UHDB31: A Dataset for Better Understanding Face Recognition across Pose and Illumination Variation

Ha A. Le and Ioannis A. Kakadiaris
Computational Biomedicine Lab
Department of Computer Science, University of Houston
{hale4,ikakadia}@central.uh.edu

Abstract

Face datasets are a fundamental tool to analyze the performance of face recognition algorithms. However, the accuracy achieved on current benchmark datasets is saturated. Although multiple face datasets have been published recently, they only focus on the number of samples and lack diversity on facial appearance factors, such as pose and illumination. In addition, while 3D data have been demonstrated improved face recognition accuracy by a significant margin, only a few 3D face datasets provide high quality 2D and 3D data. In this paper, we introduce a new and challenging dataset, called UHDB31, which not only allows direct measurement of the influence of pose, illumination, and resolution on face recognition but also facilitates different experimental configurations with both 2D and 3D data. We conduct a series of experiments with various face recognition algorithms and point out how far they are from solving the face recognition problem under pose, illumination, and resolution variation. The dataset is publicly available and free for research use[1].

1. Introduction

Challenging face datasets with diverse pose and illumination conditions are indispensable in evaluating face recognition systems. Along with recent advances in deep learning, face recognition accuracy measured on current benchmark datasets is saturated. In fact, automatic face recognition systems based on deep convolutional neural networks surpassed the human level of identifying faces under the well-constrained condition: frontal pose and standard illumination [3]. Face datasets published several years ago (*e.g.*, LFW [15]) are no longer challenging. They are not only limited in the number of samples, they also lack large pose and illumination variations. Several large face datasets have been published recently (*e.g.*, IJB-A [18], MegaFace [17], and MS-Celeb-1M [13]). However, these datasets primarily focus on the number of samples, and they are not coupled with pose or illumination ground-truth, which is essential to analyze the strengths and weaknesses of a face recognition algorithm. A face dataset with precise pose and illumination ground-truth is in great demand to evaluate face recognition algorithms.

In addition to 2D datasets, 3D datasets are also significant to face recognition research. With advances in hardware technology, we may soon witness a revolution in 3D data. Three-dimensional cameras are becoming cheaper and more popular. The structure sensor [24], for example, is an iPad external accessory that can perform a 3D scan of objects or people in just seconds. Smartphones integrated with a 3D camera are under development and will become available in a few years. Several researches have demonstrated that 3D facial data can improve face recognition accuracy by a significant margin [32]. A few 3D face datasets have been published in the past decade, but they have several deficiencies. In the meantime, 3D face reconstruction from 2D images is a solution to compensate for the shortage of 3D data and has been actively investigated in the past few years. Since 3D face reconstruction algorithms require training and testing using both 2D and 3D data, a dataset with both 2D and 3D data is vital.

To further investigate unconstrained face recognition, as well as stimulate research, we have created a challenging dataset, UHDB31. It allows researchers to measure the influence of pose, illumination, and resolution on their algorithms and facilitates different experimental configurations, including 3D-3D, 3D-2D, and 2D-2D[2]. Despite the small number of subjects, UHDB31 provides challenging data samples for face recognition due to its wide variation of poses and illuminations. UHDB31 data samples are equally distributed among 21 different poses and three different illuminations. The pose and illumination conditions of each

[1] UHDB31 can be found at: http://cbl.uh.edu/repository-data

[2] By X-Y (*i.e.* 3D-3D, 3D-2D, and 2D-2D), we refer to the use of X data as gallery and Y data as probe.

data sample are observed and recorded during data acquisition to provide precise ground-truth. In addition, high resolution 2D and 3D data are captured simultaneously to provide symmetric 3D-2D data, which creates a connection between 2D and 3D data. In addition to the raw captured data, four sets of lower resolution images are created by downscaling the original data to evaluate face recognition algorithms under different resolutions. Moreover, twelve manually annotated landmarks of both 2D and 3D data are provided to assess face detection and face alignment algorithms.

Instead of focusing on large number of subjects, UHDB31 focuses on identifying the strengths and weaknesses of a face recognition algorithm based on comprehensive ground-truth. We design three evaluation protocols to separately evaluate pose variation, illumination variation, and the combination variation of both pose and illumination. Each protocol is evaluated on two types of face recognition: 3D-2D and 2D-2D. Extensive experiments are conducted on UHDB31 using four baseline face recognition methods. The experimental results indicate the strengths and weaknesses of each algorithm and show that face recognition under pose and illumination variation is still a challenging problem.

The contributions of the paper are:

- We acquired the data and created a face dataset that allows evaluation of face recognition algorithms across poses, illuminations, and resolutions, and facilitates different experimental configurations, including 3D-3D, 3D-2D, 2D-2D and 3D reconstruction. The dataset is publicly available and free for research use.
- We conducted a set of experiments to identify strengths and weaknesses of several state-of-the-art face recognition algorithms.

The rest of the paper is structured as follows. In Section 2, we provide a brief overview of the existing datasets, including 2D and 3D datasets. Section 3 explains the data acquisition procedure and data specifications in detail. A series of experiments is conducted and evaluated in Section 4. Section 5 offers our conclusions.

