Illumination Alignment Using Lighting Ratio:
Application to 3D-2D Face Recognition

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

Abstract—Illumination alignment is accomplished by relighting or unlighting methods that compensate the appearance differences due to variations in illumination conditions between a pair of face images or among a set of face images. In this paper, we propose a lighting ratio based method that avoids parameter tuning process by using an image-specific low-pass filtering. In addition, it includes an optimization step that brings the images into a similar energy level. The proposed method is simple and does not require any prior knowledge of the number, position, type of lighting sources or the 3D face geometry. Experimental results indicate that our method is effective and efficient for illumination-robust face recognition.

I. INTRODUCTION

Illumination manipulation and alignment are widely used in film post-production systems and face recognition systems. We use the term alignment to refer to relighting of the target images to a preset texture with a constant value (i.e., unlighting), or relighting of the target images to another facial texture (i.e., relighting). In general, face relighting involves a pair of images, and it aligns the illumination conditions on a target image to the lighting of a reference image. On the other hand, face unlighting involves a single image, and it removes the lighting effects from these face textures.

Two-dimensional face recognition has been beset by several challenges, including illumination variations. Quotient Image [13] is the ratio between an image and a linear combination of several sets of images from different subjects which are lit by independent lights. It has been proven to be a simple and effective way to normalize illumination under certain assumptions (e.g., lambertian surface, absence of cast shadow). Wang et al. [16] proposed the Self-Quotient image which replaced the linear combination of three images by the smoothed input image in order to ease the assumptions posed by Quotient image and resulted in improved performance. However, the choice of spatial sliding window size and parameters of Gaussian kernel were empirically determined.

In this paper, we propose using the lighting ratio to approximate the factors leading to illumination differences and hence align the dominant lighting conditions on facial textures. Lighting ratio is the ratio between an input image and its smoothed version with adjusted lighting 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 ratio via low-pass filtering in the frequency domain. In order to choose the cut-off frequency for the filter used for images under various lighting conditions, we propose an image-specific low-pass filter. The lighting ratio is adjusted to minimize the Frobenius norm between the outcome of the division and a reference texture to further reduce the contrast and exposure differences on facial images due to skin type, camera parameters, and lighting conditions.

The lighting ratio based illumination alignment methods are then used in a 3D-2D face recognition system to address illumination challenges. 3D-2D face recognition systems use 3D geometry and facial texture in the gallery for recognition of 2D probe images from ordinary cameras [15]. The 3D textured data in the gallery can be used to register 2D images under various poses and normalize head orientations into a frontal pose. Surveys for 3D-2D face recognition can be found in [20][12][9][1].

Our main contributions are: (i) developed a method that uses the lighting ratio to normalize illumination variation via either unlighting or relighting; and (ii) provided comparisons between unlighting, relighting, and bi-directional relighting in terms of photo-realism and face recognition performance.

The paper is organized as follows: Section II discusses related work. The proposed lighting ratio based illumination normalization method is detailed in Section III. Section IV presents the experimental results. Section V concludes the paper.

II. LITERATURE REVIEW

Illumination alignment methods can be categorized into subspace-based or image processing-based.

Subspace-based methods model the lighting in a low-dimensional space. Linear models are built and fit to face images. Relighting is performed by using the model parameters from the reference illumination to synthesize the target illumination. Shim et al. [14] built a subspace model on each pixel under various lighting conditions per subject and per pose. The lighting, pose and reflectance was jointly inferred using an EM-like process. However, the reference image and the target image should originate from the same subject, which limits its application on face recognition. Zhang and Samaras [18] 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 a MRF framework to remove the lighting [17]. However, the methods in [18][17] required that subjects in the gallery and the probe have approximately the same pose.
Image processing based methods relight the target image by multiplying it with the relighting factors computed from both the target and reference images. Han et al. [8] compared illuminations on both target and reference images to obtain the relighting ratio. The illuminations were computed from homomorphic wavelet filtering. Chen et al. [4] first applied an Active Appearance Model (AAM) to warp the reference to the target image. The relighting was achieved via solving a locally constrained global objective function. Chen et al. [5] used an Active Shape Model (ASM) to align the two images, and then used edge-preserving filters to obtain the large-scale layers of two images and convolved them to compute the relighting factors. Approaches based on quotient images require alignment of the two images, generally accomplished by AAM or ASM, where a large number of manually labelled landmarks are necessary. Biswas et al. [3] 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 estimates and thus resulted in a more realistic albedo. However, the accuracy of the estimate highly depends on the training albedo set, which should be captured under strictly controlled lighting conditions. Zhao et al. [19] proposed to use self lighting ratio to normalize the illumination conditions on face images. In this paper, we advance the self lighting ratio to lighting ratio in terms of definition and computation in order to address illumination alignment problem and generalize the proposed method.

