Bidirectional relighting for 3D-aided 2D Face Recognition

G. Toderici[1], G. Passalis[1,2], S. Zafeiriou[4], G. Tzimiropoulos[4],
M. Petrou[3,4], T. Theoharis[1,2], I.A. Kakadiaris[1]

[1]Computational Biomedicine Lab, Dept. of CS, ECE, Biomedical Engineering, Univ. of Houston, Houston, TX, USA

[2]Computer Graphics Lab, University of Athens, Greece   [3]ITI-CERTH, Greece

[4]Department of Electrical and Electronic Engineering, Imperial College London, UK

Abstract

In this paper, we present a new method for bidirectional relighting for 3D-aided 2D face recognition under large pose and illumination changes. During subject enrollment, we build subject-specific 3D annotated models by using the subjects' raw 3D data and 2D texture. During authentication, the probe 2D images are projected onto a normalized image space using the subject-specific 3D model in the gallery. Then, a bidirectional relighting algorithm and two similarity metrics (a view-dependent complex wavelet structural similarity and a global similarity) are employed to compare the gallery and probe. We tested our algorithms on the UHDB11 and UHDB12 databases that contain 3D data with probe images under large lighting and pose variations. The experimental results show the robustness of our approach in recognizing faces in difficult situations.

1. Introduction

Face recognition is one of the most widely researched topics in computer vision due to a wide variety of applications that require identity management. Most existing face recognition studies are focused on 2D images with nearly-frontal-view faces and constrained illumination; however, 2D facial images are affected by varying illumination conditions and changes in pose. Thus, although these existing methods are able to provide satisfactory performance under constrained conditions, they are challenged by unconstrained pose and illumination conditions.

FRVT 2006 explored the feasibility of using 3D data for both enrollment and authentication [11]. The algorithms using 3D data have demonstrated their ability to provide good recognition rates; however, for practical purposes, it is unlikely that large-scale deployments of 3D systems will take place in the near future due to the high cost of the necessary hardware. Nevertheless, it is not unreasonable to assume that an institution may want to invest in a limited number of 3D scanners, if having 3D data for enrollment can yield higher accuracy for 2D face authentication/identification.

In this paper, we propose a face recognition method which makes use of 3D face data for enrollment, while requiring only 2D data for authentication. During enrollment, unlike the existing methods (e.g., [3]) that use a 2D image to infer a 3D model in the gallery, we use 2D+3D data (2D texture plus 3D shape) to build subject-specific annotated 3D models. To achieve this, we first fit an Annotated Face Model (AFM) to the raw 2D+3D data using a subdivision-based deformable framework. A geometry image representation is extracted using the UV parameterization of the model. In the authentication phase, we use a single 2D image as the input to map the subject-specific 3D AFM. Given the pose in the 2D image, an Analytical Skin Reflectance Model (ASRM) is then applied to the gallery AFM in order to transfer the lighting from the probe to the texture in the gallery. The matching score is computed using the relit gallery texture and the probe texture.

Our contributions are the following: (i) using 2D+3D data to build a subject-specific 3D model during enrollment, which is able to more accurately characterize the subject identity than the existing 2D/3D methods which use a 2D image to infer the 3D gallery model, (ii) a bidirectional face *relighting* algorithm which allows us to achieve better face recognition performance than the traditional *unlighting* methods, (iii) a new view-dependent distance metric, and iv) a new correlation-based distance metric. We used UHDB11 and UHDB12 because other publicly available 3D face databases, including FRGC v.2, do not have probe images under both lighting variation and pose variation.

The rest of this paper is organized as follows: in Section 2 we briefly review related methods, in Section 3 we present the methods for enrollment using 2D+3D data and for authentication using 2D images, while in Section 4 we provide qualitative and quantitative results. We provide our conclusions in Section 5.

2. Related Work

The literature in 3D and 2D+3D Face Recognition has rapidly increased in recent years. An excellent survey was given by Bowyer *et al*. [4]. The most closely related work

