumination, face pose, verification, identification, computer vision

## 1 Introduction

Databases for face recognition focus either on the shortcomings of conventional 2D images (*e.g.*, acquisition conditions, inter-person variability) or on the benefits of using alternative sensors and modalities (*e.g.*, 3D scans and facial models). In addition, both modalities are traditionally employed in multimodal, 2D+3D systems [4] where shape and texture features are combined to compute face similarity. Databases that contain both types of data (*i.e.*, 2D and 3D), apart from the 2D texture of a 3D face [11], were designed in order to primarily explore the merits of 3D-3D versus 2D-2D or multi-view 2D-2D recognition when both gallery (or target) and probe (or query) sets originate from the same modality.

A different paradigm for face recognition, proposed to combine the detail captured by 3D data with the practical applicability of a 2D recognition system, is a 3D-2D, cross-modal framework, where the probe and gallery sets are acquired from different sensors. One of the conjectures posed and tested by the Face Recognition Grand Challenge [11] involved the effectiveness of high-resolution 2D images versus 3D scans for recognition. By combining the two in an asymmetric, cross-sensor framework, efficient,



**Fig. 1.** 3D-2D face recognition using 3D samples for the gallery and 2D probes. Color represents similarity values in  $[0, 1]$ .

discriminative and practical systems and methods can arise for face verification or face identification (Fig. 1).

By inter-changing the type of data for the enrollment and recognition sessions, different practical scenarios can be formulated. A 3D-2D framework, where the need for 3D acquisition hardware is restricted to enrollment only, can facilitate the acquisition, storage and distribution of highly-descriptive databases of 3D models. On a 2D-3D framework, the abundance of existing face databases, composed primarily of 2D data, can provide reference enrollment sets for matching 3D facial scans. 3D-2D (or 2D-3D) face recognition is distinct from *asymmetric* recognition, where shape features are matched against texture features (heterogeneous 3D FR), *3D-aided* recognition, where real or reconstructed 3D models are used for registration or normalization (3D-aided 2D FR), or *multi-modal* recognition, where shape and texture data are fused on a single facial signature (3D+2D FR).

Research on 3D-2D face recognition has been focusing primarily on methods, whose performance is evaluated on existing 3D databases, with a subset of the registered textures of the 3D data used as 2D input. However, a realistic 3D-2D configuration may involve different resolutions, sensors, and acquisition sessions between enrollment and recognition sets. To facilitate research on 3D-2D and 3D-aided 2D face recognition, we created database UHDB11 (University of Houston Database 11), that aims to address 3D-2D database requirements and limitations related to high-resolution data, full-face 3D data and identity-independent variability in pose and lighting conditions. In spite of the small identity sample size, compensated by the relatively large number of probe instances per individual (approx. 70) and the roughly-quantized head-pose (12) and illumination variations (6), database UHDB11 provides challenging data instances for pairwise comparisons and complements the limited (in number) available 3D-2D databases such as FRGC v2.0 [11].

**Why is UHDB11 database useful?** (i) It provides a 3D+2D modality, asymmetric 3D/2D probe-gallery matches, and data from three modalities per subject (3D shape, 3D texture, 2D image), (ii) it covers different acquisition conditions (posed subjects apart from light conditions), (iii) it simulates realistic application scenarios (3D enrollment and uncontrolled 2D recognition), and (iv) it is statistically more challenging, diverse and information rich than evaluation cohorts of 10 times larger number of subjects [3].



Fig. 2. Sample 2D image and 3D mesh data from the UHDB11 database.

Table 1. Databases for 2D-2D and 3D-3D face recognition with variations in: **P** : Pose, **V**: Viewpoints, **I**: Illumination, **E**: Expressions, **M**: Multiple instances/captures, **T**: Time-lapse, **S**: Sessions, **O**: Occlusions, **AU**: Action Units (the levels of variation are given in parentheses).