2. Related Work

Face recognition is one of the most well-studied topics in computer vision. The advance of face recognition has a close connection with the number of published datasets. In the scope of this paper, we provide a brief overview of well-established datasets in the last decade to highlight their targeted challenges and contrast them to UHDB31. We classify these datasets into two groups: 2D datasets, and 3D datasets. A non-exhaustive list of published datasets can be found in Table 1.

2.1. Two-dimensional Datasets

Two-dimensional datasets are designed for 2D-2D face recognition experiments, where both the gallery and the probe are 2D images. They can be classified into two small groups: uncontrolled and lab-acquired datasets. An uncontrolled dataset provides only images and their corresponding labels. The images are varied in pose, illumination, expression, occlusion and resolution. The most popular uncontrolled dataset is Labeled Faces in the Wild (LFW), which contains 13,233 face images of 5,748 subjects collected from the internet. LFW was published in 2007, and has been a standard benchmark in face recognition for a decade. However, face images in LFW were collected using the Viola-Jones face detector, which can only detect frontal or near-frontal face images. LFW is, therefore, no longer challenging for recent face recognition algorithms. For example, FaceNet [29] achieves 99.63% face verification rate on LFW. Ng *et al*. [23] proposed a method to clean up large data from search queries and created the FaceScrub dataset with 106,863 images of 530 celebrities. Similarly, Yi *et al*. [34] presented the CASIA-Webface dataset with 494,414 images of 10,575 celebrities. VGG-Face [25] dataset was also collected from the internet, but it focuses on the number of samples per subject. On average, VGG-Face has 374.8 images per subject, while CASIA-Webface and FaceScrub have only 46.8 and 201.6 images per subjects, respectively. FaceScrub, CASIA-Webface, and VGG-Face are only used for training since they do not provide an evaluation protocol. Multiple datasets have been proposed along with their face recognition challenges. Klare *et al*. [18] presented IJB-A dataset and two evaluation protocols that support both open-set identification and verification. Although IJB-A includes large variation within an image, it is not a large scale dataset. It contains only 25,813 images of 500 subjects. MegaFace [17] is the first face dataset that contains more than one million images. In the MegaFace challenge, FaceScrub dataset is used as the test set. The difficulty of MegaFace challenge comes from its one million distractors added to the gallery. MegaFace is extended to MF2 [22] dataset, which contains 4,753,320 images of 672,057 subjects. The biggest published dataset is MS-Celeb-1M [13] with 10 million images of 100,000 subjects.

Unlike the aforementioned datasets, UHDB31 is a lab-acquired dataset. A lab-acquired dataset provides not only images and their corresponding labels but also the detailed variation information (*e.g*., pose and illumination). Due to difficulties in setup conditions, few lab-acquired datasets have been published. Multi-PIE [12], the most popular lab-acquired dataset, contains 337 subjects, captured under 15 view points and 19 illumination conditions in four recording sessions for a total of more than 750,000 images. In comparison with Multi-PIE, UHDB31 has a wider pose variation (180° in yaw rotation and 60° in pitch rotation), more

Table 1: Summary of 2D and 3D Datasets with variations in: **P** : Pose, **I**: Illumination, **E**: Expression, **O**: Occlusion, **R**: Resolution, **T**: Twin (the levels of variation are given in parentheses).

Dataset	Year	Group	# Subjects	# Samples	Variation
LFW [15]	2007	2D	5,749	13,233	Uncontrolled
FaceScrub [23]	2014	2D	530	106,863	Uncontrolled
CASIA-Webface [34]	2014	2D	10,575	494,414	Uncontrolled
VGG-Face [25]	2015	2D	2,622	982,803	Uncontrolled
IJB-A [18]	2015	2D	500	25,813	Uncontrolled
MegaFace [17]	2016	2D	690,572	1,027,060	Uncontrolled
MS-Celeb-1M [13]	2016	2D	100,000	10M	Uncontrolled
MF2 [22]	2017	2D	672,057	4,753,320	Uncontrolled
Multi-PIE [12]	2008	2D	337	750,000	P(15), I(19), E(6)
SCface [11]	2009	2D	130	4,160	P(9)
GBU [19]	2010	2D	437	6,510	I(3)
FRGC v2.0 [27]	2005	3D	466	28,049	I(2), E(2)
BU-3DFE [35]	2006	3D	100	7,500	P(2), E(7)
ND-2006 [9]	2007	3D	888	13,450	E(6)
Bosphorus [28]	2008	3D	105	9,250	P(13), E(34), O(4)
CASIA-3D FaceV1 [5]	2008	3D	123	9,248	P(11), I(5), E(5)
Texas 3DFRD [14]	2010	3D	118	4,596	N/A
3D-TEC [31]	2011	3D	214	856	E(2), T(214)
UMB-DB [6]	2011	3D	143	2,946	E(4), O(6)
Florence Faces [2]	2011	3D	53	212+	P(3)
UHDB11 [30]	2013	3D	23	3,312	P(12), I(6)
UHDB31 (Ours)	**2017**	**3D**	**77**	**25,872**	**P(21), I(3), R(5)**

precise pose information, and free for research use. SCface [11] is a dataset of 4,160 face images of 130 subjects, in which 1,170 images were captured under nine view points and the standard illumination condition. Another 2,990 images were acquired in uncontrolled indoor environment using five video surveillance cameras of various qualities. Compared to SCface, UHDB31 has more pose and illumination variations. GBU [19] contains 6,510 images of 437 subjects. GBU face images were captured under frontal pose, but with three different lighting conditions: indoor, outdoor and ambient. Compared to GBU, UHDB31 varies not only in illumination but also in pose.