III. MINIMIZING ILLUMINATION DIFFERENCES USING THE LIGHTING RATIO

A. Lighting Ratio

The self-Quotient image \( Q \) is defined as an intrinsic property of a face image \( I \) [16], by \( Q = I/I \), and \( \hat{I} \) is the smoothed version of \( I \). Computing the ratio between \( I \) and \( \hat{I} \) is a popular way to remove the lighting effects from the face. Wang et al. [16] used the spatial 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, the range of a self-quotient image spans from zero to infinity, 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 [16]. Meanwhile, using a spatial filter to obtain the smoothed version of \( I \) raises two different concerns. First, it is difficult to determine the appropriate sliding window size for filtering since the size depends on the face scale and the illumination conditions. Second, high dynamic range and improper exposure problems have not been addressed. These challenges affect the contrast and offset of the facial texture. 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 [2], where under the Lambertian assumption, the low frequency part in the image captures mostly the illumination conditions on the facial image. Thus, an alternative way to overcome the aforementioned problems is to decompose the facial texture into the lighting ratio and facial albedo (Eq. 1). The process can be written as:

\[
A = \frac{I}{(\alpha I + \beta)} = \frac{I}{\alpha \mathcal{F}^{-1}(\Gamma(\mathcal{F}(I))) + \beta}
\]

where \( A \) denotes the lighting ratio, \( \Gamma() \) denotes an image-specific low-pass filtering operation, \( \mathcal{F} \) and \( \mathcal{F}^{-1} \) denote the Fourier transform and the inverse Fourier transform, and \( \alpha, \beta \) are scale and offset parameters. The input image is denoted by \( I \), while \( \hat{I} \) denotes the smoothed version of the image \( I \). The parameters \( \alpha, \beta \) compensate for the contrast and offset variations in \( I \). These two parameters are set through the optimization step in the algorithm in Sec. III-C. All of the operations (i.e., addition, division, and multiplication) are performed in a pixel-wise manner.

B. Image-Specific Low-pass Filtering

It is hard to choose an arbitrary cut-off frequency for low-pass filters for all facial images since the illumination conditions vary in different facial images. Instead, we propose to automatically compute an image-specific energy threshold \( P \) (in the range \([0,1]\)) for each image. The basic idea is that the number of lights, light direction, and intensity change the frequency distribution mostly in the low frequencies. The portion of energy useful for identification is distributed mostly in middle frequencies, while the skin details (e.g., facial component edges, wrinkles) are distributed mostly in high frequencies and remain relatively constant. The energy variations among fixed pose facial textures are highly related to the illumination conditions. Thus, by varying the cut-off energy instead of the cut-off frequency, we can estimate the illumination conditions more accurately. Note that the energy in Fourier spectrum \( F_g \) is defined as the \( L_2 \) norm of the \( F_g \) magnitude. Thus, the cut-off frequency is iteratively increased until the energy of the passed frequency components is greater than a portion \( P \) of the energy of all frequencies of the image.

For each image, we adaptively compute \( P \) based on the energy associated with the illumination conditions in the image. Histogram equalization \( \mathcal{H}() \) 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 and across various illumination conditions. Thus, we approximate \( P \) using the energy of the histogram equalized adjusted image as a reference:

\[
P \propto E(I)/E(I) \approx \kappa E(I)/E(\mathcal{H}(I)),
\]

where \( \kappa \) is a constant, \( E(I) \) is the energy of the input image \( I \), and \( E(\mathcal{H}(I)) \) is the energy of the input image processed
by the histogram equalization. The energy is defined as the $L_2$ norm of pixel values of their gray level images. The term $E(I)$ consists of two parts: the energy due to the identification and the energy due to the illumination conditions. The energy of identification is quite constant across images and the energy of the Histogram Equalization (HE) adjusted image has a relative constant value. Thus, when performing the division, $P$ is able to represent the energy variations caused by illumination conditions.

