Illumination Normalization Using Self-lighting Ratios for 3D2D Face Recognition

Xi Zhao, Shishir K. Shah, and Ioannis A. Kakadiaris

Computational Biomedicine Laboratory, Department of Computer Science,
University of Houston, 4800 Calhoun, Houston, TX 77204
zhaoxi1@gmail.com, shah@cs.uh.edu, ioannisk@uh.edu

Abstract. 3D-2D face recognition is beginning to gain attention from the research community. It takes advantage of 3D facial geometry to normalize the head pose and registers it into a canonical 2D space. In this paper, we present a novel illumination normalization approach for 3D-2D face recognition which does not require any training or prior knowledge on the type, number, and direction of the lighting sources. Estimated using an image-specific filtering technique in the frequency domain, a self-lighting ratio is employed to suppress illumination differences. Experimental results on the UHDB11 and FRGC databases indicate that the proposed approach improves the performance significantly for face images with large illumination variations.

Keywords: Lighting ratio, illumination suppression, 3D-2D face recognition.

1 Introduction

The research on 3D-2D Face Recognition has rapidly increased in recent years. Since the facial geometry is invariant to camera viewpoints, 3D facial data can be used to overcome the difficulties of head pose variations. Currently, limited by the high cost of 3D face scanners, it is impractical to deploy a large number of 3D scanners in real world face recognition applications. This emphasizes the advantage of using the alternative approach of 3D-2D face recognition setups, which use 3D + 2D facial data as the gallery and 2D facial images as the probe. An excellent survey of face recognition was presented by Bowyer *et al.* [1]. Recent reviews are included in [2][24][3][4]. Illumination variations remain a challenge for improving the overall robustness and performance of 3D-2D FR systems. In this paper, we provide an efficient solution for illumination normalization to enhance the performance of 3D-2D face recognition systems.

The basic idea of our method is to relight the input images to a preset texture with a constant value so that the relit input images are symmetric and close to each other. This is achieved by computing the point-wise division between the original texture and the *lighting ratio*, an estimate of the illumination conditions. Assuming that most of the illumination effects vary slowly on the facial textures, and that the majority of the energy of illumination is distributed among the low

frequencies, we estimate the lighting ratios via low-pass filtering in the frequency domain. In order to choose the cut-off frequency for the filter so that it can be adaptively used for images under various lighting conditions, we propose an image-specific low-pass filtering technique. The lighting ratio is adjusted to minimize the L_2 norm between the outcome of the division and the preset texture to further reduce the contrast and exposure difference on faces due to skin type, camera parameters, and lighting conditions.

Our main contributions are: (i) developed a novel illumination normalization method based on self-lighting ratio with the aim of enhancing face recognition performance; (ii) developed an image-specific low-pass filtering method that does not require preset parameters; (iii) proposed to use a preset texture with a constant value as the reference for illumination suppression.

The rest of the paper is organized as follows: Section 2 discusses the related work in the literature. Section 3 provides an overview of a 3D-2D face recognition system. The proposed method is detailed in Section 4. Section 5 presents the experimental results. Section 6 summarizes our findings.

2 Literature Reviews

The illumination normalization methods for face recognition can be generally divided into two catogories: (i) normalizing the lighting conditions for a pair of images (relighting) [5][6][7][8][9][10], (ii) normalizing the lighting effects for a set of images (unlighting) [11][12][13][14]. To estimate the lighting effects, either subspace-based models or image processing techniques have been adopted.

Subspace-based methods model the lighting in a low-dimensional space. Shim *et al.* [5] built a subspace model for each pixel under various lighting conditions per subject and per pose. The lighting, pose, and reflectance were jointly inferred using an EM-like process. Zhang and Samaras [15] recovered person-specific basis images by combining spherical harmonics illumination representation with 3D morphable models. The basis images were subdivided into small regions and incorporated into an MRF framework to remove the lighting and relit faces under arbitrary unknown lighting conditions in [7]. Blanz and Vetter [16] imposed linear constraints on both the albedo and the shape of the face. They proposed a 3D morphable model to represent each face as a linear combination of 3D basis exemplars. The recovery of lighting parameters was realized as an optimization problem that aimed to minimize the difference between the input and the reconstructed image. Sungho *et al.* [13] minimized the L_1-norm between an image and a linear combination of principal lighting eigenvectors for unlighting. The lighting eigenvectors were obtained using PCA on eight illumination maps with eight light sources distributed evenly. Kumar *et al.* [6] built eight-dimensional subspace models for texture and illumination, respectively.