2.2. Three-dimensional Datasets

Three-dimensional datasets, which also called multi-modal datasets, are designed for multiple experimental configurations, including 3D-3D, 3D-2D, 2D-2D and 3D reconstruction. The quality of 3D data depends on the 3D sensor. Multi-modal datasets have a clear transition before and after the release of Kinect, a low-cost small-size acquisition device, in year 2010.

Before the release of Kinect, multi-modal datasets focused on the quality of 3D facial models. FRGC v2.0 [27] is the most popular 3D face dataset in the literature. It contains 4,007 facial models and 24,042 images of 466 subjects. These data were captured in 4,007 sessions, and each session contains one facial model, four controlled images, and two uncontrolled images. All images were captured under two expressions and two illuminations. FRGC v2.0 was later extended to the ND-2006 [9] dataset, which contains 13,450 facial models of 888 subjects. BU-3DFE [35] is the first 3D face dataset focusing on facial expression. It includes 100 subjects with 2,500 facial expression models and 2,500 two-view texture images. Each subject performed seven expressions in front of a 3D face scanner, and two corresponding facial texture images captured at two view-points ($\pm 45°$) were associated with each expression shape model. The Bosphorus [28] dataset represents a new comprehensive multi-expression, multi-pose 3D face data enriched with realistic occlusions. The dataset is a combination of 13 poses, 34 expressions, and four occlusions. In total, it contains 4,625 pairs of facial models and images of 105 subjects. Similarly, the UHM-DB [6] dataset was focused on partial occlusions. It contains 1,473 pairs of facial models and images of 143 subjects. Each subject was acquired with four different facial expressions, and with the face partially occluded by various objects such as eyeglasses, hats, scarves and hands. CASIA-3D FaceV1 [5] consists of 4,624 pairs of facial models and images of 123 persons. The dataset is a combination of 11 poses, five illuminations, and five expressions. Texas 3DFRD [14] contains 1,149 pairs of high resolution, pose normalized, preprocessed, and perfectly aligned color and range images of 118 subjects. The 3D-TEC dataset [31] contains 3D facial data for 107 pairs of identical twins. The Florence Faces dataset [2] consists of more than 212 facial models and 159 video sequences of 53 subjects. Each subject is scanned four or five times, including two frontal scans, two profile scans and a scan with glasses if present. UHDB11 [30] consists of 1,625 pairs of facial models and images of 23 subjects, and each subject was captured 72 times: 12 poses and six illuminations. In comparison with these 3D datasets, UHDB31 contains 1,617 facial models and 24,255 images of 77 subjects captured under 21 poses, three illuminations, and five resolutions.

After the release of Kinect, multiple 3D datasets were acquired using Kinect (*e.g.*, FaceWarehouse [4], and Kinect-FaceDB [21]). A complete list of 3D datasets acquired using Kinect can be found in [10]. Due to the low precision in depth data, 3D facial models of these datasets do not contain detailed facial geometries (*e.g.*, wrinkles, and eyelids). Unlike these datasets, UHDB31 provides high quality 3D data.

3. The UHDB31 Dataset

3.1. Data Acquisition

UHDB31 data were captured using a 3dMD system [1] that includes seven modular units of 21 cameras and an industrial-grade flash system. All the cameras are arranged in a half-ellipse shape with the semi major axis and semi minor axis of 1.2 *m* and 1 *m*, respectively. The participants were asked to sit in a chair placed in the center of the ellipse and look straight forward at the center camera. In a single capture, 21 pairs of a 180° facial model and a high resolution images are automatically generated by a single coordinate system from all synchronized stereo pairs.

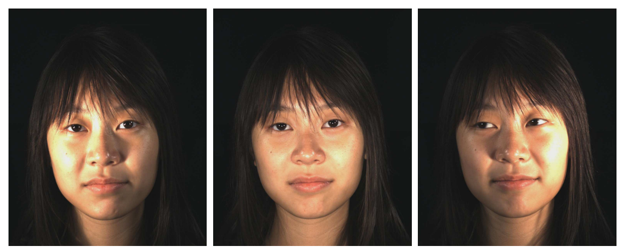
Indoor illumination variations are obtained using multiple diffuse light sources, from incandescent light bulbs with an approximate color temperature of 2,800K. Three indoor lighting conditions are created by altering the light sources in three directions, left, center, and right, from the subject perspective. Figure 1 depicts our data acquisition system.