is by Riccio and Dugelay [14], who proposed to use geometric invariants on the face in order to establish a correspondence between the 3D gallery face and the 2D probe. Some of the invariants were manually selected. This algorithm did not utilize the texture information registered with the 3D data from the scanner, thus did not take full advantage of the input data. Blanz and Vetter [3] employed a morphable model technique in order to acquire the geometry and texture of faces from 2D images. Wang *et al.* [20] used a spherical harmonic representation [1] with the morphable model for 2D face recognition. In contrast to our method that uses 2D+3D data to build a 3D subject-specific model for the gallery, their methods used a 2D image to build a 3D model for the gallery based on a 3D statistical morphable model. Yin and Yourst [23] used frontal and profile 2D images to construct 3D shape models. Compared with these methods, our approach is able to more accurately model the subject identity due to the use of more information (2D+3D). Smith and Hancock [15] presented an approach for albedo estimation from 2D images also based on a 3D morphable model. The normals of the fitted model were then used for the computation of shading, assuming a Lambertian reflectance model. Biswas *et al.* [2] proposed a method for albedo estimation for face recognition using two-dimensional images. However, their approach made the assumption that the image did not contain shadows, and did not handle specular light. The relighting approach of Lee *et al.* [8] also suffers from the self-shadowing problem. Zhou *et al.* [24] use nearest-subspace patch matching to warp near frontal face images to frontal and project this face image into a pre-trained low-dimensional illumination subspace. This method requires training of patches in many different illumination conditions. Huang *et al.* [5] extract histograms of Local Binary Patterns (LBP) from 2D gallery and probe. LBP is also extracted from the gallery 3D range image. Then, the LBP histograms are compared using Chi-square distance;,while the LBP gallery image is matched to the LBP probe image using Canonical Correlation Analysis (CCA). Their experimental results were limited to frontal pose and a small database.

The proposed method, having significantly fewer constraints and limitations than previous approaches, widens the applicability of such methods. Especially when compared with relighting methods designed for face recognition (e.g., [17]), the proposed method offers significantly higher visual quality in cases where specular highlights over-saturate the images.

3. Methods

3.1. 2D+3D Enrollment

The enrollment pipeline (Algorithm 1) is depicted in Fig. 1. We employ the Annotated Face Model (AFM) proposed by Kakadiaris *et al.* [6] to generate geometry im-

ages (regularly sampled 2D images with three channels) encoding geometric information (x, y and z components of a vertex in $\mathbb{R}^3$). In this paper, the number of channels in the geometry image is seven (three channels for representing the actual geometry of the face, three for representing the texture information, and one for the visibility map). Specifically, we first fit the AFM to the input 3D data [6].

Algorithm 1 Enrollment with 3D data

Input: 3D facial mesh, 2D facial image, subject ID.

1. Pre-process the 3D facial mesh.
2. Register AFM to the 3D facial mesh.
3. Fit AFM to 3D facial mesh.
4. Lift texture from the 2D facial image using the fitted AFM.
5. Compute visibility map.
6. Store the fitted AFM, texture, and visibility map in the gallery as metadata for subject ID.

Once the fitting is complete, we represent the AFM as a geometry image. For each vertex in the geometry image, we compute the closest point on the data. The texel corresponding to this point in the data is used to create the corresponding texture image for the fitted AFM. Additionally, we compute a visibility map. If the closest point on the data does not have a valid texel assigned (i.e., if the 3D point was not visible to the 2D image sensor), we assign a value one to the corresponding location in the visibility map. Otherwise, we assign a value of zero.
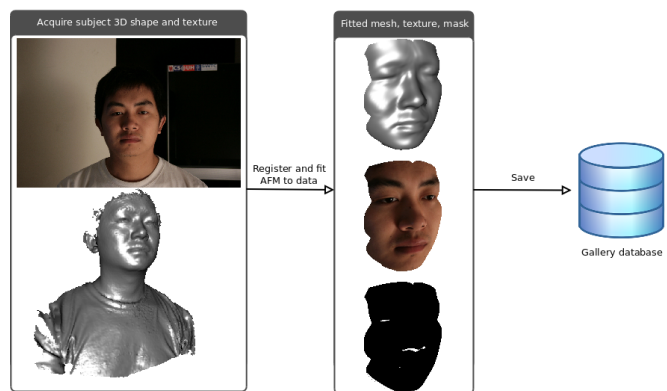

Figure 1. Depiction of the enrollment procedure for our 3D-aided 2D face recognition system. First, subject's 3D and 2D data are acquired. Then, the fitted AFM (mesh, texture, and visibility mask) are computed. Finally, the extracted data are saved in the gallery.

3.2. 2D Authentication

In the authentication stage (Algorithm 2), the input to our method is a 2D image. Seven fiducial landmarks (two eye
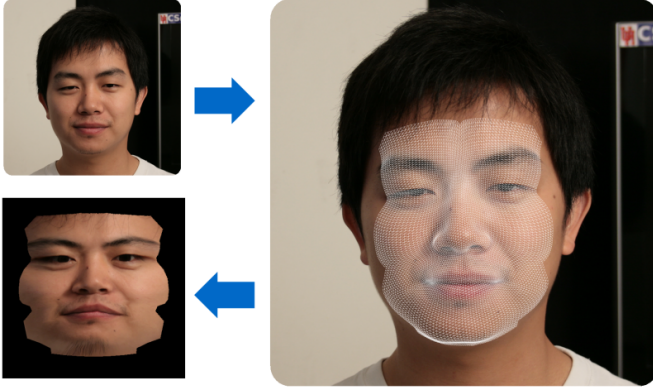
Figure 2. Converting raw 2D images to textures in the geometry image space: Raw 2D image $\rightarrow$ Fitted AFM of the same subject registered and superimposed over the image $\rightarrow$ Image converted to texture in geometry image space. The conversion is done by matching a set of landmarks on the AFM and on the 2D image.