Image processing based methods estimate the lighting mainly via spatial smooth filtering or using Quotient image approaches. Quotient Image [17] is the ratio between an image and a linear combination of three images lit by independent light sources. Wang *et al.* [18] proposed the Self-Quotient image
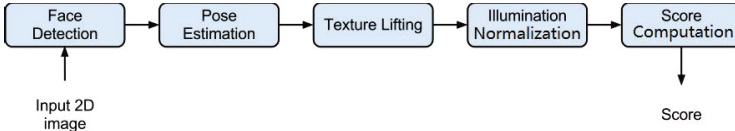
Fig. 1. Flowchart of the 3D-2D face recognition pipeline

which replaced the linear combination of three images by the smoothed input image to ease the assumptions posed in Quotient image and presented preliminary results on face recognition. Han *et al.* [8] computed illuminations using a homomorphic wavelet filtering and computed the Quotient image between two illumination images for relighting. Chen *et al.* [10] used edge-preserving filters to obtain the large-scale layers of two images and then convolved them to compute the ratio coefficients for relighting. Biswas *et al.* [11] added a signal-dependent non-stationary noise term to the Lambertian model and hence, computed albedo as the Linear Minimum Mean Square Error estimate of the true albedo. The noise incorporated the errors in surface normal and illumination estimation and thus resulted in obtaining a more realistic albedo using their methods. Vural *et al.* [14] applied the Ayofa-filters to filter out the illumination effect on faces.

3   Overview of a 3D-2D Face Recognition System

The face recognition experiments are conducted using the 3D-2D face recognition system proposed by Toderici *et al.* [3]. The facial data in the gallery include both a mesh and a texture while the facial data in the probe includes only texture. For each 3D mesh in the gallery, the Annotated Facial Model (AFM) [19] is fit to establish a person-specific point-to-point correspondence mapping of the 3D mesh to the UV space. For the 2D image in the gallery, each fitted model is transformed and projected to the pose appearing in the texture image using a set of landmarks. Then, using the previously established correspondences, the texture is lifted to a two-dimensional (UV) space with pose normalized to be frontal. The self-occluded parts of the face in the original pose are masked out. For each texture in the probe dataset, the same process is repeated using the fitted model from the gallery dataset that it is compared with. After processing textures in both gallery and probe, the illumination is normalized either by relighting [3] or unlighting (illumination normalization) as proposed in this paper, and a correlation-coefficient-based distance metric is computed for each pair of textures from the gallery and the probe. A flowchart of the system is presented in Fig. 1.

4   Minimizing Illumination Differences Using the Self-lighting Ratio

4.1   Definition of the Self-lighting Ratio

The self-quotient image Q is defined as an intrinsic property of a face image I of an individual [18], by $Q = \frac{I}{\hat{I}}$, where $\hat{I}$ is the smoothed version of I, F is the

smoothing kernel, and the division is pixel-wise. Wang *et al.* [18] used a Gaussian filter F to obtain the $\hat{I}$.

The self-quotient image method has demonstrated its capability in improving face recognition performance via reducing illumination differences. However, during its computation the range of self-quotient image spans from zero to infinite, since it is computed from a pixel-wise division. A small variation on the smoothed value may lead to large variations on the self-quotient image. Therefore, while this method reduces the illumination on the facial texture, it also introduces numerical artefacts, especially in the shadowed regions. For example, obvious noise can be observed on the shadow region of the face in Fig. 2 in [18]. Meanwhile, using a spatial filter to obtain the smoothed version of I raises two different problems. First, it is hard to determine the appropriate sliding window size for filtering since the size depends on the face scale and the lighting conditions. Second, there is no consensus on the appropriate kernel in the spatial domain to obtain the lighting conditions on the face. Nevertheless, a smoothing kernel in the spacial domain has the effect of a low-pass filter in the frequency domain removing high frequencies (mostly edges) while enhancing low frequencies. This is consistent with the statement in [20], where under the Lambertian assumption, the low frequency part in the image captures mostly the lighting condition on the facial image. Thus, an alternative way to overcome the aforementioned problems is to estimate the self-lighting ratio (Eq. 1). This can be written as:

$$T = \left(\alpha\frac{1}{\hat{I}} + \beta\right)I = \left(\alpha\frac{1}{\mathcal{T}^{-1}(\mathcal{T}(\xi)\mathcal{T}(I))} + \beta\right)I = L_R I \tag{1}$$

where L_R denotes the self-lighting ratio, ξ denotes a low-pass filter, $\mathcal{T}$ and $\mathcal{T}^{-1}$ is the Fourier transform and inverse Fourier transform, T is the relit image, $\hat{I}$ is the smoothed version of the image I, and α, β are scale and offset parameters. The adoption of these two parameters compensates the contrast and offset variations of facial textures I caused by lighting conditions, skin type, and camera parameters. The addition, the division, and the multiplication are pixel-wise.

4.2   Algorithm

The proposed algorithm changes the self-lighting ratio to minimize the difference between the relit image T and a predefined image I_p (set to a uniform value 60): $\arg\min\|I_p - L_R I\|_F$. Thus, the algorithm pursues the minimum Frobenius norm of the difference image between the input image $L_R I$ and I_p. There are three positive effects from this algorithm. First, all the relit images $T = L_R I$ are similar to each other in terms of global intensities since we adjust the lighting conditions according to one preset texture. Second, the lighting conditions on the relit images are more symmetric since I_p is symmetric and the parameters α, β adjust the contrast and offset variations on $\hat{I}$. Finally, the estimation of $\hat{I}$ is more accurate since it is performed using an image-specific low-pass filtering algorithm, which avoids the use of an arbitrary low-pass filtering parameter for all textures regardless of their lighting conditions.

Algorithm 1. Illumination normalization using lighting ratio

Input: Facial texture I and low-pass filter ξ
Output: Normalized texture T

1: Convert the texture I to the HSV color space and extract its V channel (V_g)
2: Obtain V_g's Fourier spectrum F_g via 2D Fast Fourier Transform
3: Compute the image-specific parameter P for ξ (Sec. 4.4)
4: Filter the magnitude of F_g by applying the low-pass filter ξ to obtain F_g'
5: Apply the inverse Fast Fourier transform on the filtered spectrum F_g' to obtain the lighting component $\hat{V}_g$
6: Minimize the L_2 norm between I_p and $(\alpha \frac{1}{\hat{V}_g} + \beta) * V_g$
7: Compute the normalized texture: $T = (\widehat{\alpha} \frac{1}{\hat{V}_g} + \widehat{\beta}) * V_g$

The detailed algorithm is depicted in Alg. 1. The symbols $\hat{\alpha}$ and $\hat{\beta}$ in Step 7 denote the optimized parameters. The Nelder-Meade simplex algorithm [21] is chosen for the minimization in Step 6. It is one of the best known algorithms for unconstrained optimization without derivatives and is quite simple in computation.

4.3 Illumination Estimation in the Frequency Domain

To filter the image in the frequency domain, we first apply the Fourier Transform to the image V: $F(k, l) = \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} V(s,t) e^{-2\pi i (\frac{ks}{N} + \frac{lt}{N})}$, where V is the image, s, t are the indices on V, and N is the image size, F is the Fourier spectrum and k, l are the indices on F. The exponential term is the basis function corresponding to each point (k, l) on the Fourier spectrum. Then, the low-pass filter ξ which has the same size of F is multiplied with the magnitude of the spectrum F in a pixel-by-pixel fashion: $|F'(k, l)| = |F(k, l)| \xi(k, l)$, where $|F(k, l)|$ is the magnitude of the input spectrum F, $\xi(k, l)$ is the filter, and $|F'(k, l)|$ is the filtered spectrum. To obtain the lighting estimates in the spatial domain, the inverse Fourier transform is applied on the filtered spectrum F': $\hat{V}(s, t) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} F'(k, l) e^{2\pi i (\frac{ks}{N} + \frac{lt}{N})}$, where F' captures the low frequencies of the image V, and $\frac{1}{N^2}$ is a normalization term in the inverse transformation.