3.2. Data Specifications

Seventy-seven subjects, 53 males, and 24 females, of different ethnic groups, participated in our data acquisition, and each subject was captured three times under three lighting conditions. Figure 2 illustrates three illumination variations. Each capture generates 21 pairs of samples of different poses. Since the captured 3D facial models are invariant

Figure 1: Depiction of the data acquisition system.

to the illumination variation, only one set of 3D facial models is retained. Thus, UHDB31 contains 1,617 facial models and 4,851 raw images in total.



(a) (b) (c)

Figure 2: Illumination variations in 2D images. (a-c) Images acquired with the light positioned to the left, in front, and to the right of the subject, respectively.

Since 21 cameras are well-aligned on a grid of size 3×7 and cover 180° around the subject, the precise pose ground-truth is measured directly from the camera viewpoints. In particular, three pitch rotations are evenly distributed from $-30°$ to $30°$ while seven yaw rotations are evenly distributed from $-90°$ to $90°$. An example of pose variations is shown in Table 2.

The capture speed of the 3dMD system is approximately 1.5 *ms*, with the geometry accuracy of less than 0.2 *mm* RMS. However, the system needs more than 9 *s* to render a high quality 3D facial model. The rendering is processed offline. In average, a facial model has 25,000 vertices and 49,500 triangles. Table 3 shows 21 facial models corresponding to 21 pose images in Table 2.

High resolution 2D images of size 2,048 × 2,448 pixels are saved in the BMP file format. To evaluate face recognition algorithms on a variety of image qualities, four other

Table 2: Depiction of the pose variations in 2D images. Rows correspond to seven yaw rotations, and columns correspond to three pitch rotations.

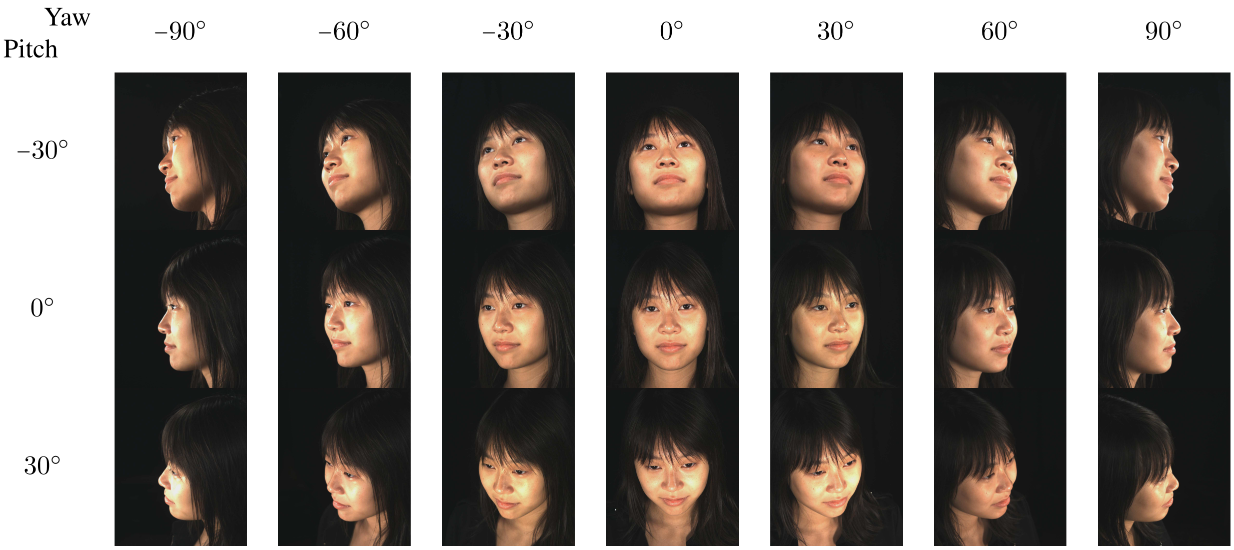
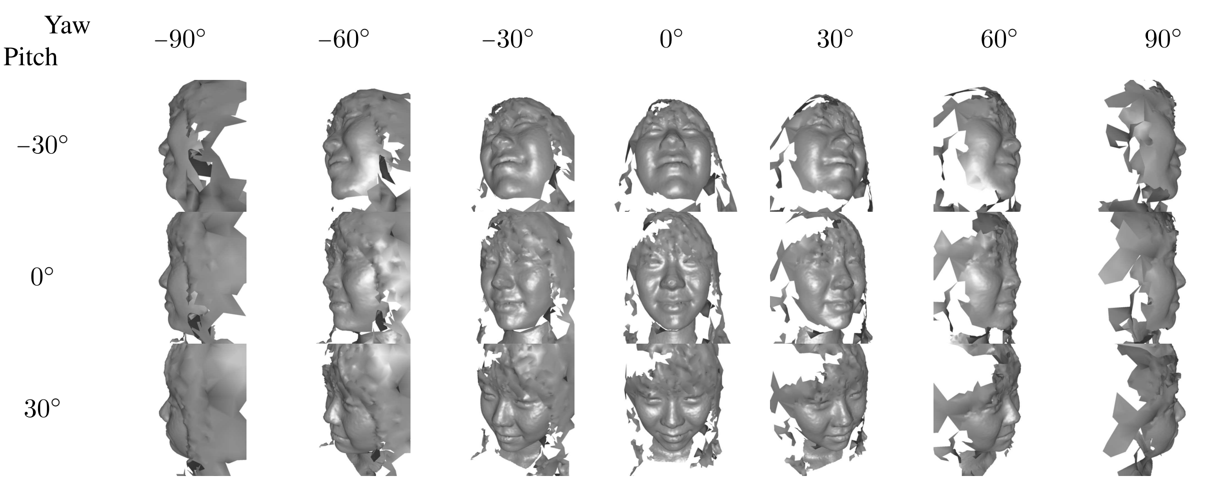