Algorithm 2 Authentication using 2D images

Input: 2D facial image and claimed subject ID.

1. Retrieve "claimed ID" AFM from the gallery.
2. Locate the seven landmarks on the 2D facial image.
3. Register the AFM to the 2D facial image using the corresponding landmarks (Fig. 2).
4. Compute the visibility map.
5. Bidirectionally relight the enrollment 2D facial texture to match the probe 2D facial texture.
6. Compute the CWSSIM and GS (Section 3.2.3) scores between the relit texture and the probe texture.
7. Threshold the score to make an ACCEPT/REJECT decision.

---

inner corners, two eye outer corners, nose tip, and two nose corners) are manually labelled. Once the pose is estimated (using these landmarks and their corresponding locations on the AFM along with camera information acquired before the capture of the 2D image), the texture is mapped onto the AFM (Fig. 2). We use an analytical skin reflectance model to bidirectionally relight the gallery texture using the stored AFM mesh, in order to match the illumination of the probe texture (Fig. 3).

3.2.1 Analytical Skin Reflectance Model

We use a hybrid bidirectional reflectance distribution function (BRDF) to model skin reflectance. We did not employ a bidirectional surface scattering reflection distribution function (BSSRDF) model since the test data did not have sufficient resolution to estimate a subsurface scattering component. The ASRM uses the Lambertian BRDF to model the diffuse component and the Phong BRDF to model the specular component. The Lambertian BRDF is

the simplest, most widely used, physics-based model for diffuse reflectance. The model assumes that the surface is equally bright from all directions. The intensity of the light at a surface point is proportional to the angle between surface normal and incident light directions (denoted as θ) $I_d = E \cos \theta$, where E is the intensity of the light source. The Lambertian BRDF does not take into account the specular reflections caused by the oily layer of the skin. To accommodate this we use the BRDF proposed by Phong [12]. The intensity given by the specular reflection at a surface point is $I_s = E \cos^n \phi$, where ϕ is the angle between the view vector and the reflected light and n is a parameter that controls the size of the highlight. Note that each facial area has different specular properties, therefore we use a specular map based on the annotation of the AFM (for details see [6]).

3.2.2 Bidirectional Relighting

The illumination parameters and the ASRM can be optimized in two different ways: estimate the albedo (unlighting, see our past study [16]) and transfer illumination (relighting). In both cases the texture must be represented in the AFM's UV space.

Generally, the texture M_T is the result of the lighting applied on the unknown albedo M_A and is given by: $M_T = I_s + (I_d + I_a) \cdot M_A$, where I_a is the ambient component, I_d the diffuse component and I_s the specular component (assuming white specular highlights). Solving this equation for the albedo yields: $M_A = \frac{M_T - I_s}{I_d + I_a}$. However, for many practical applications, the albedo itself is not required, and is used only as an intermediate step for relighting. We advocate the use of bidirectional relighting without first estimating the albedo. This means that the optimization directly estimates the parameters for two lights (one that *removes* the illumination from the gallery image and one that *adds* the illumination from the probe image). The goal is to match the illumination conditions of a gallery texture to that of a probe texture. The following metric is minimized:

$$\mathcal{D} = \left| \mathcal{M}'_\mathcal{T} - \mathcal{I}'_f - (\mathcal{I}'_\lceil + \mathcal{I}'_\dashv) \frac{\mathcal{M}_\mathcal{T} - \mathcal{I}_f}{\mathcal{I}_\lceil + \mathcal{I}_\dashv} \right|, \quad (1)$$

where I_a, I_d, and I_s are the parameters of the light illuminating the gallery; I'_a, I'_d and I'_s are the parameters of the second light illuminating the probe, while M'_T is the target texture. This process is depicted in Fig. 4. The relighting method is bidirectional, meaning that probe and gallery textures can be interchanged.