Database	Ref.	Year	Type	# Subjects	Variation	# Samples
FERET	[10]	2000	2D-2D	1,199	T, M	14,126
CMU PIE	[15]	2002	2D-2D	68	V (13), E (4), I (43)	>41,000
FRGC v2.0	[11]	2005	2D-2D	466	M, I (Controlled/Uncontrolled)	24,042
LFiW	[9]	2007	2D-2D	5,749	Uncontrolled	13,233
CMU Multi-PIE	[7]	2008	2D-2D	337	V (15), E (6), I (19), S (4)	>750,000
GBU	[12]	2011	2D-2D	437	I (Outdoor, Indoor Ambient)	6,510
FRGC v2.0	[11]	2005	3D-3D	466	E (7)	4,007
BU-3DFE	[18]	2006	3D-3D	100	E (7)	2,500
UND ND-2006	[5]	2007	3D-3D	888	M, E (6)	13,450
Bosphorus	[14]	2008	3D-3D	105	P (13), E (6), AU (28), O (4)	4,652
3D TEC	[17]	2011	3D-3D	214	Twins, E (2)	428

The rest of this paper is organized as follows: In Sec. 2, we provide a brief overview of existing databases for 2D-2D and 3D-3D face recognition along with reference cohorts previously used for 3D-2D experiments. In Sec. 3, we outline the purpose and potential use of the proposed 3D-2D database. The data specifications and six potential experimental configurations for 3D-2D benchmarking are presented in Sec. 4, while a baseline evaluation using a system for 3D-2D face recognition is provided in Sec. 5. Limitations and extensions of UHDB11 are discussed in Sec. 6.

## 2 Face recognition databases

The type and variation in existing face recognition databases depends on the challenges aimed to be addressed by each one [6]. Categories include facial expressions [13], uncontrolled acquisition conditions [9], illumination and viewpoints [15, 7], different sensors and resolutions [11, 4], identity mismatches due to different sessions [12] and even twin subjects [17]. Although a comprehensive review of available databases is out of the scope of this paper, we provide a brief overview of well-established sets (Table 1) in order to highlight their targeted challenges and contrast them to UHDB11.

### 2.1 Databases for 2D-2D recognition

The FERET database [10] was one of the first systematically collected sets for large scale 2D face recognition evaluations. It contains 14,126 images of 1,199 individuals,

with time-lapse and multiple sessions for some of the subjects (365 duplicate sets). The average inter-pupil distance (IPD) in the FERET faces is 68 pixels with a standard deviation of 8.7 pixels. Controlled light conditions were also used in the CMU PIE database [15], which in addition included viewpoint variations, resulting in different head poses, using a fixed camera grid and a still subject. Data are color images of dimension  $640 \times 486$  pixels. The extended CMU Multi-PIE [7] increased the number of subjects (337), acquisition sessions and resolution of the collected images to  $3,072 \times 2,048$ , with an IPD typically more than 400 pixels. The Face Recognition Grand Challenge [11] defined recognition frameworks on the 2D sets of FRGC v2.0 (Experiments 1 and 4) on probe sets of size 16,028 and 8,014 for controlled and uncontrolled (*e.g.*, outdoor) conditions. Multiple viewpoints or sessions per subject were also provided for multi-still recognition (Experiment 2). High-resolution, frontal images, of average IPD  $175 \pm 36$  pixels, from the uncontrolled set of the FRVT 2006 Notre Dame data were grouped in three partitions of increasing difficulty (easy, average and hard matching pairs) on the GBU database [12]. Labeled Faces in the Wild (LFIW) [9] was proposed as a large scale, unconstrained recognition benchmark, where both probe and gallery faces can vary considerably with respect to non-identity or imaging factors. The images are of resolution  $250 \times 250$  pixels and were collected for 5,749 subjects using face detection on web data.

## 2.2 Databases for 3D-3D recognition

FRGC v2.0 is a multi-modal database which was designed to explore the independent and comparative role of high resolution images, multiple captures, and 3D facial imagery for face recognition [11]. The size of the FRGC v2.0 3D set is 4,007 samples, corresponding to 1-22 instances for each of 466 subjects. Samples are range data of resolution  $640 \times 480$  obtained from a structured light sensor, with a registered texture image of the same dimensions and an average IPD of 160 pixels. Four facial landmarks are provided as meta-data. The largest database of 3D facial data is UND ND-2006 [5], originally proposed for multi-instance gallery and facial component studies, containing 13,450 scans for 888 subjects. Variations of identity among twins are captured in the 3D TEC [17] database that contains the 3D facial data for 107 pairs of identical twins.

