Improving Shadow Suppression for Illumination Robust Face Recognition

Wuming Zhang, Xi Zhao, Jean-Marie Morvan, Liming Chen

Improving Shadow Suppression for Illumination Robust Face Recognition

Wuming Zhang, Xi Zhao, Jean-Marie Morvan and Liming Chen, *Senior Member, IEEE*

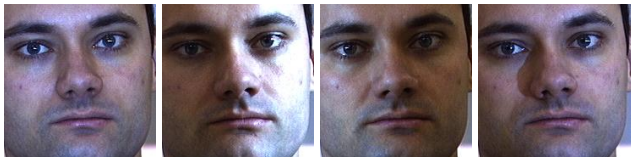
Abstract—2D face analysis techniques, such as face landmarking, face recognition and face verification, are reasonably dependent on illumination conditions which are usually uncontrolled and unpredictable in the real world. The current massive data-driven approach, *e.g.*, deep learning-based face recognition, requires a huge amount of labeled training face data that hardly cover the infinite lighting variations that can be encountered in real-life applications. An illumination robust preprocessing method thus remains a very interesting but also a significant challenge in reliable face analysis. In this paper we propose a novel model driven approach to improve lighting normalization of face images. Specifically, we propose to build the underlying reflectance model which characterizes interactions between skin surface, lighting source and camera sensor, and elaborate the formation of face color appearance. The proposed illumination processing pipeline enables generation of the Chromaticity Intrinsic Image (CII) in a log chromaticity space which is robust to illumination variations. Moreover, as an advantage over most prevailing methods, a photo-realistic color face image is subsequently reconstructed, which eliminates a wide variety of shadows whilst retaining the color information and identity details. Experimental results under different scenarios and using various face databases show the effectiveness of the proposed approach in dealing with lighting variations, including both soft and hard shadows, in face recognition.

Index Terms—Face recognition, lighting normalization, illumination and texture analysis

1 Introduction

Face analysis has received considerable attention due to the enormous developments in the field of biometric recognition and machine learning. Beyond its scientific interest, face analysis offers unmatched advantages for a wide variety of potential applications in commerce and law enforcement compared to other biometrics, such as easy access or avoidance of explicit cooperation from users [1]. Nowadays, conventional cases have attained quasi-perfect performance in a highly constrained environment wherein poses, illuminations, expressions and other non-identity factors are controlled. However, these approaches suffer from a very restricted range of application fields due to the non-ideal imaging environments frequently encountered in practical cases: users may present their faces without a neutral expression, or human faces may come with unexpected occlusions such as sunglasses, or, yet again, the images are captured from video surveillance that may group all difficulties such as low resolution images, pose changes, lighting condition variations, *etc.* In order to be adaptive to these challenges in practice, both academic and industrial research have understandably shifted their focus to unconstrained real-scene face images.

- *(Corresponding author: Xi Zhao.)*
- *W. Zhang and L. Chen are with the Laboratory of LIRIS (CNRS UMR 5205), Department of Mathematics and Computer Science, Ecole Centrale de Lyon, University of Lyon, 69310 Ecully, France. E-mail: {wuming.zhang, liming.chen}@ec-lyon.fr*
- *X. Zhao is with the School of Management, Xi'an Jiaotong University, Xi'an 710049, China. E-mail: zhaoxi1@hotmail.com*
- *J. M. Morvan is with Université Lyon 1, Institut Camille Jordan, CNRS UMR 5208, 43 blvd du 11 Novembre 1918, F-69622 Villeurbanne-Cedex, France, and King Abdullah University of Science and Technology, Visual Computing Center, Bldg 1, Thuwal 23955-6900, Saudi Arabia. E-mail: jean-marie.morvan@kaust.edu.sa*



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

Fig. 1. An example of varying lighting conditions for the same face. (a) Front lighting; (b) Specular highlight due to glaring light coming from right side; (c) Soft shadows and (d) hard-edged cast shadow.

Compared with other nuisance factors such as pose and expression, illumination variation impinges more strongly upon many conventional face analysis algorithms that assume a normalized lighting condition. As depicted in Fig. 1, the lighting condition can be fairly complicated due to numerous issues, *e.g.*, the intensity and direction of the lighting, or the overexposure or underexposure of the camera sensor. Moreover, it has already been proven that in face recognition, differences caused by lighting changes could be even more significant than differences between individuals [2]. The current state of the art massive data-driven approach, *e.g.*, deep learning-based face recognition [3], requires a huge amount of labeled face data which, however, are unable to cover the infinite illumination variations that can occur in real-life applications. Therefore, illuminant-invariant approaches based on lighting normalization continue to be crucially important for further widening the application field of face recognition.