In order to improve performance under low lighting conditions, instead of computing the difference in the RGB color space, we choose to use a Hue-Saturation-Intensity (HSI) model with the intensity weighed twice the amount of hue and saturation. We observe both visually and quantitatively improved relighting performance when using this color space and weighting scheme instead of simply computing an L_2 norm in RGB color space.
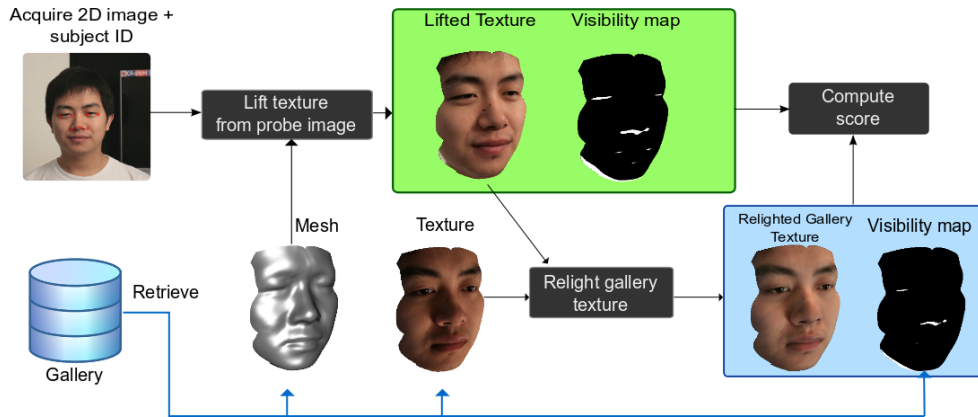
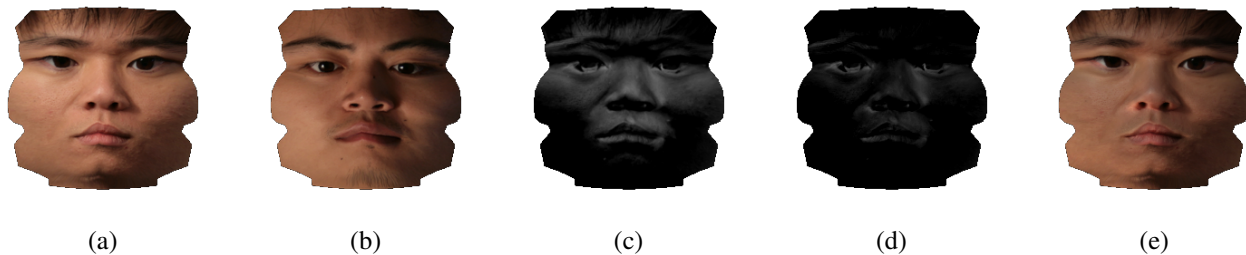Figure 3. The authentication phase of the 3D-aided 2D face recognition system.

(a)　　　　　　(b)　　　　　　(c)　　　　　　(d)　　　　　　(e)

Figure 4. Optimization for relighting (textures are in geometry image space): (a) M'_T: texture of subject A ; (b) M_T: texture of subject B; (c) Texture difference between subjects (before optimization); (d) Texture difference between subjects (after optimization); (e) Subject A with illumination of subject B ($I'_s + (I'_d + I'_a)\frac{M_T - I_s}{I_d + I_a}$).

The above equations describe an ASRM for a single point light and the objective function to be minimized. The ASRM is implemented as a Cg shader and for self-shadowing the shadow mapping technique is used [16]. To model multiple point lights, the contribution of each light's ASRM must be summed.

3.2.3　Distance Metrics

We introduce two simple distance metrics in order to evaluate the benefits of using unlit or relit images versus raw images in a face recognition scenario.

CWSSIM: We first apply a wavelet transform to the texture channel from the geometry images. This transform decomposes the images using the complex version [13] of the steerable pyramid transform (CWSPT), a linear multi-scale, multi-orientation image decomposition algorithm. The image is then divided into highpass and lowpass subbands. The lowpass subband is then fed into a set of steerable bandpass filters, which produce a set of oriented subbands and a lower-pass subband. This lower-pass subband is subsampled by two and recursively applied to the same set of steerable bandpass filters. Such pyramid wavelet representation is translation-invariant and rotation-invariant. To maintain reasonable image resolution and computational complexity, our algorithm applies a three-scale, 10-orientation complex

steerable pyramid transform to decompose each component of the image. Only the oriented subbands at the farthest scale are stored, allowing us to compare the subband coefficients of the two images directly without the overhead of reconstruction. These coefficients are compared using the Complex Wavelet Structural Similarity (CWSSIM) index algorithm, a translational insensitive image similarity measure inspired by the SSIM index algorithm [21]. CWSSIM iteratively measures the similarity indices between the two sliding windows placed in the same positions of the two images and uses the weighted sum as a final similarity score. This score is measured by a variation of the CWSSIM index equation originally proposed by Wang and Simoncelli [22].

In order to be able to handle variations in pose, we compute the normal map of the AFM when registered to the 2D input. The normal map is used to determine which pixels are not visible to the camera. The hidden pixels are used to create a thresholding map in the UV space of the AFM. We compute the CWSPT of the resulting image. This allows us to determine the contribution of each hidden pixel in the final score of the CWSSIM. Since the CWSSIM is computed using a sliding window, we use only those pixels for which the magnitude of the thresholded CWSPT map is below the upper bound τ.