BU-3DFE [18] was the first 3D database developed for facial expression analysis. It includes facial meshes for 100 subjects, under a neutral and six prototypical expressions, on four intensity levels. The 3D model data were acquired using the 3dMD system [2], cropped to an effective facial region size of 13K – 21K polygons, with an associated texture of  $512 \times 512$  pixels. The locations of 83 facial points are also specified per sample. Similarly, the Bosphorus database [14] consists of 3D data captured using a structured-light 3D sensor, for the study of multi-expression, multi-pose 3D recognition. Data variability includes facial expressions from six prototypical emotions and 28 groups of similar action units (lower, upper and combinations), pose variations in pitch and yaw angles and occlusions. The texture images are of relatively high resolution ( $1,600 \times 1,200$  pixels) and the segmented 3D facial areas correspond to approximately 35K vertices. A set of 24 manually labeled facial landmarks is provided as meta-data.

**Table 2.** Databases and experimental cohorts for 3D-2D face recognition.

	Source	Subj.	Gallery Size	Probe Size	Probe Type	Resolution
FRGC-Exp 5 [11]	FRGC v2.0	466	4,007	16,028	2D Image	$1,200 \times 1,600$
FRGC-Exp 6 [11]	FRGC v2.0	466	4,007	8,014	2D Image	$2,272 \times 1,704$
Huang <i>et al.</i> [8]	FRGC v2.0	466	466	3,541	2D Texture	$640 \times 480$
Al-Osaimi <i>et al.</i> [3]	FRGC v2.0	250	250	470	2D Texture	$640 \times 480$
GBU/FRGC	GBU, FRGC v2.0	437	186	(set) 2,170	2D Image	$3,008 \times 2,000$
UHDB11-Exp. 1	UHDB11	23	23	1,602	2D Image	$3,888 \times 2,592$

### 2.3 Cohorts for 3D-2D recognition

Several partitions or combinations of existing sets with 2D and 3D data can be employed for 3D-2D experiments (Table 2). Two of the configurations for FRGC v2.0 [11] specify a 3D-2D scenario, where the gallery is the full set of 3D data (4,007) and each probe is a single 2D image. For controlled (Experiment 5) and uncontrolled (Experiment 6) acquisition, the probe set size is 16,028 and 8,014, of pixel resolution  $1,200 \times 1,600$  and  $2,272 \times 1,704$  respectively.

Huang *et al.* [8] designed an asymmetric 3D-2D recognition configuration from the 4,007 3D samples of FRGC v2.0 that correspond to 466 subjects. The first neutral-expression instance from each subject was used to form a 3D gallery of 466, and the texture images of the remaining 3,541 instances, cropped to an effective size of  $175 \times 190$ , form the 2D probe set. Evaluations involve 2D-2D (texture-based), 3D-2D (asymmetric) and fused matching. Similarly, Al-Osaimi *et al.* [3] used 470 textures as probes, matched against 250 3D galleries in an open recognition cohort, for studying illumination normalization. Alternative cohorts can be formed through the overlap of subject identity sets for databases originating from the same Institutes. For example, using the 2D images from GBU [12] (2,170 per partition) and the corresponding 3D data from FRGC v2.0, following the identity naming convention of the UND data, a 3D-2D experiment of 186 3D targets and 2,170 probes can be defined for the difficult set.

Compared to these cohorts, UHDB11 proposes to address limitations regarding resolution, variability and type of 3D/2D data. FRGC v2.0, for example, offers a large-scale framework, however with: (i) frontal-only 2D probes, (ii) lower resolution, (iii) no 2D images available for the 3D gallery (3D/2D same session) and (iv) no quantization of the space of (illumination) variations that can facilitate condition-specific quantitative analysis. With respect to [3], empirical statistical evidence from 3D-2D and 2D-2D recognition suggests that UHDB11 (i) provides richer information (higher pairwise score variance) from less subjects, and (ii) is more challenging (in terms of verification performance and match-non-match score separability).