Table 3: The 3D facial models corresponding to the 2D images in Table 2.



sets of data are created by down-sampling the raw images. Five subsets of different resolutions and average inter-pupil distances (IPDs) are summarized in Table 4. The mean and standard deviation of IPDs are computed using manually annotated landmarks presented in the following subsection. Figure 3 shows the distribution of IPDs on raw images.

3.3. Facial Landmarks

Facial landmark localization, also known as face alignment, is an essential module in a face recognition system. To facilitate the evaluation of facial landmark localization algorithms, we manually annotate 12 landmarks for each 2D image and 3D facial model. The 12 landmarks consist of four inner and outer eye corners, a nose tip, four inner and outer corners of two nostrils, a nose middle and two mouth corners. Figure 4 shows examples of 12 landmarks superimposed on a 2D image and a 3D facial model.

4. Evaluations

4.1. Evaluation Protocols

The multiple types of data and the wide range of poses, illuminations, and resolutions of UHDB31 enable different experimental configurations on face recognition. Three

Table 4: Five subsets of UHDB31 with five different resolutions. IPD: mean and standard deviation of inter-pupil distances.

Subset	Resolution (pixels)	IPD (pixels)
UHDB31.R2048	2,048 $\times$ 2,448	398.41 $\pm$ 152.14
UHDB31.R1024	1,024 $\times$ 1,224	199.20 $\pm$ 76.07
UHDB31.R0512	512 $\times$ 612	99.61 $\pm$ 38.03
UHDB31.R0256	256 $\times$ 306	49.82 $\pm$ 19.01
UHDB31.R0128	128 $\times$ 153	24.91 $\pm$ 9.50

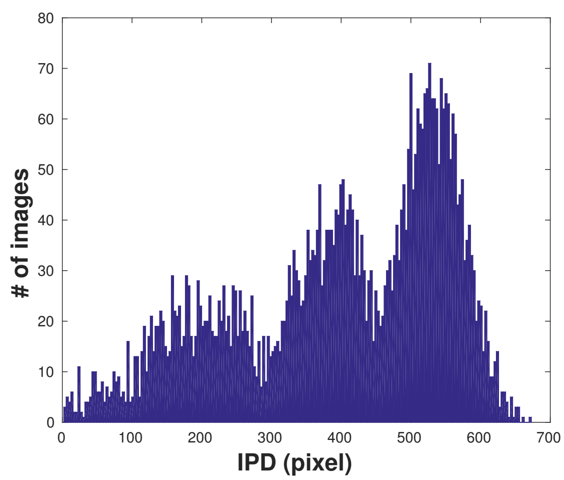
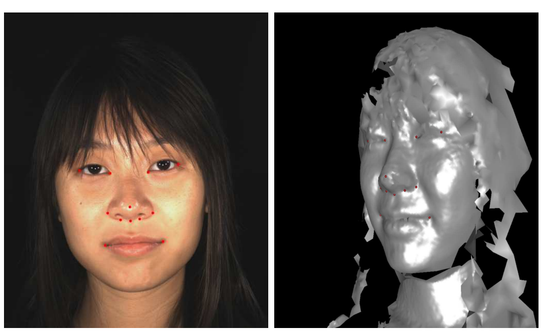


Figure 3: Distribution of IPDs on raw images.



(a) (b)

Figure 4: Twelve manually annotated landmarks superimposed in: (a) a 2D image, and (b) a 3D facial model.

evaluation protocols have been defined, including pose, illumination, and combination, as the standard benchmark of UHDB31. Each protocol consists of two types of experiments: 2D-2D and 3D-2D. Each experiment is conducted on five resolution subsets separately to evaluate the impact of image quality on face recognition methods. The detailed descriptions of three evaluation protocols are described in Table 5, in which P01 - P21 denote the 21 poses, P11 is the frontal pose, and I01, I03, and I05 denote illumination from the left, center and right light sources, respectively. The subset of P11 (frontal pose) and I03 (center light) is used as the gallery in both 3D-2D and 2D-2D experiments. However, 77 facial models corresponding to P11 are used as a complement to the gallery in 3D-2D experiments. The subsets of P01 - P21 and I03 are used as the probes in the pose evaluation protocol, while the subsets of P11 and two illuminations I01 and I05 are used as the probes in the illumination evaluation protocol. The probes used in the combination evaluation protocol are the subsets of all combination of 21 poses and three illuminations. The subset used as the gallery is removed from probes.