The choice of the low-pass filter is also important in estimating the illumination conditions. The ideal low pass filter introduces a 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). 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. Therefore, we adopt a Gaussian filter $\xi$ in this work.

### C. Unlighting and Relighting Algorithm

The unlighting algorithm based on the lighting ratio is presented in Alg. 1. We define a constant image $I_p$ (all pixel values set to a constant value $\psi$) and minimize the objective function: $\arg\min_{\alpha, \beta} \| I_p - A \|_F$. There are three positive effects from this algorithm. First, the lighting ratios $A$ are similar to each other in terms of global energy since we adjust the illumination conditions according to one preset texture. Second, the illumination conditions on $A$ are more uniform since $I_p$ is symmetric and the parameters $\alpha, \beta$ adjust the contrast and offset variations on the raw images $I$. Finally, the estimation of $I$ is more accurate since it is performed using an image-specific low-pass filtering operation, which avoids the use of an arbitrary low-pass filtering parameter for all textures.

#### Algorithm 1 Illumination normalization using lighting ratio

**Input:** Facial texture $I$ and low-pass filter $\xi$  
**Output:** Lighting Ratio $A$

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 the 2D Fast Fourier Transform
3. Compute the image-specific parameter $P$ for $\xi$ (Sec. III-B)
4. Filter the magnitude of $F_g$ by applying the low-pass filter $\xi$ to obtain $F_g'$
5. Apply the inverse Fast Fourier transform on the filtered spectrum $F_g'$ to obtain the illumination component $V_g$
6. Compute the ($\alpha, \beta$) that minimize the Frobenius norm between $I_p$ and $V_g/(\alpha V_g + \beta)$
7. Compute $A$ with $\alpha, \beta$ as in Eq 1

The relighting algorithm works in a similar way as unlighting. Instead of using the preset image, we use facial textures in the probe as reference images. Meanwhile, relighting is achieved by multiplying the lighting ratio of the target image $A_g$ with the illumination condition estimate of the reference image $I_p'$. The relighting algorithm changes the lighting ratio to minimize the Frobenius norm between the relit image $I_g'$ and an reference image $I_p'$:

$$\arg\min_{\alpha, \beta} \| I_p - I_g' \|_F = \arg\min_{\alpha, \beta} \| I_p - I_p A_g \|_F$$  \hspace{1cm} (3)

We assume that the gallery and probe facial textures have a dense correspondence. This is accomplished by the 3D-2D face recognition system, where the texture lifting module outputs pairs of co-registered gallery and probe facial textures with normalized pose in a two-dimensional space [15]. Alternatively, an Active Shape Model [7] or an Active Appearance Model [6] can be applied to register two facial textures via an affine transform [5]. However, these two methods normally require more than 80 landmarks manually labelled for training while the UR2D system [15] only needs nine. Moreover, AAM and ASM are sensitive to head pose, while the UR2D texture lifting module is capable of handling varying poses by taking advantage of the 3D face data and is able to output a visible mask to exclude the self-occluded regions in the original pose. The detailed algorithm is depicted in Alg. 2.

#### Algorithm 2 Relighting using lighting ratio

**Input:** Target facial texture $I_g$, reference facial texture $I_p$, low pass filter $\xi$  
**Output:** Relit target $T_g$