## 3 UHDB11 intended purpose and uses

The intended use of the developed database is primarily the performance and robustness evaluation of 3D-2D face recognition algorithms under sample variability related to pose and/or lighting. The enrollment set, referred also as gallery or target, is the collection of data with known identity and the recognition set, referred as probe or query,



**Fig. 3.** Database samples acquired with single light (3) and frontal pose (5) as raw 3D (top), cropped in a sphere of radius 150 units centered at the tip of the nose, and 2D textures (bottom). The samples correspond to the gallery set of the full UHDB11 3D-2D cohort (Exp. 1).

is the collection of samples of unknown identity that are acquired off- or on-line. The recognition set can be restricted to the gallery identities (closed set) or include previously unseen samples (open set). For *face verification*, where a claimed identity is verified on the set of gallery identities, a decision for a positive match is obtained on a one-to-one matching basis. For *face identification*, the closest ranking identity is retrieved after one-to-many matching.

In a 3D-2D recognition setting the gallery data are 3D and the probe data 2D. If the gallery is represented by both 2D and 3D modalities, template face signatures can be extracted using any aspect of face appearance (texture and/or shape). Conversely, a 2D-3D setting yields comparisons of 3D probes to 2D gallery images. By specifying close-set data partitions and inter-changing the role and number of probe and gallery instances, both settings can be explored through the proposed database. In terms of acquisition conditions, UHDB11 can be used to quantify the effects of pose (head rotation in 2D views) and illumination (intensity, direction, shadows) on recognition performance. Different cohorts can be formed adapted to pose, lighting or their combinations. In addition, methods developed for multiple gallery instances, for enhancing the target set or normalizing score values, can be assessed since multiple captures per subject are available. A suggested list of experimental configurations is provided in Sec. 4.2.

With both 3D and 2D data available, the comparison of 3D-2D frameworks to their 2D-2D or 3D-3D counterparts on the same data is advocated [11, 16]. We use the following definitions: *Image* refers to a 2D still obtained via a photographic camera, *shape* refers to the 3D data component acquired by a 3D scanner and *texture* to the corresponding 2D component, registered to the 3D. UHDB11 can be used to evaluate image-image algorithms for conventional 2D face recognition and 3D model-based, image-image methods for 3D-aided 2D face recognition. Likewise, it can assist benchmarking of asymmetric (*e.g.*, texture-image, shape-image) [8] or multimodal schemes (shape/texture-image) [4] for performance evaluation across features and modalities.



**Fig. 4.** Illumination variations in 2D images. Rows correspond to light arrays positioned left and right from the subject respectively. Columns correspond to lower, middle, upper positioned lights. Different lights are indexed 1-6, from upper left to lower right.

## 4 Database description

### 4.1 Acquisition and specifications

Three-dimensional data were captured using a 3dMD two-pod optical scanner [2] and 2D data using a commercial Canon DSLR camera. Lighting condition variations were obtained using multiple diffuse lights, from incandescent light bulbs with approximate color temperature of 2,800K. Both 3D and 2D data were acquired from 23 subjects of different ethnic groups and gender, under all combinations of six illumination conditions and twelve head poses, four yaw and three roll rotations, resulting in 72 different pose/light variations per subject. From the  $23 \times 72$  pool, 31 instances have been removed due to inconsistent specifications or corrupted data, yielding 1,625 samples across subjects and acquisition conditions.

Six directional lights (indexed by 1-6) were positioned on the sides of the camera pair in two stacked, vertical arrays (lights 1-3 left and lights 4-6 right), simulating directional light conditions on the face, apart from the ambient room light. Figure 4 illustrates the lighting conditions formed on a frontal face from the left (top row) and the right (bottom row) light arrays. For each illumination condition, a subject was asked to face four different points in the room, generating rotations on the Y axis (yaw angles). For each rotation on Y, three images with rotations on the Z axis (roll angles) were acquired, assuming a coordinate system with the Z axis pointing towards the camera. In Fig. 5, rows correspond to yaw and columns to roll pose variations.

**Table 3.** UHDB11 face sizes (inter-pupil distance in units).

	Min	Max	Mean	Median	Std. Dev.
Gallery (3D)	61	70	66	66	2
Gallery (2D)	349	440	404	406	25
Probe (2D)	253	511	402	400	34



**Fig. 5.** Head pose variations in 2D images. Rows correspond to yaw and columns to roll rotations. Different poses are indexed 1-12 (upper left to lower right).

Since poses in images are the result of actual head movement (*i.e.*, camera configuration is fixed) as opposed to varying the camera viewpoints (as in most 2D databases), precise ground truth for yaw and roll degrees is not available (*e.g.*, through camera information and pose from calibration parameters). Poses are quantized with respect to the virtual positions and instructions given to subjects during acquisition. A finer distribution of the rotation angles can be obtained through automatic pose estimation using 3D-2D landmark correspondences.