In this paper, we propose a novel model driven-based lighting normalization approach for the purpose of lighting variation robust 2D face recognition. Specifically, we first divide the whole face into highlighted and non-highlighted

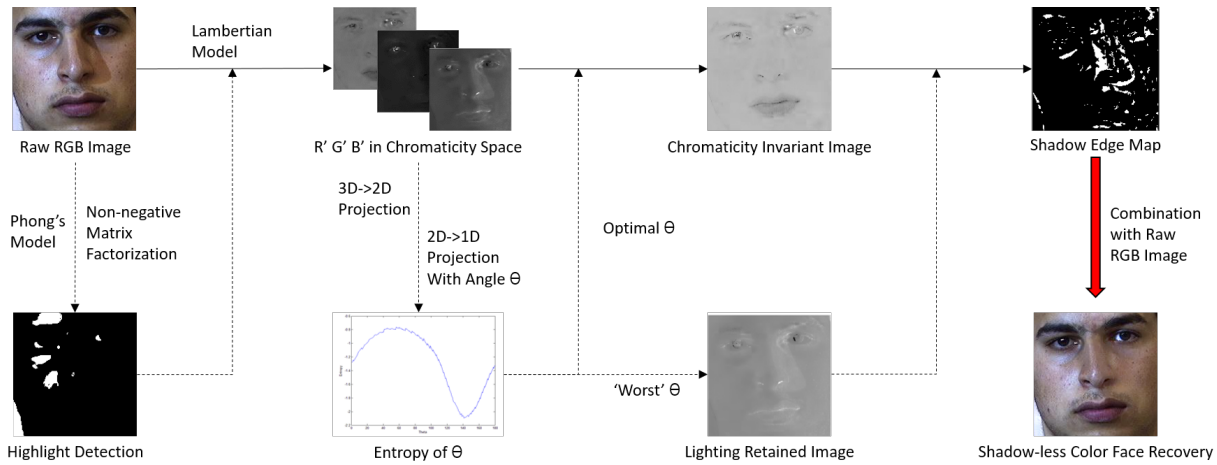


Fig. 2. Overview of the chromaticity space-based lighting normalization process and shadow-free color face recovery process.

regions. Second, we approximate Lambertian surfaces and Planckian lighting in order to investigate image formation rules. Then, a pixel-level transformation in log space is constructed with a view to pursuing a chromaticity invariant representation. The final step is to extend this chromaticity invariance to color space by taking shadow edge detection into account. An overview of the proposed processing method is illustrated in Fig. 2. Ultimately, the experiments are carried out based on lighting normalized images, and favorable experimental results have been achieved on the CMU-PIE and the FRGC face database. Our specific contributions are listed as follows.

1) We introduce and develop a chromaticity-based physical interpretation for modeling the face imaging process, which takes highlight detection as preprocessing and is able to separate the illumination effect from intrinsic face reflectance.
2) We present a novel application of the chromaticity invariant image for shadow-free color face reconstruction rather than gray-scale level de-lighting, demonstrating the potential for recovering a photo-realistic face while eliminating the lighting effect.
3) We evaluate the proposed method on two benchmarking datasets across illumination variations and demonstrate that it can help improve performance of state-of-the-art methods especially on hard shadows, both qualitatively and quantitatively.

The remainder of this paper is structured as follows: Section 2 briefly overviews related work in illumination invariant face recognition; Section 3 describes the color formation principles of human faces in the RGB space, while Section 4 details an illumination-normalized intrinsic image formation algorithm in chromaticity space; in Section 5 this invariance is further studied to enable full color shadow-free face recovery; promising experimental results and conclusions are given in Section 6 and Section 7, respectively.

2 Related Work

Over the years, a surge of qualitative and quantitative studies on illumination invariant research has been observed due to the suitability and efficacy of such techniques in face analysis. These techniques could be roughly divided into three categories according to their diverse theoretical backgrounds: holistic normalization methods, invariant feature extraction methods, and 3D model-based methods.

Holistic normalization-based approaches used to be common in early algorithms. These attempt to redistribute the intensities of the original face image in a more normalized representation, which is less prone to lighting changes by applying a simple gray-scale intensity adjustment. Histogram Equalization (HE) and Histogram Matching (HM) [4] initiated these methods by adopting an image preprocessing stage on histogram level. Shan *et al.* [5] developed Gamma Intensity Correction (GIC) for normalizing overall image intensity at a given illumination level, and introduced an intensity mapping: $G(x,y) = cI(x,y)^{1/\gamma}$ where c is a gray stretch parameter and γ is the Gamma coefficient. Notwithstanding their ease of implementation and the apparent beneficial effects on lighting normalization, these methods fail to further satisfy the increasingly rigorous demands on accuracy as they are global and do not take into account the in-depth image formation principles. This means that they only average holistic intensity distribution and cannot satisfactorily handle soft shadow, hard shadow or highlight, respectively.

In view of this deficiency of holistic normalization, invariant feature extraction methods are proposed. Extraction of illumination-invariant components from the frequency domain is the mainstream approach yielding implementation of wavelet-based denoising [6] or logarithmic discrete cosine transform (LDCT) [7]. Derived from Land's Retinex model [8] and its variants, which indicate that a face image could be broken down into its smoothed version and its illumination invariant features, Riklin-Raviv and Shashua [9] proved that Quotient Image (QI), *i.e.* a ratio image between a test image and a linear combination of three prototype images based on the Lambertian model, is illumination free. The algorithm is then generalized by Wang *et al.* [10] to the Self Quotient Image (SQI), which replaced the prototype images by a smoothed version of the test image itself. SQI achieved predominant performance while suffering from a lack of edge-preserving capability caused by their weighted

Gaussian filter. Chen *et al.* [11] utilized the TV-L^1 model for factorizing an image and succeeded in overcoming this drawback. Local Normalization (LN) was proposed by Xie *et al.* [12] to cope with uneven lighting conditions by reducing or even removing the effect of noise. Gradientface [13] and Weberface [14] compute the ratio of x-gradient to y-gradient and the ratio of the local intensity variation to the background of a given image, respectively, to obtain illumination invariant representations. An integrative preprocessing chain was created by Tan and Triggs [15], who successively merged Gamma correction, difference of Gaussian filtering, optional masking, and contrast equalization. All these approaches achieved impressive performance on removing soft shadows, yet encountered problems with hard-edged cast shadows especially caused by self-occlusion around the nose. Moreover, these methods cannot be extended to color space, resulting in limited applications in the real world.