A window of size 3×3 traverses the image one step

at a time. At each step, we extract all wavelet coefficients, resulting in two sets of coefficients $p_w = \{p_{w,i} | i = 1, ..., N\}$ and $g_w = \{g_{w,i} | i = 1, ..., N\}$, drawn from the probe image and the gallery image, respectively. For the same window, the coefficients from the visibility map are $m_w = \{m_{w,i} | i = 1, ..., N\}$. The distance metric can be written as follows:

$$
\tilde{S}(p_w, g_w) = 1 - \left(\frac{2 \sum_{i=1}^{N} |p_{w,i}||g_{w,i}| + K}{\sum_{i=1}^{N} |p_{w,i}|^2 + \sum_{i=1}^{N} |g_{w,i}|^2 + K} \right)
$$
$$
\times \left(\frac{2| \sum_{i=1}^{N} p_{w,i} g_{w,i}^* | + K}{2 \sum_{i=1}^{N} |p_{w,i} g_{w,i}^*| + K} \right)^r \cdot Q, \qquad (2)
$$

where w is the current step of the window, N is the number of coefficients in the window, and r is an experimentally determined exponent. The parameter K is a small positive value which is used to make the result numerically stable. The complex conjugate of $g_{w,i}$ is denoted as $g_{w,i}^*$, and $|m_w|$ is the cardinality of the set m_w. The variable Q is defined as

$$
Q = \frac{|\{m_{w,i} | m_{w,i} > \tau\}|}{|m_w|} \qquad (3)
$$

The first component (of the subtracted term) measures the equivalence of the two coefficient sets while the second reflects the consistency of phase changes. If $p_{w,i} = g_{w,i}$ for all i's, the distance is 0. The weighted sum of the local scores from all windows provides the distance score: $Score(F^P, F^G) = \sum_w (b_w \cdot \tilde{S}(p_w, g_w))$ where F^P, F^G are the probe and gallery images and b_w is a predefined weight depending on which subband the local window lies.

GS: We propose a new metric measure for global similarity using a correlation coefficient which is largely insensitive to serious mismatches induced by parts of the two images that do not match; thus, it is particularly suitable for measuring the similarity of face data and metadata (i.e., the texture channels of the geometry images) which may vary substantially, not only due to different capturing conditions, but also due to significant appearance changes of the individual subjects. Assume we are given two images $F_i(\mathbf{x}), i = 1, 2$. For each image, we compute $G_i = \nabla_x F_i + j \nabla_y F_i$, where $\nabla_x F_i$ and $\nabla_y F_i$ are the gradients along the horizontal and vertical direction respectively, and $O_i = G_i / |G_i| = e^{j\Phi_i}$, where Φ_i simply captures the orientation of image gradients at each spatial position. Our correlation coefficient is given by

$$
C = \Re \left\{ \int O_1(\mathbf{x}) O_2^*(\mathbf{x}) d\mathbf{x} \right\} = \int \cos(\Phi_1(\mathbf{x}) - \Phi_2(\mathbf{x})) d\mathbf{x}
$$
$$
= \int \cos(\Delta\Phi(\mathbf{x})) d\mathbf{x} \qquad (4)
$$

To understand how C can be used to cancel out the effect of outliers, suppose that the two images match only partially, that is $F_1(\mathbf{x}) = F_2(\mathbf{x})$, only for $\mathbf{x} \in \Omega$. For $\mathbf{x} \in \Omega$, we may observe that $\Delta\Phi(\mathbf{x}) = 0$. For $\mathbf{x} \notin \Omega$, it is not unreasonable

to assume that $\Delta\Phi(\mathbf{x})$ is uniformly distributed in $[0, 2\pi)$, since the images do not match, and, therefore, we expect that differences in gradient orientation can take any value in the range $[0, 2\pi)$ with equal probability. Thus, we can write

$$
C = \int_{\mathbf{x} \in \Omega} \cos(\Delta\Phi(\mathbf{x})) d\mathbf{x} + \int_{\mathbf{x} \notin \Omega} \cos(\Delta\Phi(\mathbf{x})) d\mathbf{x} = \Omega + 0 \quad (5)
$$

Overall, the value of C will be the contribution from the areas in the two images that match solely. Finally, note that gradient orientations are relatively robust for images under non-uniform illumination variations [9].

4. Results

For practical purposes, in all experiments we use a resolution of 256×256 for our geometry images. The threshold for τ in 3.2.3 is 0.5. Our implementation of bidirectional relighting on a consumer-level graphics hardware takes, on average, 5 seconds per relighting.

4.1. Bidirectional Relighting

We provide two examples, one demonstrating bidirectional relighting, and the second demonstrating the relighting of a gallery (2D facial images) to multiple probe textures. In all cases, textures from the same subject are used.