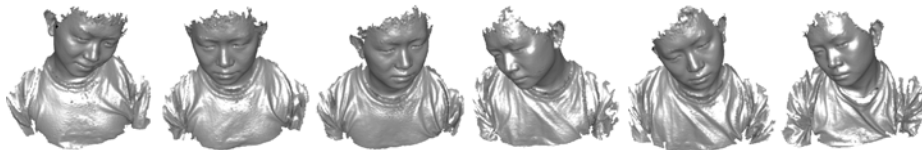
The 2D images are high-resolution color JPEGs of size  $3,888 \times 2,592$  pixels, and the textures for the 3D facial meshes are bitmaps (BMP) of dimension  $2,732 \times 1,948$  pixels. The raw 3D data include the upper torso and subject head, whereas the head area corresponds roughly to 10K – 15K vertices and 20K – 30K triangles, depending on head size. Data pre-processing can include isolating the 3D facial area by defining a sphere around the most extruded facial location (tip of nose). The resulting 3D sets (Fig. 3,7) range from 9.5K (min) to 17.5K (max) vertices and 18K to 35K triangles, depending on face size.

Mesh data were reconstructed from the stereo pair as a single point cloud, triangulated in a polygonal mesh. The 3dMD acquisition system is composed from two stereo cameras (pods), under different viewpoints and provides a  $180^\circ$  face coverage (ear-to-ear), with approximately 1.5 ms capture speed at the highest resolution. Data are captured as a single point cloud from the stereo pair with a surface accuracy of less than 0.2 mm RMS [2].



**Table 4.** UHDB11 database evaluation configurations and experiments.

Setting	Type	Gallery		Probes	
		# Instances (per subject)	# Light - # Pose	# Instances	# Light - # Pose
<b>Exp. 1</b>	Full	23 (1)	1-1	1602	6-12
<b>Exp. 2</b>	Multi-sample	138 (6)	6	1487	6-12
<b>Exp. 3</b>	Light Var.	23 (1)	1-1	115	5-1
<b>Exp. 4</b>	Pose Var.	23 (1)	1-1	250	1-11
<b>Exp. 5</b>	Yaw Var.	23 (1)	1-1	382	6-3
<b>Exp. 6</b>	Op. Light	23 (1)	1-1	814	3-12

**Fig. 6.** Sample 3D data with head pose, corresponding to the single-subject poses of the multi-sample gallery (Exp. 2 in Table 4).

## 4.2 Cohorts and 3D-2D experiments

The multitude of illumination/pose conditions on UHDB11 enables the specification of different 3D-2D verification and identification tasks. As a reference, we define six experimental settings (Table 4), for which the verification rate at  $10^{-3}$  false acceptance rate and rank-1 identification rate can be used, at a minimum, as performance metrics for benchmark comparisons.

**Exp. 1 - Full set, frontal galleries.** The default 3D-2D configuration is designed as a closed-set experiment, with 23 gallery datasets (3D mesh/texture, 2D image), shown in Fig. 3, and 1,602 2D probe images. The gallery is composed of a single scan, texture and image per subject, of frontal pose and unique light condition (3). The remaining, multiple 2D images per subject are used for the probe set.

**Exp. 2 - Multi-sample galleries:** For the effect of multi-instance enrollment, a configuration with multiple 3D data for each of the 23 subjects is used as gallery. Six unique pose-light instances per subject are used as gallery data, forming a total set of 138. Figure 6 depicts an example of multi-sample 3D data for a single subject. The remaining 1,487 samples are used as a closed 2D probe set.

**Exp. 3 - Light variation:** To verify resilience in illumination variations, the probe set is composed by the 2D subset with frontal pose, across the five lighting conditions (1,2,4-6) that are not in the default gallery set (Exp. 1).

**Exp. 4 - Pose variation:** To isolate the influence of pose, the probe set is composed by the 2D subset with gallery light conditions (3), across the non-frontal pose variations (1-4, 6-12) that are not in the default gallery set (Exp. 1).

**Exp. 5 - Yaw variation:** For the influence of the more common, yaw-only viewpoint variations, the zero-roll poses (2, 8, 11) at the gallery light condition (3) are used for the probe set, with the default, frontal-pose gallery (Exp. 1).