With the ever-advancing development of 3D data acquisition technologies, many researchers turned their attention to 3D model estimation based upon physical principles for dealing with lighting problems. Basri *et al.* [16] proved that a convex Lambertian object obtained under a large variety of lighting conditions can be approximated by a 9D linear subspace. Blanz and Vetter [17] first proposed the 3D Morphable Model (3DMM) to estimate and synthesize lighting conditions by means of a linear combination of prototype models. A publicly available 3D Morphable Face Model - the Basel Face Model (BFM) [18] - was then constructed to realize the widespread use of 3DMM. Wang *et al.* [19] presented the Spherical Harmonic Basis Morphable Model (SHBMM), fusing 3DMM and spherical harmonic illumination representation [16]. Based on physical lighting models, Zhao *et al.* [20] decomposed lighting effects using ambient, diffuse, and specular lighting maps and estimated the albedo for face images with drastic lighting conditions. 3D-based lighting independent methods are powerful and accurate in comparison with 2D-based ones. However, they are easily confined to data acquisition and are limited by the unavoidable high computational cost. Even if we can compromise by considering only 2D images and normalizing their lightings using 3D models, data registration between 2D and 3D remains likewise an inconvenience.

To summarize, the proposed approach in this paper, which is actually a fusion of holistic normalization and the reflectance model, introduces, for the first time, the usage of the chromaticity invariant image in the field of face analysis to reconstruct a shadow-free color face image without using 3D priors. Compared with existing methods, we have constructed a comprehensive framework which combines the physical interpretation of face imaging and the simplicity of implementation. Moreover, since the proposed method removes shadow in color space, it can jointly function with other gray-scale level techniques to improve lighting normalization performance.

3 Skin Color Analysis

In this section, we formulate a physics-based reflectance model for approximating pixel-based face skin colors. To begin with, we recapitulate the definition and properties of the two most commonly used reflectance models. A non-negative matrix factorization (NMF) based method is then implemented to locate the highlighted facial region, which is less informative for precise model formulation. A product-form representation, which could account for diffuse color, is finally proposed as the cornerstone for our approach.

3.1 Reflectance Model: Lambert vs. Phong

Despite the appearance of several more comprehensive and more accurate BRDF models in recent years, these models are practically constrained by computational burden and become strongly ill-posed with respect to inverse estimation of material reflectance, thus greatly restricting their application in general lighting normalization tasks. Instead, classical models like Lambert and Phong [21] still occupy a prime position in this field due to their ease of implementation.

As a common assumption, Lambert and Phong both adopt the concept of ideal matte surface, obeying Lambert's cosine law where the incident lighting arriving at any point of an object surface is uniformly diffused in all observation directions. Furthermore, Phong's model extends Lambertian reflectance by adding a specular highlight modelization term, which is merely dependent on the object's geometric information and lighting direction at each surface point. The representation of the Lambertian model and Phong's model can be formulated by equation (1) and (2), respectively,

$$L_{diffuse} = S_d E_d(\boldsymbol{n} \cdot \boldsymbol{l}) \tag{1}$$

$$L_{diffuse} + L_{specular} = S_d E_d(\boldsymbol{n} \cdot \boldsymbol{l}) + S_s E_s(\boldsymbol{v} \cdot \boldsymbol{r})^{\gamma} \tag{2}$$

where S_d and S_s denote the diffuse and specular reflection coefficients; E_d and E_s represent the diffuse and specular lighting intensities; $\boldsymbol{n}$, $\boldsymbol{v}$, $\boldsymbol{l}$ and $\boldsymbol{r} = 2(\boldsymbol{n} \cdot \boldsymbol{l})\boldsymbol{n} - \boldsymbol{l}$ refer to the normal vector, the viewer direction, the direction of incident light and the direction of the perfectly reflected ray of light for each surface point; γ is a shininess constant.

Despite the fact that the human face is neither pure Lambertian (as it does not account for specularities) nor entirely convex, the simplifying Lambertian assumption is still widely adopted in face recognition studies [16], [22], [23] as the face skin is mostly a Lambertian surface [24]. Nevertheless, premising the work on this assumption would be suboptimal because the specular highlight widely occurs in practice and could not be ignored in face images due to the inevitable existence of the oil coating and semi-transparent particles in the skin surface. To address this problem, we decide to first detect the highlight region on each face image using the Phong-type model. The classical Lambertian reflectance will then be applied afterwards to the skin color analysis for the non-highlighted region.