1. Convert the textures $I_g$ and $I_p$ into the HSV color space and extract their $V$ channels, named $V_g$ and $V_p$, respectively
2. Obtain the Fourier spectrum $F_g$, $F_p$ from $V_g$ and $V_p$ via 2D Fast Fourier Transform
3. Filter the magnitudes of $F_g$ and $F_p$ by applying the low pass filter $\xi$ to obtain $F_g'$ and $F_p'$, whose energy is a portion $P$ of the overall energy of $F_g$ and $F_p$
4. Apply inverse Fast Fourier transform on the filtered spectrum $F_g'$ and $F_p'$ to obtain the lighting components $V_g$ and $V_p$
5. Compute the ($\alpha, \beta$) that minimize the Frobenius norm between $V_p$ and $\hat{V}_p A_g$
6. Compute the relit gallery: $T_g = \hat{V}_p A_g$ using $\alpha, \beta$

The optimum values $\hat{\alpha}$ and $\hat{\beta}$ provide the best scale and offset to make the relit gallery as similar as possible to the probe in terms of global energy. The Nelder-Mead simplex algorithm [11] is chosen for the minimization because it is one of the best known and efficient algorithms for unconstrained optimization without derivatives.

### IV. Experimental Results

#### A. Datasets

In order to assess the robustness of our algorithm under various head poses, we first tested it on the publicly available dataset UHDB11 [15]. It consists of 1,625 3D facial scans
Fig. 1. Depiction of the relighting result. The first row depicts the original target face texture. The second row depicts the reference face texture. The third row depicts the relit target texture using the ideal filter with a fixed cut-off frequency. The fourth row depicts the relit target texture using the Gaussian filter with a fixed cut-off frequency. The last row depicts the target texture relit using our image-specific low-pass filtering algorithm.

captured by 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 number of samples in the gallery is only 23 textured 3D facial data but the number of probe samples is 1602 images. This setup results in an experiment with 36,846 comparisons of pairs. 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 illumination conditions. The term \( \psi \) was empirically set to 80.

B. Relighting and Unlighting

Figure 1 depicts the result of relighting. The first row includes three target faces, while the second row includes three reference faces. The following rows depict the relit faces in the first row using data in the second row as the reference. Note that the lighting on relit faces using the ideal filter (i.e., the third row) is not distributed as smoothly as in the images relit using the Gaussian filter (i.e., the fourth row). Comparing the relit faces using the fixed cut-off frequency as in the fourth row, we can observe that the lighting effects appear to be better estimated and transferred using our proposed image-specific low-pass filtering technique, which avoids an arbitrary choice of the cut-off frequency and selects the parameters according to specific image’s lighting condition.

Fig. 2. Depiction of the unlighting results. The first column depicts the original face textures in the UV space. The second column depicts the self-ratio images. The third column depicts the unlit face textures using our method.
Figure 2 depicts the comparison between the self-ratio and the proposed unlighting algorithm (UHLR-u). Note that the lighting effects on the unlit facial texture are reduced and distributed more evenly than those on the raw facial texture. When compared to the images obtained using 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 ratio adjusts 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 illumination conditions. Thus, the differences among unlit textures are mainly caused by the different identities.

### C. Face Recognition

We use the UR2D face recognition system [15]. This system uses an asymmetric face recognition pipeline which is capable of handling images that have variations in head pose. The 3D textured data in the gallery can be used to register 2D images under various poses, either in the gallery or in the probe, into a canonical 2D image space where the head pose is normalized to a frontal pose.

For each 3D mesh in the gallery, a set of 3D landmarks are first detected and then the AFM is fitted. As a result, a point-to-point correspondence from the vertices on the fitted mesh to pixels on the 2D space is established. For a 2D image in the gallery, we first detect the face and find the same set of landmarks as those on the mesh. In order to estimate the pose of the face, the pose estimation module uses the two sets of 2D and 3D landmarks and outputs the rotation and translation parameters along with a 3D-to-2D projection matrix for the mesh. Then, the corresponding 3D mesh is projected on the image in order to lift the texture. Using the 3D coordinate system of the AFM as a bridge, the face texture is mapped onto a space where the pose is normalized to be frontal. For each image in the probe, we first detect the face and the same set of 2D landmarks. Then, a fitted mesh in the gallery is used for pose estimation and texture lifting. The mesh is selected from gallery with which the probe image is being compared. Then, illumination normalization is performed followed by the computation of a correlation-coefficient-based distance metric for each pair of textures from the gallery and the probe. In the bi-directional relighting experiment, two scores are computed. First, one score is computed for each pair of relit-gallery and probe and then the second score is computed for each pair of gallery and relit-probe. These two scores are fused by the sum rule for the final score. The sum rule is selected because of its most resilience to estimation errors according to a sensitivity analysis [10].