Figures 5(a,b) depict two textures of the same subject. Note that only 5(a) has a corresponding 3D mesh, while 5(b) is a simple 2D image and uses the mesh of 5(a). We performed bidirectional relighting to transfer the illumination conditions from one to the other. The result is two synthetic textures depicted in Figs. 5(c,d). The difference of the synthetic textures from the respective target textures is depicted in Figs. 5(e,f). After a visual inspection, no significant visible artifacts were introduced by the relighting process.

Figure 6 depicts a multiple relighting example. Again, only the source texture (top row) has a corresponding 3D mesh; the remaining textures are using the source's 3D mesh. Four textures with different illumination conditions are depicted (Fig. 6 (bottom row)) as target textures. The proposed method estimates the four synthetic relighted images depicted in Fig. 6 (middle row). These results show that the proposed method is robust under varying lighting conditions, since in all cases the relighting was qualitatively successful.

4.2. Face Recognition

In this scenario, 2D+3D data are acquired during enrollment, and 2D data during authentication (one-to-one matching). This is a reasonable installation where only a few enrollment stations are needed (which may be expensive) along with many authentication stations (which must be inexpensive).

Figure 5. Bidirectional relighting for a specific subject: (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 minus ST2; (f) RT2 minus ST1.
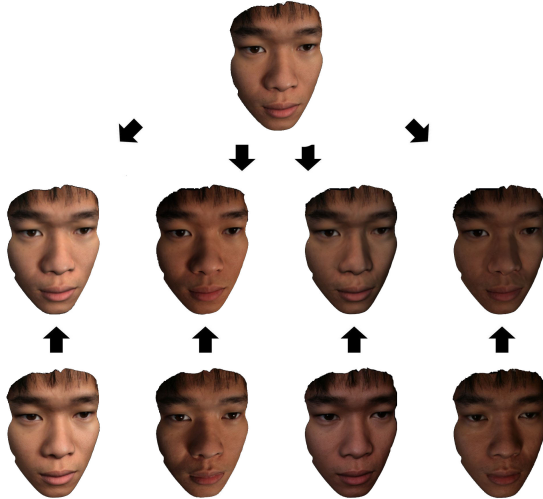

Figure 6. Examples of relighting. Top row: source texture. Middle row: synthetic relighted texture. Bottom row: target texture.

Figure 7. Examples from database UHDB12: Sample 2D images captured with varying lighting conditions for the same subject.

Database UHDB11 [18]: In order to analyze the impact of the variation in both pose and lighting, we acquired data from 23 subjects under six illumination conditions. For each illumination condition, we asked the subject to face four different points inside the room. This generated rotations on the Y axis. For each rotation on Y, we also acquired three images with rotations on the Z axis (assuming that the Z axis goes from the back of the head to the nose, and that the Y axis is the vertical axis through the subject's head); thus, we acquired images under six illumination conditions, four Y rotations, and three Z rotations per subject. For each image we concurrently acquired the 3D mesh. Fig. 8 depicts the variation in pose and illumination for one of the subjects from UHDB11. There are 23 subjects, resulting in

23 gallery datasets (3D plus 2D) and 1,602 probe datasets (2D only).

Database UHDB12 [19]: The 3D data were captured using a 3dMD^{TM} two-pod optical scanner, while the 2D data were captured using a commercial Canon^{TM} DSLR camera. The system has six diffuse lights that allow the variation of the lighting conditions. For each subject there is a single 3D scan (and the associated 2D texture) that is used as a gallery dataset and several 2D images that are used as probe datasets. Each 2D image is acquired under one of the six possible lighting conditions depicted in Fig. 7. There are 26 subjects, resulting in 26 gallery datasets (3D plus 2D) and 800 probe datasets (2D only).

Figure 8. Examples from database UHDB11 with variation of lighting and pose.

Authentication: We performed a variety of authentication experiments. We evaluated both relighting and unlight-
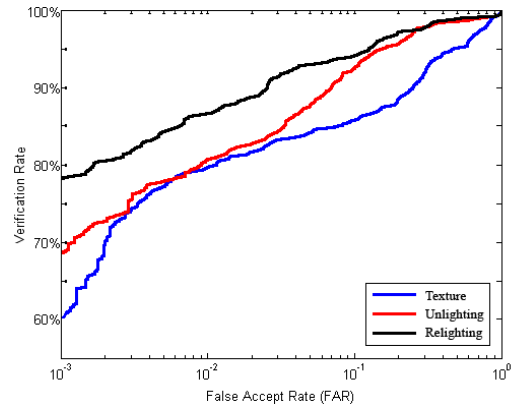