3.2 Specular Highlight Detection

As was proven in [25], variations in density and distribution of skin pigments, such as melanin and hemoglobin, simply scales the skin reflectance function, *i.e.* $S_d(\boldsymbol{x}, \lambda) = \beta(\boldsymbol{x}) S_d(\lambda)$. Here $\boldsymbol{x}$ denotes the spatial coordinates. Furthermore, as stated in [26], the spectrum of surface-reflected light for specular spots in face skin can be considered to be equal to the spectrum of source lighting, *i.e.* $S_s = 1$, otherwise $S_s = 0$ for non-highlighted regions. With these

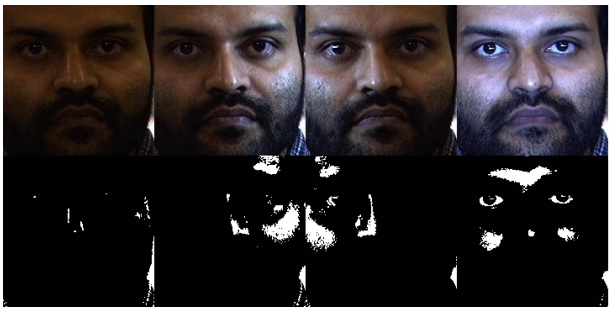


Fig. 3. Specular highlight detection results on images under various lighting conditions. Top: raw images; bottom: detected highlight masks.

caveats in mind, each component in Phong's model could be divided into an achromatic term (decided only by geometric parameters) and a chromatic term (parametrized by λ):

$$L(\boldsymbol{x},\lambda) = (\boldsymbol{n}\cdot\boldsymbol{l})\beta(\boldsymbol{x})E_d(\lambda) + (\boldsymbol{v}\cdot\boldsymbol{h})^{\gamma}S_s(\boldsymbol{x})E_s(\lambda) \quad (3)$$

More specifically, the RGB responses could be rewritten as spatial coordinates determined by geometrical dependency in space, spanned by the colors of light and surface:

$$\begin{bmatrix} R(\boldsymbol{x}) \\ G(\boldsymbol{x}) \\ B(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} R_d & R_s \\ G_d & G_s \\ B_d & B_s \end{bmatrix} \times \begin{bmatrix} k_d(\boldsymbol{x}) \\ k_s(\boldsymbol{x}) \end{bmatrix} \quad (4)$$

where the first term of the right-hand side is a 3×2 matrix representing RGB channel magnitudes for diffuse and specular reflection, while the second achromatic term is a 2×N matrix (N denotes the number of pixels) containing diffuse and specular coefficients.

Remarkably, all these matrices are non-negative and $k_s(\boldsymbol{x})$ is sparse due to the fact that only a small portion of face contains specularity. It then becomes natural to consider the use of Non-negative Matrix Factorization (NMF) [27] for solving such a $\boldsymbol{V} = \boldsymbol{W}\cdot\boldsymbol{H}$ problem. Implementation is easy: we set the inner dimension of factorization to 2 and apply a sparse constraint for $k_s(\boldsymbol{x})$ by restricting its L_1 norm while fixing its L_2 norm to unity as a matter of convenience.

As illustrated in Fig. 3, performance of highlight detection using the proposed method proves to be robust, irrespective of lighting intensity and lighting direction for face images under different illumination environments. In particular, NMF may not be able to distinguish specular and diffuse properties under low illumination, but in this case the shininess constant γ in equation (2) becomes very small and thus the specular reflection can be ignored.

3.3 Skin Color Formation

After successful separation of the surface-reflected region from the body-reflected region, our focus will be to investigate skin color formation on the non-highlighted area using the Lambertian model. Conceptually, three primary factors may be involved in a comprehensive image formation scene: source lighting, object surface, and imaging sensor. Each factor is physically modeled, based on which the definitive color representation will be straightforwardly derived.

First, we assume that the source illuminations are Planckian, which could cover most lighting conditions such as daylight and LED lamps, *i.e.* the spectral radiance of lighting could be formulated by $B(\lambda,T) = \frac{2hc^2}{\lambda^5}\frac{1}{e^{hc/\lambda k_B T}-1}$ where $h = 6.626\times10^{-34}J\cdot s$ and $k_B = 1.381\times10^{-23}J\cdot k^{-1}$ are the Planck constant and the Boltzmann constant, respectively; λ characterizes the lighting spectrum; T represents the lighting color temperature, and $c = 3\times10^8 m\cdot s^{-1}$ gives the speed of light in the medium. Additionally, since the visible spectrum for the human eye always falls on high frequencies where $hc/\lambda \gg k_B T$, spectral power distribution $E(\lambda,T)$ of illumination with an overall intensity I tends to Wien's approximation [28]:

$$E(\lambda,T) \simeq I\frac{k_1}{\lambda^5}e^{-\frac{k_2}{\lambda T}} \quad (5)$$

where $k_1 = 2hc^2$ and $k_2 = \frac{hc}{k_B}$ refer to the first and second radiation constants. Moreover, as proven in [29], the Planckian characteristic can be approximately considered linear, thus allowing us to generalize this assumption to a bi-illuminant or multi-illuminant scene.

The assumption for skin surface has already been formulated, *i.e.* the skin is a Lambertian surface and follows the reflection rule specified in (1). With the sensor response curve $F_i(\lambda)$ corresponding to three color channels, the spectral reflectance function of skin surface $S(\lambda)$ and the aforementioned spectral power distribution $E(\lambda)$, the final output of camera sensors $\boldsymbol{C} = \{R,G,B\}$ could be represented as an integral of their product over the spectrum:

$$C_i = \int F_i(\lambda)E(\lambda)S(\lambda)(\boldsymbol{n_k}\cdot\boldsymbol{l})d\lambda, \quad i = 1,2,3 \quad (6)$$

where $(\boldsymbol{n_k}\cdot\boldsymbol{l})$ describes the inner product between surface normal and illumination direction. Given a specific scene and geometry, this product value for each surface point is fixed to a constant α.