Figure 3 depicts the ROC curves for six different methods: (i) Baseline (B): computing the distance metric using the raw output of texture lifting; (ii) relighting Gallery based on lighting ratio (UHLR-g): computing the distance metric between the probes and the relit galleries using the proposed algorithm; (iii) relighting Probe based on lighting ratio (UHLR-p): computing the distance metric between the galleries and the relit probes using the proposed algorithm; (iv) bi-directional relighting based on lighting ratio (UHLR-b): computing the sum of two distance metrics using the relit textures (one score from relit gallery according to probe and one from relit probe referring to gallery) using the proposed algorithm; (v) BRDF-based relighting (RO): computing the distance metric between the probes and the relit galleries using method in [15]; (vi) Self-ratio method (SR): computing the distance metric between the unlit gallery and the unlit probe using Self Ratio image method in [16]; (vii) unlighting based on lighting ratio (UHLR-u): computing the distance metric between the unlit gallery and the unlit probe using the proposed algorithm. The verification rates at $10^{-3}$ FAR are 52.3% (B), 63.9% (UHLR-g), 61.7% (UHLR-p), 69.1% (UHLR-b), 64.1% (RO), 61.0% (SR) and 68.9% (UHLR-u), respectively. All the illumination normalization methods demonstrated their ability to improve the face verification performance. When we relight galleries referred to probes, the UHLR-g and relighting method in [15] achieved almost similar verification rate, but our method does not require 3D normals. The scheme of relighting galleries referred to probes (UHLR-g) achieved better verification rate than the one relighting probes referred to galleries (UHLR-p). The best result is achieved by the proposed method when used in a bidirectional scheme. This is because the distance metrics are in the same range, the sum rule combined the discriminability from both and boosted the verification rate. The UHLR-u algorithm has also demonstrated improvement over the self-quotient method because of a better estimation of illumination conditions via the image-specific low-pass filtering and a better normalization process with the adoption of parameters $\alpha$ and $\beta$.

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 [1] (i.e., the same 250 facial scans from 250 subjects as the gallery and 470 facial scans as the probe). Figure 4 depicts the ROC curves of the score metrics from
of the baseline, the proposed relighting method (UHLR-g and UHLR-b), and the proposed unlighting method (UHLR-u). Our algorithm achieves verification rates of 48.1% (UHLR-g), 52.9% (UHLR-b) and 53.8% (UHLR-u) at $10^{-3}$ False Accept Rate. The UHLR-u algorithm performs better than the UHLR-b since many textures in this dataset have more than one dominant lighting sources, which spans more lighting effects out of the low frequency components. Thus, the cut-off frequency rises accordingly in the image specific filtering. When conducting the relighting, identification information in the low-medium frequencies from one subject may be mixed with the identification in the higher higher frequencies from another subject, which causes the ambiguity in identification. However, compared to the verification rates of 20.43% and 34.89% achieved in [1], depicted as the red and green dot lines in Fig.3, the proposed approach still achieves a better performance.

When one relighting algorithm is required in face recognition, we opt to use the UHLR-b because of the reduced computational complexity compared with [15]. The UHLR-b reduces the computational cost from around 150 s for one call of the relighting algorithm [15] to 14.3 s on average on a computer with Xeon E5507 @2.27Ghz CPU and 8 GB memory. This decreases the time for the relighting to 10% of the original time. Even when the UHLR-b has to run the relighting algorithm twice per pair, we still can decrease the time for the relighting module by 80%.

V. CONCLUSION

We proposed a lighting-ratio based illumination alignment approach which can be used for both relighting and unlighting, without requiring 3D geometry information and light information. Compared with the self-ratio method, our major improvements are the adoption of image-specific low-pass filtering and the global correction on contrast and offset refer to another image. We tested the algorithm on the publically available UHDB11 framework, and FRGC databases. The results of the proposed framework demonstrate the effectiveness and accuracy.

REFERENCES