Figure 9. ROC curve on authentication experiment on UHDB12 (varying illumination).
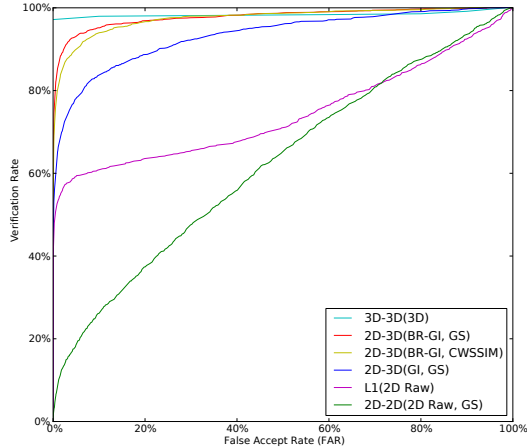
Figure 10. ROC curve for an authentication experiment using data from UHDB11 (varying illumination and pose). Note that the Equal Error Rate which the 3D-aided 2D face recognition algorithm achieves is half that of the leading commercial product available at this time.
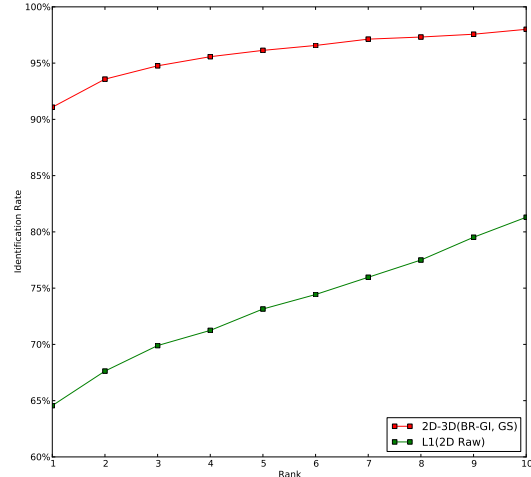

Figure 11. Identification performance of the 3D-aided 2D face recognition approach versus the performance of a leading commercial 2D face recognition product.

ing. In the unlighting case, both gallery and probe images were unlit (thus becoming albedos). In the relighting case, the gallery image was relit according to the probe image. The results for UHDB12 (using our algorithm, the CWSSIM metric and Z-normalization) are summarized using a Receiver Operating Characteristic (ROC) curve (Fig. 9). Note that face recognition benefits more from relit images than from unlit images. It achieves a 10% higher authentication rate at 10^{-3} False Accept Rate (FAR) than unlighting. The performance using the raw texture is also included as a baseline. Even though these results depend on the UHDB12 and the distance metric that was used, they indicate clearly that *relighting is more suitable for face recognition than unlighting*. The reason behind this is that any unlighting method produces an albedo for which the ground truth is not known; therefore, the optimization procedure is more prone to errors.

Before we proceed, we will explain E normalization. Let $\{F_i^G(\mathbf{x})\}$ with $i = 1, \ldots, N$ be the set of gallery images and $\{F_i^P(\mathbf{x})\}$ with $i = 1, \ldots, K$ be the probe set. We create a dissimilarity matrix in the gallery set $\mathbf{D^1} = [C_n(F_i^G(\mathbf{x}), F_j^G(\mathbf{x}))]$, where $C_n(F_i(\mathbf{x}), F_j(\mathbf{x}))$ is the normalized correlation coefficient $C(F_i(\mathbf{x}), F_j(\mathbf{x}))$ in order to force C_n to satisfy the following properties: (i) reflectivity (i.e., $C_n(F_i(\mathbf{x}), F_i(\mathbf{x})) = 0$) and (ii) positivity (i.e., $C_n(F_i(\mathbf{x}), F_j(\mathbf{x})) > 0$ if $F_i(\mathbf{x}) \neq F_j(\mathbf{x})$). It can be easily seen that because C is symmetric, C_n satisfies the symmetry property $C_n(F_i(\mathbf{x}), F_j(\mathbf{x})) = C_n(F_j(\mathbf{x}), F_i(\mathbf{x}))$, as well. Since C_n is a proper dissimilarity measure, we can extract features by applying the method proposed in [10]. The features of the probe set can be derived by creating the dissimilarity matrix between the gallery and probe set $\mathbf{D^2} = [C_n(F_i^G(\mathbf{x}), F_j^P(\mathbf{x}))]$ and then applying the method

in [10]. Finally, we create a new dissimilarity matrix $\mathbf{D}^3$ using these features.

In order to assess the robustness of the 3D-aided 2D face recognition approach with respect to both lighting and pose variation, we employed UHDB11. Figure 10 depicts the ROC curve for UHDB11 for four different methods: (i) 3D-3D: Using the UR3D algorithm by the University of Houston where both the gallery and probe are 3D datasets (shape only no texture) [6]; (ii) 2D-3D(BR_GI, GS): Our proposed 2D-3D algorithm using bidirectionally relit images, GS distance metric, and E-normalization; (iii) 2D-3D(BR_GI, CWSSIM): Our proposed 2D-3D algorithm using bidirectionally relit images, CWSSIM distance metric, and E-normalization;(iv) 2D-3D(GI, GS): Our proposed 2D-3D algorithm using raw texture from the geometry images, GS distance metric, and E-normalization; (v) 2D-2D(2D Raw, GS): Computing the GS distance metric for the raw 2D data, and E-normalization; (vi) L1(2D Raw, GS): Results from the L1 IdentityToolsSDK [7]. Note that our proposed 2D-3D(BR_GS, GS) outperforms one of the best commercial products.