**Fig. 7.** The reference UHLS9 facial landmark set (nine points on eyes, mouth and nose) annotated on 3D meshes and 2D images.

**Exp. 6 - Light direction:** For the influence of opposing illumination conditions, the probe set is formed from the frontal pose instances (5), illuminated from an opposite to the default gallery (Exp. 1) light direction (4, 5, 6).

### 4.3 Facial landmarks

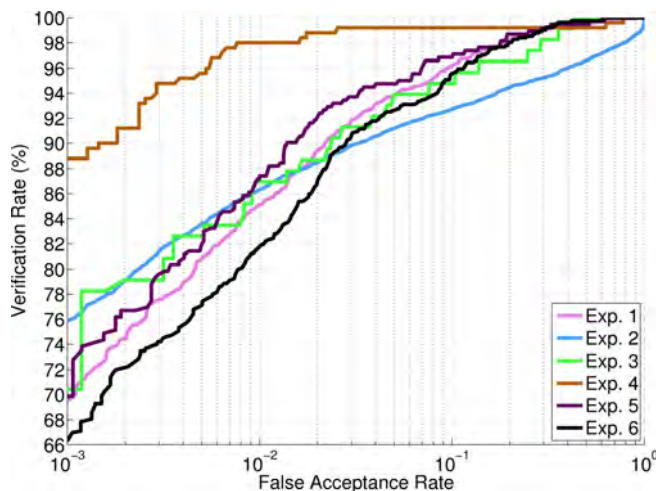
The locations of nine facial point landmarks are provided as a reference with UHDB11 for the purpose of landmark detection, registration or pose estimation. Landmarks have been manually annotated by experts for both 3D and 2D data, examples of which are shown in Fig. 7 superimposed on 3D meshes and 2D images. The reference set (UHLS9) consists of the outer and inner eye corners, nose tip and nose inner corners and mouth corners. For 3D, both the mesh and overlaid texture have been used for accurate localization. Hidden points due to self-occlusions have been excluded from the 2D sets, with the final valid number of landmarks varying from four to nine across the database.

## 5 Baseline performance evaluation

As baseline evaluation, we report recognition performance on the UHDB11 experiments using a reference 3D-2D system [16], where the tested conjecture was that 3D-2D surpasses 2D-2D and can approximate shape-based 3D-3D recognition. Results are reported using light normalization through a bidirectional relighting module. Raw similarity matrices have been normalized using Z-score normalization, although improved results have been reported using a Multi-Dimensional Scaling projection of the distance matrix.

### 5.1 System description

For gallery processing, raw 3D data are fitted to a 3D deformable face model and model-registered images for the texture components are generated using the model surface parametrization. For recognition, a 2D probe is transformed to a geometry texture image, given a fitted gallery model and the 3D-2D landmark correspondence between image and mesh. The 3D-2D projection parameters are estimated for each probe-gallery pair, so that a mapping between mesh and image points is defined. To match the illumination of the probe texture, the gallery texture is bidirectionally relit using an analytic



**Fig. 8.** ROC curves (in logarithmic FAR range) for 3D-2D face verification using the reference system, on the different experiments of UHDB11 database.

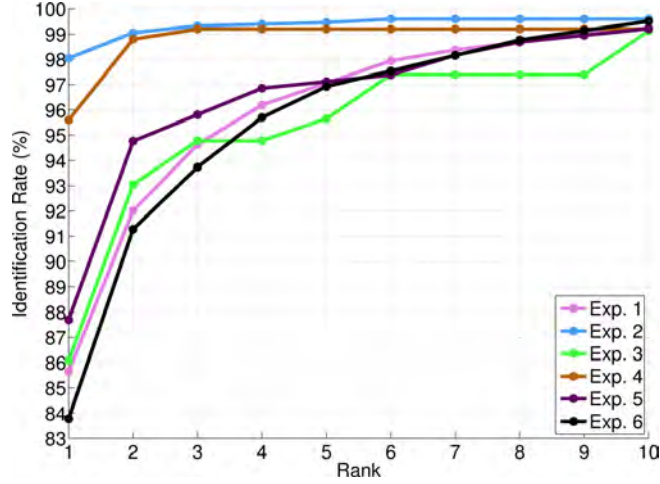
skin reflectance model and an iterative optimization scheme. The final similarity score is based on the correlation of the gradient orientations of the pose and light-normalized texture pair.