Besides the experimental configurations presented in Table 5, UHDB31 can be used for 3D-3D or partial 3D-3D face recognition. Although the primary purpose of UHDB31 is face recognition evaluation, we can also apply the dataset in different evaluation tasks, including face detection, landmark localization, 3D pose reconstruction, and gender estimation.

4.2. Baseline Methods

For 2D-2D experiments, three face recognition methods, including a commercial off-the-shelf (COTS) software, a well-known face recognition system VGG-Face [25], and a face recognition pipeline UR2D, are used as the baseline methods. The UR2D pipeline is a combination of five state-of-the-art modules, including face detection using DPM Headhunter [20], 3D face reconstruction using UH-E2FAR [7], pose estimation using GoDP [33], texture lifting using the method proposed by Kakadiaris *et al.* [16], and feature extraction using PRFS [8].

For 3D-2D experiments, a face recognition pipeline UR3D2D is used as the baseline method. UR3D2D pipeline has the same modules as UR2D. However, UR3D2D directly uses 3D facial models for pose estimation and texture lifting on the gallery images.

The code for VGG-Face was downloaded from the author's website [26]. Cosine distance is used for matching since it provides better results than L_2 distance in all three systems.

4.3. Evaluation Results

In this section, we report the results of pose protocol on UHDB31 using the baseline face recognition methods and discuss the influence of pose and resolution on face recognition performance. In all the following tables, we drop the percentage sign for simplicity.

Table 6 shows the rank-1 identification rates of four baseline methods evaluated on pose protocol. The rank-1 iden-

Table 5: Summary of evaluation protocols: P01 - P21 denote 21 poses corresponding to 21 poses in Table 2 (following the order top to bottom, then left to right). P11 is the frontal pose; I01, I03 and I05 denote illumination from the left, center and right light sources, respectively. A pair of samples includes a 2D image and its corresponding 3D facial model.

Protocol	Type	Gallery			Probes		
		Pose	Light	# Samples	Poses	Lights	# Samples
Pose	2D-2D	P11	I03	77	P01 - P10, P12 - P21	I03	1,540
	3D-2D	P11	I03	77 pairs	P01 - P10, P12 - P21	I03	1,540
Illumination	2D-2D	P11	I03	77	P11	I01, I05	154
	3D-2D	P11	I03	77 pairs	P11	I01, I05	154
Combination	2D-2D	P11	I03	77	P01 - P10, P12 - P21	I01, I05	3,080
	3D-2D	P11	I03	77 pairs	P01 - P10, P12 - P21	I01, I05	3,080

Table 6: Rank-1 identification rates of four baseline methods evaluated on the pose protocol. The four numbers in each cell of the table are in the following order: COTS, VGG-Face, UR2D and UR3D2D. The best result for each pose is highlighted in bold.

Pitch \ Yaw	−90°	−60°	−30°	0°	30°	60°	90°
30°	3.9/30.4/ 57.9/**58.7**	41.8/81.0/ **94.8/94.8**	78.4/**99.0/** 98.2/98.4	89.1/96.9/ 98.7/**99.0**	80.0/**99.2/** 98.4/98.7	44.7/86.5/ 95.6/**95.8**	11.9/45.5/ 55.6/**56.4**
0°	9.9/64.9/ 86.5/**86.8**	54.5/95.8/ 99.2/**99.5**	92.2/**100.0/** 99.7/**100.0**	-	90.6/**100.0/** 98.7/99.0	66.5/99.5/ 99.5/**99.7**	13.5/76.1/ **87.8/87.8**
−30°	2.9/30.7/ **56.9/56.9**	14.3/81.8/ 93.5/**93.8**	65.2/94.6/ 98.7/**99.0**	91.4/97.1/ 99.7/**100.0**	81.6/97.4/ 98.7/**99.0**	19.2/84.7/ 97.7/**97.9**	5.5/36.1/ 68.3/**68.8**

tification rates are saturated on near-frontal poses. VGG-Face, UR2D and UR3D2D achieve 98%, 98.9%, and 99.1% rank-1 identification rates on an average of nine near-frontal poses (P07 - P15), respectively. The COTS achieves only 47.9% on average because the software fails to detect large pose faces. Since both UR2D and UR3D2D utilize a 3D face reconstruction module, they achieve better results than VGG-Face in terms of pose. By using 3D facial models, UR3D2D improves UR2D result slightly.

Image quality also affects the performance of face recognition algorithms. As shown in the Fig. 5, rank-1 identification rates of baseline methods are significantly increased when the image resolution is changed from 128 × 153 pixels to 256 × 306 pixels. The increment is saturated when the image resolution reaches the size of 1,024 × 1,224 pixels. Face recognition methods based on deep convolutional neural networks require rescaling the input face image into a predefined size (*e.g.*, VGG-Face uses the input image size of 224 × 224 pixels). When the face size is larger than the predefined size of each face recognition method, the performance stops improving.