4.4 Image-Specific Low-Pass Filtering

It is hard to choose an arbitrary cut-off frequency for all facial images for low-pass filters since the lighting conditions vary in different facial images. Instead, we propose to compute an image-specific energy threshold P (in the range [0,1]) for each image automatically. The basic idea is that the number of lights, lighting direction, and intensity change the frequency distribution mostly in the low frequencies. While the portion of energy useful for identification is distributed mostly in middle frequencies, and the skin details which are distributed mostly in high frequencies remain relatively steady, the energy variations among fixed posed facial textures are highly related to the lighting conditions. Thus, by varying the cut-off energy instead of the cut-off frequency, we can estimate the lighting conditions in a more accurate manner for each image. Note that the energy

in Fourier spectrum F_g is defined as the L_2 norm of the F_g magnitude and we iteratively increase the cut-off frequency until the energy of the passed frequency components is greater than a portion (P) of the energy of all frequencies of the image in the implementation.

We adaptively compute P for each image, which varies with the energy from lighting conditions in the image. Histogram equalization (HE) adjusts the intensity distributions on the histograms of the texture which is able to change the energy of the image to a level quite constant among different face textures across various lighting conditions. Thus, we approximate the P using the energy of the HE adjusted image as a reference:

$$P \propto E(I)/E(I_{eq}) \approx \kappa E(I)/E(I_{eq}), \tag{2}$$

where κ is a constant (set to 0.05), $E(I)$ is the energy of the input image I, and $E(I_{eq})$ is the energy of the input image processed by the histogram equalization I_{eq}. The energy is defined as the L_2 norm of their V_g images. Since the energy identification and the skin details are relatively constant as well as the energy of the HE adjusted image, P is able to reflect the energy variations caused by lighting conditions.

The choice of the low-pass filter is also important in estimating the lighting conditions. The ideal filter is a simple low-pass filter, which suppresses all frequencies higher than a threshold frequency C and keeps the lower frequencies unchanged. However, it introduces an ringing artefact that occurs along the edges of the filtered image in the spatial domain. Thus, we opt to use more sophisticated low-pass filters (*e.g.,* a Gaussian filter or a Butterworth filter). The Gaussian filter has the same shape in the spatial and frequency domains and therefore does not incur the ringing artefact. The Butterworth filter is an approximation of Gaussian filter and outputs a similar result as the one obtained by the Gaussian filter. However, considering the computational complexity, the Butterworth filter is a better choice for wide low-pass filtering, while the Gaussian filter is more appropriate for narrow low-pass filtering, which is our case. Thus, we adopt a Gaussian filter ξ in this work.

5   Experimental Result

5.1   Datasets

To assess the robustness of our algorithm under various head poses and illumination conditions, we tested it on the publicly available UHDB11 [22] and FRGC v2.0 [23] datasets. The UHDB11 dataset contains 1,625 3D facial scans captured by a 3dMD scanner and 1,625 images captured by a Canon DSLR camera. Facial data from 23 different subjects were acquired under six indoor illumination conditions, four yaw rotations, and three roll rotations per subject. The head pose varies from ±50 degrees in the roll direction, and ±30 degrees in the pitch direction. We also evaluate our algorithm on the FRGC v2.0 dataset, which contains a large number of co-registered face images and 3D meshes with controlled and uncontrolled lighting conditions.

Fig. 2. Depiction of the unlighting results. The first and fourth rows depict the original face textures in the UV space. The second and fifth rows depict the self-ratio images. The third and sixth rows depict the unlit face textures using our method.