A widely used assumption in computer graphics, which is subsequently adopted here, is that camera sensors are sufficiently sharp and that their spectral sensibility could be characterized by Dirac delta function $F_i(\lambda) = f_i\delta(\lambda-\lambda_i)$. This satisfies $\int F_i(\lambda)d\lambda = f_i$ and turns the integral representation in (6) into a multiplicative form in (7).

$$C_i = \alpha f_i E(\lambda_i)S(\lambda_i), \quad i = 1,2,3 \quad (7)$$

Eventually, a comprehensive representation of color formation emerges after a combination of (5) and (7):

$$C_i = \alpha I k_1 f_i \lambda_i^{-5} e^{-\frac{k_2}{\lambda_i T}} S(\lambda_i), \quad i = 1,2,3 \quad (8)$$

An apparent truth about this formula is that the color value for one skin surface point can be practically compartmentalized into three segments: a constant part ($\alpha I k_1$), a lighting (T) invariant yet channel (λ_i) related part ($f_i\lambda_i^{-5}S(\lambda_i)$), and a lighting related part ($e^{-\frac{k_2}{\lambda_i T}}$). This thought-provoking observation instantly prompts us to first carry out a normalization processing to remove the constant part and then separate the channel related part and the lighting related part. Not surprisingly, the property of intensity normalization in chromaticity space, together with the attendant investigation of the chromaticity invariant image, have been brought to our attention.

4 Chromaticity Invariant Image

The target of inferring an illumination-invariant face image based upon the previously derived skin model in chromaticity space is discussed and achieved in this section. We first recall the definition of chromaticity, whereafter an intrinsic characteristic of the chromaticity image in log space is studied, leading to the following gray-scale chromaticity invariant face image formation.

4.1 Skin Model in Chromaticity Space

Chromaticity [29], [30], generally considered as an objective specification of the quality of color regardless of its luminance, is always defined by intensity normalized affine coordinates with respect to a tristimulus color space, such as CIEXYZ or RGB utilized in our case. The normalization mapping mainly contains two modalities: L1-normalization: $\boldsymbol{c} = \{r, g, b\} = \{R, G, B\}/(R+G+B)$ and geometric mean normalization: $\boldsymbol{c} = \{r, g, b\} = \{R, G, B\}/\sqrt[3]{R*G*B}$. In both normalization methods, all colors are regularized to equiluminous ones, which helps to attenuate the effect of the intensity component.

For computational efficiency and further extension, the geometric-mean-normalized chromaticity is implemented as a processing pipeline for skin color in (8). The $\boldsymbol{c} = \{r, g, b\}$ values in chromaticity space are given as follows:

$$c_i = \frac{f_i \lambda_i^{-5} S(\lambda_i)}{(\prod_{j=1}^{3} f_j \lambda_j^{-5} S(\lambda_j))^{\frac{1}{3}}} \frac{e^{-\frac{k_2}{\lambda_i T}}}{e^{\frac{1}{3}\sum_{j=1}^{3} -\frac{k_2}{\lambda_j T}}}, \quad i = 1, 2, 3 \qquad (9)$$

Within this chromaticity representation, all constant terms are normalized. The two remaining terms consist of a channel-related one and a lighting-related one. If we switch our focus back to the process of highlight detection in the previous section, which aims at separating specular reflection from diffuse reflection, the explanation could be sufficient: only under the assumption of the Lambertian model can we be capable of normalizing the constant terms benefiting from the multiplicative representation of color.

So far, we solidify and parametrize an exhaustive color formation model in a concise form. More specifically, this representation could be naturally considered as an aggregation of a lighting-invariant part and another lighting-related part, thus providing us with the opportunity to further explore illumination invariant components.

4.2 Chromaticity Invariant Image Generation

When investigating the characteristics of the skin model in chromaticity space, both its multiplicative form and the exponential terms easily guide us to logarithm processing, which is capable of transforming (9) into:

$$\psi_i = \log(c_i) = \log \frac{W_i}{W} + (-\frac{k_2}{\lambda_i} - \frac{1}{3}\sum_{j=1}^{3} -\frac{k_2}{\lambda_j})/T, \qquad (10)$$

with the lighting-invariant components $W_i = f_i \lambda_i^{-5} S(\lambda_i)$ and $W = (\prod_{j=1}^{3} f_j \lambda_j^{-5} S(\lambda_j))^{\frac{1}{3}}$.

It is noticeable that all three chromaticity color channels in log space are characterized by the same lighting color T, which implies the potential linear correlation among these values. Let us now consider another fact: $c_1 * c_2 * c_3 = 1$ since they are geometric mean normalized values, it could be equally inferred that in log space we have $\psi_1 + \psi_2 + \psi_3 = 0$, illustrating that all chromaticity points $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3)$ in 3D log space actually fall onto a specific plane perpendicular to its unit normal vector $\boldsymbol{u} = 1/\sqrt{3}(1, 1, 1)$.