In summary, the experimental results allow us to identify the strengths and weaknesses of four baseline methods. We hypothesize that one of the reasons the COTS has poor results on UHDB31 because its face detection module fails

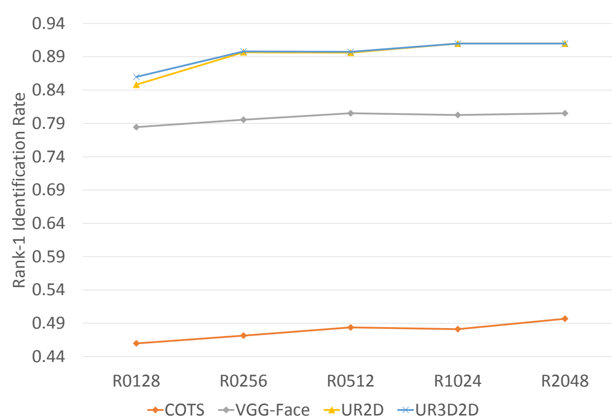


Figure 5: The growth of rank-1 identification rates along with the increase of image resolution.

on large face poses. VGG-Face proves their performance on near-frontal poses. Both UR2D and UR3D2D have verified their strengths on large face poses in the evaluation on the pose protocol due to the improvement of the 3D reconstruction method [7]. Face recognition performance is improved if we use 3D facial models as a complement for gallery images. However, the improvement is not distinctly shown in our experiments. It appears that UH-E2FAR module estimates an accurate 3D model to be used for recognition.

5. Conclusion

In this paper, we presented UHDB31, a face dataset with two types of data and rich ground-truth information, including poses, illuminations, resolutions and landmarks. Additionally, we defined three evaluation protocols for both 3D-2D and 2D-2D face recognition. The evaluation protocols allow identification of the strengths and weaknesses of face recognition algorithms under pose, illumination, and resolution variations. Our experiments with four baseline methods demonstrated that face recognition performance is limited under large face pose. In addition to 3D-2D and 2D-2D face recognition, UHDB31 can also be applied to other evaluation tasks.

UHDB31 has the follow limitations: (i) small number of subjects, (ii) limited number of lighting conditions, absence of (iii) facial expressions and (v) occlusion. In our future work, we will focus on addressing UHDB31 limitations by complementing the dataset with additional data captured from more subjects and multiple acquisition conditions, especially outdoor lighting conditions.

6. Acknowledgement

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2015-ST-061-BSH001. This grant is awarded to the Borders, Trade, and Immigration (BTI) Institute: A DHS Center of Excellence led by the University of Houston, and includes support for the project "Image and Video Person Identification in an Operational Environment" awarded to the University of Houston. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

References

[1] 3dMD LLC. 3dMD (http://www.3dmd.com), 2004. 4

[2] A. D. Bagdanov, A. D. Bimbo, and I. Masi. The Florence 2D/3D Hybrid Face Dataset. In *Proc. Joint ACM Workshop on Human Gesture and Behavior Understanding*, pages 79–80, Scottsdale, AZ, Dec. 1 2011. 3, 4

[3] A. Blanton, K. C. Allen, T. Miller, N. D. Kalka, and A. K. Jain. A Comparison of Human and Automated Face Verification Accuracy on Unconstrained Image Sets. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 229–236, Las Vegas, NV, Jun. 26 - Jul. 1 2016. 1

[4] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou. FaceWarehouse: A 3D Facial Expression Database for Visual Computing. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):413–425, Nov. 04 2014. 4

[5] Chinese Academy of Sciences Institute of Automation. CASIA-3D FaceV1 (http://biometrics.idealtest.org), 2008. 3

[6] A. Colombo, C. Cusano, and R. Schettini. UMB-DB: A database of partially occluded 3D faces. In *Proc. IEEE International Conference on Computer Vision Workshops*, pages 2113–2119, Barcelona, Spain, Nov. 6-13 2011. 3

[7] P. Dou, S. K. Shah, and I. A. Kakadiaris. End-to-end 3D face reconstruction with deep neural networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, Jul. 21-26 2017. 6, 7

[8] P. Dou, L. Zhang, Y. Wu, S. K. Shah, and I. A. Kakadiaris. Pose-robust face signature for multi-view face recognition. In *Proc. IEEE 7^{th} International Conference on Biometrics Theory, Applications and Systems*, pages 1–8, Arlington, VA, Sep. 8-11 2015. 6

[9] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn. Using a Multi-Instance Enrollment Representation to Improve 3D Face Recognition. In *Proc. IEEE First International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, Crystal City, VA, Sep. 27-29 2007. 3

[10] M. Firman. RGBD Datasets: Past, Present and Future. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 661–673, Las Vegas, NV, June 26 - July 1 2016. 4

[11] M. Grgic, K. Delac, and S. Grgic. SCface surveillance cameras face database. *Multimedia Tools and Applications*, 51(3):863–879, feb 2011. 3