5.2   Illumination Normalization

Figure 2 depicts the comparison between the self-ratio and the proposed algorithm. It can be observed that the lighting effects on the unlit facial texture is reduced and distributed more evenly than those on the raw facial texture. When compared to the images form the self-ratio method, the lighting effects are more constant across our unlit images and the image energies of our relit textures are similar. This is because the lighting ratios adjust the overall intensity levels to the preset I_p. In addition, the low-pass filtering designates image-specific parameters for the low-pass filter leading to a more accurate estimate of the subject differences lighting conditions, thus the differences among unlit textures are mainly caused by identifications.
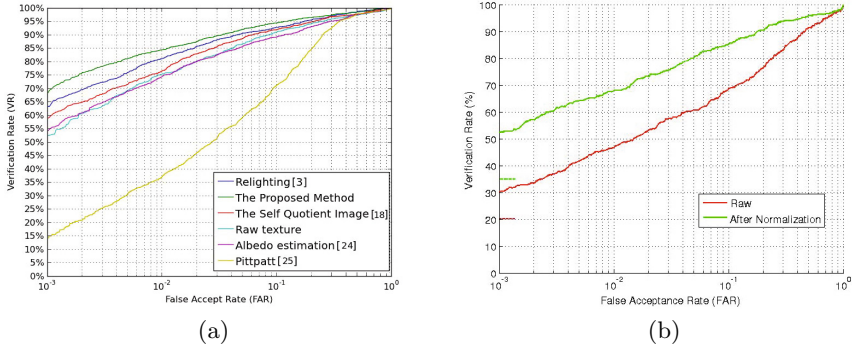
Fig. 3. Depiction of the ROC curves on the (a) UHDB11 dataset; (b) FRGC dataset

5.3 Face Recognition

Figure 3(a) depicts the ROC curves for six different methods tested on the UHDB11 dataset: (i) computing the distance metric for a relit texture using the method in [3]; (ii) computing the distance metric for the unlit texture by the proposed algorithm; (iii) computing the distance metric for the self-quotient image method proposed in [18]; (iv) computing the distance metric for the raw texture output from texture lifting; (v) computing the distance metric for the unlit texture using the albedo estimation method proposed in [25]; and (vi) using the 2D-2D PittPatt [26] face recognition system. The verification rates at 10^{-3} FAR are 64.1%, 68.9%, 61.0%, 52.3%, 54.2% and 13.8%, respectively. The gap in performance between PittPatt and other algorithms is due to pose variations in the dataset, since the other algorithms use the lifted texture in UV space with pose normalized while PittPatt recognition is performed on the face ROI detected from original images. The method proposed by Biswas *et al.* [25] results in an increased verification rate compared to the raw lifted texture. One possible reason for the relatively low performance is that their method requires a training dataset to learn the pixel intensity statistics and our training dataset is not the same as theirs. Our method has demonstrated improvement over the self-quotient method because of a better estimation of lighting conditions via the image-specific low-pass filtering and a better normalization process with the adoption of parameters α and β. Compared to the relighting method which requires the 3D face geometry, our algorithm still demonstrated its effectiveness by achieving 5% higher on verification rate.

To test the robustness of the proposed method, we also evaluated our algorithm on the FRGC v2 dataset. We used the same experimental setup as the one used in the Al-Osaimi study [4] (the same 250 facial scans from 250 subjects as the gallery and 470 facial scans as the probe). Figure 3(b) depicts the ROC curves using raw and normalized scores from our illumination normalization algorithm. Our algorithm achieves verification rates of 30.0% and 53.8% at 0.001 False Accept Rate. Compared to the verification of 20.43% and 34.89% achieved

in [4], shown as the red and green dot lines in Fig.3b, the proposed approach performs better.

6   Conclusion

We proposed a novel approach to normalize illumination conditions on facial textures, without requiring 3D geometry information and prior knowledge of lighting conditions. This method has been incorporated into a 3D-2D face recognition system, thus providing the capacity to handle illumination variations on faces with different poses. We tested the algorithm on the UHDB11 and FRGC v2.0 datasets, and the results demonstrate its robustness and accuracy.