Up to now, the dimensionality of target space has been reduced to 2. It is now reasonable to introduce a 3D-2D projection in order to make the geometric significance more intuitive. Derived from the projector $\boldsymbol{P_u^{\perp}} = \boldsymbol{I} - \boldsymbol{u^T u} = \boldsymbol{U^T U}$ onto this plane, $\boldsymbol{U} = [\boldsymbol{u_1}; \boldsymbol{u_2}]$ is a 2×3 orthogonal matrix formed by two nonzero eigenvectors of the projector, which is able to transform the original 3D vector $\boldsymbol{\psi}$ into 2D coordinates $\boldsymbol{\phi}$ within this plane. This transformation process is portrayed in (11).

$$\boldsymbol{\phi} = \boldsymbol{U\psi^T} = [\boldsymbol{u_1} \cdot \boldsymbol{\psi^T}; \boldsymbol{u_2} \cdot \boldsymbol{\psi^T}], \qquad (11)$$

with $\boldsymbol{u_1} = [\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0], \boldsymbol{u_2} = [\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}]$.

Along with the substitution of (10) in (11), we are able to derive the 2D coordinates of chromaticity image pixels analytically as follows:

$$\boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2}(d_1 + (-\frac{k_2}{\lambda_1} + \frac{k_2}{\lambda_2})/T) \\ \frac{\sqrt{6}}{6}(d_2 + (-\frac{k_2}{\lambda_1} - \frac{k_2}{\lambda_2} + \frac{2k_2}{\lambda_3})/T) \end{pmatrix} \qquad (12)$$

with $d_1 = \log(\frac{W_1}{W_2}), d_2 = \log(\frac{W_1 W_2}{W_3^2})$.

The property of linearity in the projected plane could be straightforwardly deduced via a further analysis of (12):

$$\phi_2 = \frac{\sqrt{3}}{3} \frac{\lambda_1(\lambda_2 - \lambda_3) + \lambda_2(\lambda_1 - \lambda_3)}{(\lambda_1 - \lambda_2)\lambda_3} \phi_1 + d \qquad (13)$$

where d is an offset term determined by $\{W_1, W_2, W_3\}$. Considering that W_i depends merely on object surface reflectance and remains constant for a given geometry even under varying lighting conditions, the points projected onto this plane should take the form of straight lines with the same slope. Moreover, points belonging to the same material should be located on the same line, where the length of each line shows the variation range of lighting with respect to this material. Accordingly, the distance between each pair of parallel lines reflects the difference between object surface properties behind them. The similar idea of color lines was also discussed in [31] by simply slicing the RGB histogram.

The above inference is evidenced and supported by illustrations in Fig. 4. First, Fig. 4b shows that all chromaticity image points fall onto the same plane, the normal vector of which, depicted with a fine blue line, is $u = 1/\sqrt{3}(1, 1, 1)$; then, we choose two sub-regions in the original image for the linearity study since the whole image contains excessive points for demonstration. Fig. 4c and Fig. 4d represent, respectively, the projected 2D chromaticity pixels in forehead and nose bridge rectangles, where two approximately parallel line-shaped clusters can be clearly observed. In particular, the chosen nose bridge area shows more lighting changes while there is only unchanged directional lighting in the forehead area for comparative analysis. Correspondingly, the straight line in Fig. 4c holds a smaller range than that in Fig. 4d.

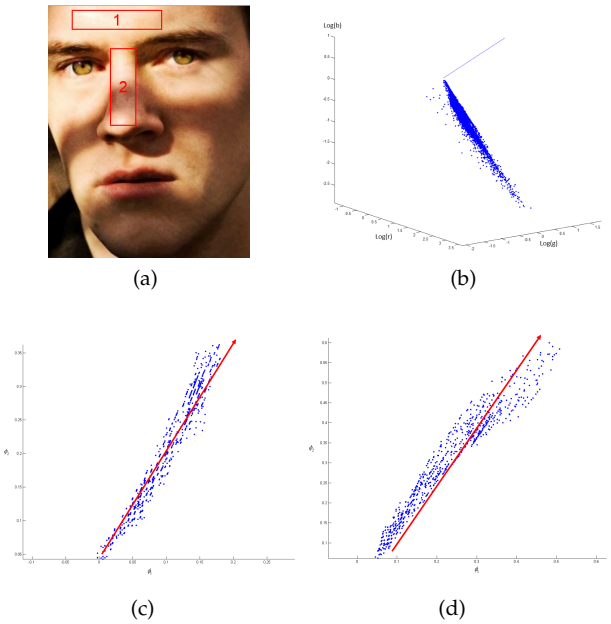


Fig. 4. Linearity of chromaticity image pixels in log space. (a) Original image. (b) chromaticity pixel values in 3D log space. (c) Pixels of the forehead area in the projected plane. (d) Pixels of the nose bridge area in the projected plane.

4.3 Entropy-based Lighting Normalization

Note that all 2D chromaticity image pixels are scattered into line-shaped clusters differentiated by their corresponding surface attributes. To estimate the intrinsic property of different materials in chromaticity images, we further proceed to reduce the dimensionality of the chromaticity space.