## 5.2 Verification and identification

Performance for 3D-2D face recognition is reported for the tasks of verification and identification using receiver operating characteristics (ROC) curves that show verification rate (VR) at varying false acceptance rates (FAR) (Fig. 8), and cumulative match characteristics (CMC) curves that show rank- $k$  recognition rate (RR $k$ ) (Fig. 9). For quantitative comparisons we report the following curve-extracted measures in Table 5: VR at FAR = 0.001 and FAR = 0.01, equal error rate (EER), *i.e.*, the point where false accept equals false reject rate, area under curve (AUC) for verification and rank-1 recognition rate (RR1) for identification.

The specified UHDB11 experiments (Table 4) constitute configurations of different sizes, thus different number of total comparisons for ROCs, ranging from 2,645 (Exp. 3) to 36,846 (Exp. 1). For all experiments the verification rate is above 66% at 0.001 and 81% at 0.01 FAR. In addition, AUC values are very high (above 0.98 in every case), and the EER are low (below 0.07 in all cases). For the full UHDB11 3D-2D set (Exp. 1), verification performance is 69.5% at 0.001 and 85.1% at 0.01 FAR, with a set size that is two to 10 times larger than experiments 3-6. This performance is superior to 2D-2D FR systems [16] and can be additionally enhanced by more sophisticated score normalization algorithms.

A relative evaluation of the different experiments allows for observations on the challenges of the proposed database, as seen by the reference system. Overall best performance, with a 88.8% VR at 0.001 FAR, is achieved in Exp. 4 ( $23 \times 250$ ) where



**Fig. 9.** CMC curves (ranks 1-10) for 3D-2D face identification using the reference system, on the different experiments of UHDB11 database.

the light conditions are the same in probe and gallery sets. This indicates the anticipated decrease in performance when imposing illumination inconsistencies in the 2D images and an increased pose-resilience of the reference system. Analogously, Exp. 3 ( $23 \times 115$ ) where frontal pose is used in both sets, demonstrates a 10.5% lower verification rate. Experiments 5 and 6 involve combinations of pose and light variations, similar to Exp. 1, but constrain the type of head-pose and light direction variation respectively. Differences due to opposing lighting directions, captured by Exp. 6 ( $23 \times 814$ ), attain the lowest relative performance, possibly due to asymmetric shadows and specularities.

Identification baselines for the different 3D-2D experiments, shown in Fig. 9 for rank 1-10 out of the 23 gallery identities, are all above 84% and up to 96% rank - 1 rate. With the main baseline (Exp. 1) at 85.6%, Exp. 5 ranks second best to Exp. 4, along with Exp. 3. Since it is not clear if this performance ordering of the six configurations is due to the reference system or the challenges in the database, the results of competing 3D-2D algorithms will further rate the difficulty of each set.

**Table 5.** Face recognition results using the reference 3D-2D system on UHDB11 database.

Method	Size	Verification				RR1(%)
		@ $10^{-3}$ FAR	@ $10^{-2}$ FAR	EER	AUC	
<b>Exp. 1</b>	$23 \times 1,602$	0.695	0.851	0.056	0.986	85.6
<b>Exp. 2</b>	$138 \times 1,487$	0.759	0.864	0.078	0.960	98.1
<b>Exp. 3</b>	$23 \times 115$	0.783	0.870	0.062	0.983	86.1
<b>Exp. 4</b>	$23 \times 250$	0.888	0.980	0.016	0.994	95.6
<b>Exp. 5</b>	$23 \times 382$	0.728	0.874	0.051	0.988	87.7
<b>Exp. 6</b>	$23 \times 814$	0.663	0.817	0.067	0.985	83.8

## 6 Limitations and extensions

Database UHDB11 is a systematic approach to bridge the availability gap between 3D-2D and symmetric 3D or 2D facial database benchmarks. The main source of limitation of the current distribution relates to the small subject number, which is restricted compared to existing 2D or 3D datasets (Table 1) or formed 3D-2D cohorts (Table 2). However UHDB11 was formed in order to facilitate the development of 3D-2D systems in the presence of non-identity variations, and contains multiple captures of the same subject. In that respect, it can be useful for evaluating light and pose normalization algorithms, on a restricted identity set, with multiple samples per subject on the probe or gallery sets.