2D-3D Identification Experiment: We also constructed an identification experiment using data from UHDB11. The results are provided in a Cumulative Matching Characteristic (CMC) curve on 23 subjects of UHDB11 (Fig. 11). Our approach outperforms the commercial 2D-only product throughout the entire CMC curve.

5. Conclusions

A new system for face recognition that uses 2D+3D data for enrollment and 2D data for authentication was proposed. This was achieved by fitting an annotated deformable model to the 3D data and by using an analytical skin reflectance

model to relight the 2D data while using the fitted AFM from the gallery. The qualitative and quantitative evaluation demonstrate that the 2D/3D method with relighting process provides robust face recognition performance under varying pose and lighting condition.

6. Acknowledgements

References

[1] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, Feb. 2003. 2

[2] S. Biswas, G. Aggarwal, and R. Chellappa. Robust estimation of albedo for illumination-invariant matching and shape recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:884–899, 2009. 2

[3] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003. 1, 2

[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. 1

[5] D. Huang, M. Ardabilian, Y. Wang, and L. Chen. Asymmetric 3d/2d face recognition based on lbp facial representation and canonical correlation analysis. In *Proc. 16th IEEE International Conference on Image Processing*, pages 3325–3328, Cairo, Egypt, Nov. 2009. 2

[6] I. A. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, Apr. 2007. 2, 3, 7

[7] L1 Identity Solutions. L1 FaceIt SDK. 7

[8] J. Lee, R. Machiraju, H. Pfister, and B. Moghaddam. Estimation of 3D faces and illumination from single photographs using a bilinear illumination model. In *Proc. Eurographics Symposium on Rendering*, pages 73–82, Konstanz, Germany, Jun. 29 - Jul. 1 2005. 2

[9] M. Osadchy, D. Jacobs, and M. Lindenbaum. Surface dependent representations for illumination insensitive image comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):98–111, 2007. 5

[10] E. Pekalska, P. Paclik, and R. Duin. A generalized kernel approach to dissimilarity-based classification. *Journal of Machine Learning Research, Special Issue on Kernel Methods*, 2(2):175–211, 2002. 7

[11] P. J. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:831–846, 2010. 1

[12] B. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975. 3

[13] J. Portilla and E. Simoncelli. A parametric texture model based on joint statistic of complex wavelet coefficients. *International Journal of Computer Vision*, 40:49–71, 2000. 4

[14] D. Riccio and J.-L. Dugelay. Geometric invariants for 2D/3D face recognition. *Pattern Recognition Letters*, 28(14):1907–1914, 2007. 2

[15] W. Smith and E. Hancock. Estimating the albedo map of the face from a single image. In *Proc. IEEE International Conference on Image Processing*, volume 3, pages 780–783, Genoa, Italy, Sep. 11-14 2005. 2

[16] G. Toderici, G. Passalis, T. Theoharis, and I. A. Kakadiaris. An automated method for human face modeling and relighting with application to face recognition. In *Proc. Workshop on Photometric Analysis For Computer Vision*, Rio de Janeiro, Brazil, Oct. 14-21 2007. 3, 4

[17] F. Tsalakanidou, S. Malassiotis, and M. Strintzis. A 2D+3D face identification system for surveillance applications. In *Proc. IEEE International Conference on Advanced Video and Signal based Surveillance*, pages 194 – 199, London, UK, Sep. 5-7 2007. 2

[18] UH Computational Biomedicine Lab. UHDB11 face database. http://cbl.uh.edu/URxD/datasets/, 2009. 6

[19] UH Computational Biomedicine Lab. UHDB12 face database. http://cbl.uh.edu/URxD/datasets/, 2009. 6

[20] Y. Wang, L. Zhang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face relighting from a single image under arbitrary unknown lighting conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1968–1984, Nov. 2009. 2

[21] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 4

[22] Z. Wang and E. Simoncelli. Translation insensitive image similarity in complex wavelet domain. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, volume II, pages 573–576, Philadelphia, PA, Mar. 18-25 2005. 4

[23] L. Yin and M. Yourst. 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*, pages 1–8, New York, NY, Nov. 8 2003. 2

[24] Z. Zhou, A. Ganesh, J. Wright, S.-F. Tsai, and Y. Ma. Nearest-subspace patch matching for face recognition under varying pose and illumination. In *Proc. 8th IEEE International Conference on Automatic Face Gesture Recognition*, pages 1 –8, Amsterdam, The Netherlands, Sept. 2008. 2