According to [32], global parsimony priors on reflectance could hold as a soft constraint. Under this assumption, only a small number of reflectances are expected in an object-specific image. We reasonably extend this assumption to our own work, which implies that lighting normalization substantially decreases the probability distribution of disorder in a human face image. Within this pipeline, we seek a projection direction, parametrized by angle θ, which should be exactly perpendicular to the direction of straight lines formed on the projected plane. Inasmuch as points of the same material across various illuminations fall on the same straight line, their 2D-1D projection onto a line with angle θ will result in an identical value which could be literally treated as an intrinsic value of this material. During this 2D-1D projection formulated in (14), the chromaticity image is finally transformed into a 1D gray-scale image.

$$\chi = \phi_1 \cos\theta + \phi_2 \sin\theta \tag{14}$$

With this in mind, the most appropriate projection direction could be found by minimizing the entropy of projected data. To begin with, we adopt the Freedman-Diaconis rule [33] for the purpose of determining the bin width as $h = 2\frac{Q(\chi)}{n^{1/3}}$, here n refers to the number of projected points. Compared with the commonly used Scott's rule, the Freedman-Diaconis rule replaces standard deviation of data by its interquartile range, denoted by $Q(\chi)$, which is thus more robust to outliers in data. Then, for each candidate projection direction, the corresponding Shannon entropy can be calculated based on the probability distribution of the projected points.

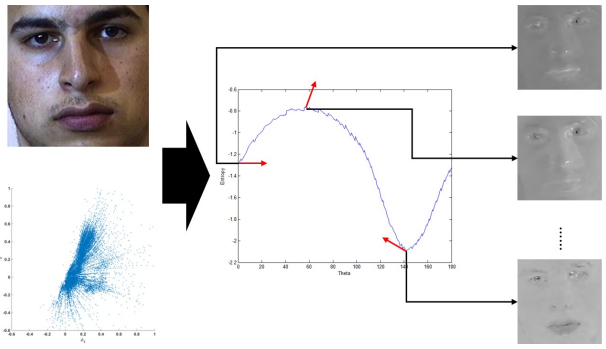


Fig. 5. Overview of chromaticity invariant image generation. Left column: original face image and its chromaticity points in 2D log space; middle column: entropy diagram as a function of projection angle, the arrows in red indicate projection directions at that point; right column: generated chromaticity images with different angle values.

Fig. 5 shows the workflow of chromaticity invariant image extraction in log space. Note that we choose three different angle samples, which are the zero point and two points leading to the minimum and maximum entropy, to visualize their generated chromaticity images. Apparently, only when the angle is adjusted to the value at which entropy is at its minimum is shadow effect significantly suppressed in its corresponding chromaticity image, *i.e.* the chromaticity invariant image.

Rather than traversing all possible θ ranging from 0 to π inefficiently, we conduct an additional analysis on the slope value of projected straight lines in (4), indicated by $k = \frac{\sqrt{3}}{3}\frac{\lambda_1(\lambda_2-\lambda_3)+\lambda_2(\lambda_1-\lambda_3)}{(\lambda_1-\lambda_2)\lambda_3}$. The theoretical value of slope is determined by trichromatic wavelengths $\{\lambda_1, \lambda_2, \lambda_3\}$, alternatively, the wavelengths of $\{R, G, B\}$ lights wherein $\{\lambda_1 \in [620, 750], \lambda_2 \in [495, 570], \lambda_3 \in [450, 495], unit : nm\}$. With simple calculations, it is interesting to note that no matter how these wavelengths change, k is always a positive value. The range of θ can therefore be restricted to $[\pi/2, \pi]$, which helps greatly reduce the computational burden.

4.4 Global Intensity Regularization

Notwithstanding illumination normalization, projected shadow-free images may suffer from global intensity differences across images caused by original lighting conditions and outliers. A final global regularization module is consequently integrated to overcome this drawback. In this step, the most dominant intensity of the resulting image is first approximated by a simple strategy:

$$\mu = (\text{mean}(\chi(x, y)^m))^{1/m} \tag{15}$$

where m is a regularization coefficient which considerably decreases the impact of large values. We take $m = 0.1$ by default following the setup in [15]. Next, this reference value is chosen to represent the color intensity of most face skin areas and is scaled to 0.5. The same scale ratio is then applied to all pixels to gain the final image.

5 Shadow-free Color Face Recovery

Though the representation of the 1D chromaticity invariant image contains successfully normalized lighting variations

across the whole face image, it is flawed due to the loss of textural details during the dimensionality reduction process, leading to low contrast images as depicted in Fig. 5. A full color image reconstruction module is therefore required both to improve the realism of generated images and enhance performance in face analysis.

5.1 In-depth Analysis of 1D Chromaticity Image

Given a chromaticity invariant image and all projection matrices, a general idea to reconstruct its color version is to project reversely its 1D lighting-normalized points to 2D/3D space in steps. However, this solution is virtually impracticable for two reasons: 1) recovery of overall intensity in each color band is an ill-posed problem since the shadow removal method is designed only for chromaticity values, 2) a large number of textural features, such as the mustache and the eyebrow, are undesirably eliminated or wrongly recognized as being skin during the forward 2D/1D projection. Thus, a further analysis of representation of RGB channels in log space is conducted.

Derived from equation (8), the logarithmic representation of RGB values, denoted by L_i, could be written as a two-component addition:

$$L_i = \log(\alpha I k_1 f_i \lambda_i^{-5} S(\lambda_i)) - \frac{k_2}{\lambda_i T}, \quad i = 1, 2, 3 \tag{16}$$

It is worth noting that the first additive component in the above equation consists of spatially varying factors, while the second additive term is lighting-dependent. Given an illumination-invariant region, the gradients at pixel (x,y) are then computed during inference:

$$\begin{aligned} \nabla_x L_i(x, y, T) &= \frac{L_i(x + \Delta x, y, T) - L_i(x, y, T)}{\Delta x} \\ \nabla_y L_i(x, y, T) &= \frac{L_i(x, y + \Delta y, T) - L_i(x, y, T)}{\Delta y} \end{aligned} \tag{17}$$

Based on evidence in [34] and [8], lighting conditions change slowly across a face image except for shadow edges. Consequently, for the partial derivative of the log-image with respect to x at any pixel (x, y) which appears out of shadow edges we obtain:

$$\nabla_x L_i(x, y, T_1) = \nabla_x L_i(x, y, T_2), \forall (T_1, T_2) \tag{18}$$

where T_1 and T_2 refer to different lighting conditions such as illuminated part and shadow part. This property holds equally for the partial derivative with respect to y.