References

1. Bowyer, K., Chang, K., Flynn, P.: A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. Computer Vision and Image Understanding 101, 1–15 (2006)
2. Zhou, Z., Ganesh, A., Wright, J., Tsai, S.F., Ma, Y.: Nearest-subspace patch matching for face recognition under varying pose and illumination. In: Proc. 8th IEEE International Conference on Automatic Face Gesture Recognition, Amsterdam, The Netherlands, pp. 1–8 (2008)
3. Toderici, G., Passalis, G., Zafeiriou, S., Tzimiropoulos, G., Petrou, M., Theoharis, T., Kakadiaris, I.: Bidirectional relighting for 3D-aided 2D face recognition. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 2721–2728 (2010)
4. Al-Osaimi, F.R., Bennamoun, M., Mian, A.S.: Illumination normalization of facial images by reversing the process of image formation. Machine Vision and Applications 22, 899–911 (2011)
5. Shim, H., Luo, J., Chen, T.: A subspace model-based approach to face relighting under unknown lighting and poses. IEEE Transactions on Image Processing 17, 1331–1341 (2008)
6. Kumar, R., Jones, M., Marks, T.: Morphable reflectance fields for enhancing face recognition. In: IEEE Conference on Proc. Computer Vision and Pattern Recognition (CVPR), pp. 2606–2613 (2010)
7. Wang, Y., Zhang, L., Liu, Z., Hua, G., Wen, Z., Zhang, Z., Samaras, D.: Face relighting from a single image under arbitrary unknown lighting conditions. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 1968–1984 (2009)
8. Han, H., Shan, S., Chen, X., Gao, W.: Illumination transfer using homomorphic wavelet filtering and its application to light-insensitive face recognition. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam (2008)
9. Chen, J., Su, G., He, J., Ben, S.: Face Image Relighting using Locally Constrained Global Optimization. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 44–57. Springer, Heidelberg (2010)
10. Chen, X., Chen, M., Jin, X., Zhao, Q.: Face illumination transfer through edge-preserving filters. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs CA (2011)

11. Biswas, S., Aggarwal, G., Chellappa, R.: Robust estimation of albedo for illumination-invariant matching and shape recovery. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 884–899 (2009)
12. Zou, X., Kittler, J., Hamouz, M., Tena, J.: Robust albedo estimation from face image under unknown illumination. In: Proc. SPIE Biometric Technology for Human Identification, Orlando, FL, p. 69440A (2008)
13. Suh, S., Lee, M., Choi, C.H.: Robust albedo estimation from a facial image with cast shadow. In: Proc. 18th IEEE International Conference on Image Processing, Brussels, Belguim, pp. 873–876 (2011)
14. Vural, S., Mae, Y., Uvet, H., Arai, T.: Illumination normalization for outdoor face recognition by using ayofa-filters. Journal of Pattern Recognition Research 6, 1–18 (2011)
15. Zhang, L., Samaras, D.: Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 351–363 (2006)
16. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 1063–1074 (2003)
17. Shashua, A., Riklin-Raviv, T.: The quotient image: class-based re-rendering and recognition with varying illuminations. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 129–139 (2001)
18. Wang, H., Li, S., Wang, Y., Zhang, J.: Self quotient image for face recognition. In: Proc. IEEE International Conference on Image Processing, vol. 2, pp. 1397–1400 (2004)
19. Kakadiaris, I., Passalis, G., Toderici, G., Murtuza, M.N., Lu, Y., Karampatziakis, N., Theoharis, T.: Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 640–649 (2007)
20. Basri, R., Jacobs, D.: Lambertian reflectance and linear subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 218–233 (2003)
21. Nelder, J.A., Mead, R.: A simplex method for function minimization. Computer Journal 7, 308–313 (1965)
22. UH Computational Biomedicine Lab: UHDB11 face database (2009), http://cbl.uh.edu/URxD/datasets/
23. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, vol. 1, pp. 947–954 (2005)
24. Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: FRVT 2006 and ICE 2006 Large-Scale Experimental Results. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 831–846 (2010)
25. Biswas, S., Chellappa, R.: Pose-robust albedo estimation from a single image. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, pp. 2683–2690 (2010)
26. Pittsburgh Pattern Recognition: PittPatt face tracking & recognition Software Development Kit 5.1 (2011)