[12] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807–813, 2010. 2, 3

[13] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. MS-Celeb-1M: Challenge of Recognizing One Million Celebrities in the Real World. In *Proc. European Conference on Computer Vision*, pages 87–102, Amsterdam, The Netherlands, Oct. 8-16 2016. 1, 2, 3

[14] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik. Texas 3D Face Recognition Database. In *Proc. IEEE Southwest Symposium on Image Analysis & Interpretation*, pages 97–100, Austin, TX, May 23-25 2010. 3, 4

[15] 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. Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, Oct. 17-20 2008. 1, 3

[16] I. A. Kakadiaris, G. Toderici, G. Evangelopoulos, G. Passalis, D. Chu, X. Zhao, S. K. Shah, and T. Theoharis. 3D-2D face recognition with pose and illumination normalization. *Computer Vision and Image Understanding*, 154:137–151, Jan 2017. 6

[17] I. Kemelmacher-Shlizerman, S. Seitz, D. Miller, and E. Brossard. The MegaFace Benchmark: 1 Million Faces for Recognition at Scale. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 4873–4882, Las Vegas, NV, June 26 - July 1 2016. 1, 2, 3

[18] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. Burge, and A. K. Jain. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 1931–1939, Boston, MA, Jun. 7-12 2015. 1, 2, 3

[19] Y. M. Lui, D. Bolme, P. J. Phillips, J. R. Beveridge, and B. A. Draper. Preliminary studies on the Good, the Bad, and the Ugly face recognition challenge problem. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 9–16, Providence, RI, Jun. 16-21 2012. 3

[20] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. Face Detection without Bells and Whistles. In *Proc. 13^{th} European Conference on Computer Vision*, pages 720–735, Zurich, Switzerland, Sep. 6-12 2014. Springer, Cham. 6

[21] R. Min, N. Kose, and J.-L. Dugelay. KinectFaceDB: A Kinect Database for Face Recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(11):1534–1548, Jul. 28 2014. 4

[22] A. Nech, I. Kemelmacher-Shlizerman, and P. G. Allen. Level Playing Field for Million Scale Face Recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, Jul. 21-26 2017. 2, 3

[23] H.-W. Ng and S. Winkler. A data-driven approach to cleaning large face datasets. In *Proc. IEEE International Conference on Image Processing*, pages 343–347, Paris, Franch, Oct. 27-30 2014. 2, 3

[24] Occipital. Structure Sensor (`https://structure.io`), 2016. 1

[25] O. M. Parkhi, A. Vedaldi, A. Zisserman, A. Vedaldi, K. Lenc, M. Jaderberg, K. Simonyan, A. Vedaldi, A. Zisserman, K. Lenc, and Others. Deep face recognition. In *Proc. British Machine Vision Conference*, pages 1–12, Swansea, UK, Sep. 7-10 2015. 2, 3, 6

[26] O. M. Parkhi, A. Vedaldi, A. Zisserman, A. Vedaldi, K. Lenc, M. Jaderberg, K. Simonyan, A. Vedaldi, A. Zisserman, K. Lenc, and Others. VGG-Face (`http://www.robots.ox.ac.uk/~vgg/software/vgg_face`), 2015. 6

[27] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 947–954, San Diego, CA, Jun. 20-25 2005. 3

[28] A. Savran, N. Alyüz, H. Dibekliolu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus Database for 3D Face Analysis. In *Biometrics and Identity Management*, pages 47–56. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. 3

[29] F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A unified embedding for face recognition and clustering. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 07-12-June, pages 815–823, Boston, MA, Jun. 7-12 2015. 2

[30] G. Toderici, G. Evangelopoulos, T. Fang, T. Theoharis, and I. A. Kakadiaris. UHDB11 database for 3D-2D face recognition. In *Proc. 6^{th} Pacific-Rim Symposium on Image and Video Technology*, pages 73–86, Guanajuato, Mexico, Oct. 28 - Nov. 1 2013. 3, 4

[31] V. Vijayan, K. Bowyer, and P. Flynn. 3D Twins and Expression Challenge. In *Proc. IEEE International Conference on Computer Vision Workshops*, pages 2100–2105, Barcelona, Spain, Nov. 6-13 2011. 3, 4

[32] Y. Wu, S. K. Shah, and I. A. Kakadiaris. Rendering or normalization? An analysis of the 3D-aided pose-invariant face recognition. In *Proc. IEEE International Conference on Identity, Security and Behavior Analysis*, pages 1–8, Sendai, Japan, Feb. 29 - Mar. 2 2016. 1

[33] Y. Wu, S. K. Shah, and I. A. Kakadiaris. GoDP: Globally optimized dual pathway system for facial landmark localization in-the-wild. *arXiv preprint*, 2017. 6

[34] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning Face Representation from Scratch. *arXiv preprint*, nov 2014. 2, 3

[35] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato. A 3D Facial Expression Database For Facial Behavior Research. In *Proc. IEEE 7^{th} International Conference on Automatic Face and Gesture Recognition*, pages 211–216, Southampton, UK, Apr. 10-12 2006. 3