Some additional limitations concern the type and variability of acquisition conditions. More challenging and even extreme poses are required for real-life recognition conditions, and these can include unconstrained head rotations, self-occlusions and change in camera viewpoint. UHDB11 involves a single, fixed camera and approximate head rotations in roll and yaw angles. Extensions will include pitch rotations, yaw rotations above  $30^\circ$  and multiple viewpoints for combined face/camera position variation and ground-truth. The acquired, indoor illumination conditions, simulate a small range of lighting conditions and can be further enhanced to include outdoor captures, multiple light sources and varying intensity.

Our on-going efforts focus on addressing these limitations by complementing the database with additional captures, an increased subject and sample size and more challenging acquisition conditions. This involves outdoor sessions with natural lighting conditions and sessions from a custom 21 3D/2D camera acquisition system with multiple reference viewpoints, reconstructions from posed or partial 3D data and an increased subject number (5-10 times) of multiple captures, age groups and ethnic groups.

## 7 Conclusions

We presented a database and evaluation framework for 3D-2D and 2D-3D face recognition, where the gallery and recognition sets have data from different facial sensing modalities. UHDB11 is publicly available [1] and aims to address data requirements regarding face resolution in 2D and 3D and identity-independent variability, in the form of lighting conditions and facial pose. For performance benchmarking under different light/pose combinations, a set of database cohorts and experiments has been designed, on which we provided a 3D-2D face identification and verification baseline at demanding operational conditions (0.001 FAR). The database may additionally be useful for asymmetric, model-based or multimodal face recognition.

## References

1. UHDB11 face database. <http://cbl.uh.edu/URxD/datasets>, 2013.
2. 3dMD. 3D Imaging Systems and Software, Nov. 2012.
3. F. R. Al-Osaimi, M. Bennamoun, and A. S. Mian. Illumination normalization of facial images by reversing the process of image formation. *Machine Vision and Applications*, 22(6):899–911, Nov. 2011.

4. K. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, Jan. 2006.
5. T. C. Faltemier, K. W. Bowyer, and P. J. Flynn. Using multi-instance enrollment to improve performance of 3D face recognition. *Computer Vision and Image Understanding*, 112(2):114–125, Nov. 2008.
6. R. Gross. Face databases. In S. Z. Li and A. K. Jain, editors, *Handbook of Face Recognition*, pages 301–327. Springer New York, 2005.
7. R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807–813, May 2010.
8. D. Huang, M. Ardabilian, Y. Wang, and L. Chen. Oriented gradient maps based automatic asymmetric 3D-2D face recognition. In *Proc. 5th IAPR International Conference on Biometrics*, pages 125–131, New Delhi, India, Mar. 29-Apr. 1 2012.
9. G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the Wild: A database for studying face recognition in unconstrained environments. In *Proc. ECCV Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, Oct. 17-20 2008.
10. P. J. Phillips, H. Moon, S. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, Oct. 2000.
11. P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, Jun. 20-25 2005.
12. P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer. An introduction to the Good, the Bad, and the Ugly face recognition challenge problem. In *Proc. 9th IEEE International Conference on Automatic Face and Gesture Recognition*, Santa Barbara, CA, Mar. 21-25 2011.
13. G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin. Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 30(10):683–697, Oct. 2012.
14. A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3D face analysis. In *Proc. 1st COST 2101 Workshop on Biometrics and Identity Management*, pages 47–56, Roskilde, Denmark, May 7-9 2008.
15. T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, Dec. 2003.
16. G. Toderici, G. Passalis, S. Zafeiriou, G. Tzimiropoulos, M. Petrou, T. Theoharis, and I. A. Kakadiaris. Bidirectional relighting for 3D-aided 2D face recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, Jun. 13-18 2010.
17. V. Vijayan, K. Bowyer, P. Flynn, D. Huang, L. Chen, O. Ocegueda, S. Shah, and I. A. Kakadiaris. Twins 3D face recognition challenge. In *Proc. International Joint Conference on Biometrics*, Washington, DC, Oct. 11-13 2011.
18. L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato. A 3D facial expression database for facial behavior research. In *Proc. 7th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 211–216, Southampton, UK, Apr. 2-6 2006.