To summarize, lighting conditions across a log-image are mainly changed on the boundary of the shadow area, *i.e.* for any pixel inside or outside this boundary, the spatial gradient is practically lighting-invariant. Motivated by this, we will derive a shadow-specific edge detection method analytically.

5.2 Shadow-Specific Edge Detection

The ability to separate shadow-specific edges from edges between different facial parts is crucial. To achieve this aim, we trace back the generation of the 1D chromaticity invariant image, where the shadow edges are removed by an orthogonal projection. Note that this projection was determined by an angle $\theta_{\min}$, which minimizes the entropy of (14). Conversely, a 'wrong' projection angle would retain or even highlight the shadow edge.

More specifically, we seek a novel direction $\theta_{\max}$ along which projection of chromaticity pixels to 1D tends to clearly preserve the chaos caused by varying lighting conditions. The $\theta_{\max}$ could be estimated by maximizing entropy. Theoretically, the freshly projected 1D image contains edges caused by both facial features and lighting variations. Thus, it would be considered to be different from the chromaticity invariant image in order to obtain the shadow-specific edge mask $M(x, y)$.

Furthermore, considering that lighting effects could be specially enhanced in one of the two dimensions described in (12), we define $M(x, y)$ as follows, while combining comparisons in both re-projected $\phi_1^{\min}, \phi_2^{\min}$ and $\phi_1^{\max}, \phi_2^{\max}$:

$$M(x, y) = \begin{cases} 1 & if\ \|\phi'_{\min}\| < \tau_1\ \&\ \|\phi'_{\max}\| > \tau_2 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

where $\|\phi'_{\min}\| = max(\|\nabla\phi_1^{\min}\|, \|\nabla\phi_2^{\min}\|)$, $\|\phi'_{\max}\| = max(\|\nabla\phi_1^{\max}\|, \|\nabla\phi_2^{\max}\|)$ and τ_1, τ_2 are two pre-defined thresholds.

It is worth mentioning that all 2D chromaticity images derived from both $\theta_{\max}$ and $\theta_{\min}$ are preprocessed by a guided filter [35] to facilitate gradient calculation on a smoother version. As regards the choice of guided filter, we use the all-ones matrix for the chromaticity invariant image to average the intensity. Conversely, the chromaticity image with shadows will take itself for guided filtering to enforce the gradient map.

5.3 Full Color Face Image Reconstruction

Inasmuch as the shadow edge mask is provided by the above detector, our focus can now be turned to the full color face image recovery. The algorithm simply continues the assumption that illumination variations mainly take place in the shadow edge area and could be ignored in other regions, *i.e.* the key to reconstructing an illumination-normalized color image is the reconstruction of a novel gradient map excluding the shadow-specific gradients.

To address this problem, we define a shadow-free gradient map $\zeta(x, y)$ for each log-RGB channel i as follows:

$$\zeta_{k,i}(x, y) = \begin{cases} \nabla_k L_i(x, y) & if\ M(x, y) = 0 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where $k \in \{x, y\}$. Apparently this novel shadow-free gradient map will lead us to a shadow-free Laplacian for each band:

$$\nu_i(x, y) = \nabla_x \zeta_{x,i}(x, y) + \nabla_y \zeta_{y,i}(x, y) \tag{21}$$

This straightforwardly computed Laplacian, when combined with the shadow-free log-image $\widehat{L}$ to be reconstructed, allows us to define Poisson's equation easily:

$$\nabla^2 \widehat{L}_i(x, y) = \nu_i(x, y) \tag{22}$$

Solving Poisson's equation is challenging. Two nontrivial priors are therefore imposed to make it soluble: first, the Neumann boundary condition is adopted which specifies the derivative values on the boundary. Here we uniformly set them to zero for convenience; second, instead of enforcing integrability of ν_i, we simply discretize relevant

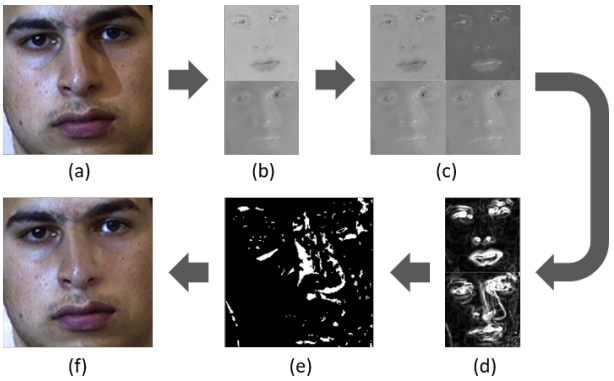


Fig. 6. Overview of edge mask detection and full color face recovery. (a) and (f) are raw and recovered face images; (b), (c) and (d) depict 1D/2D chromaticity images and edge maps, respectively. Note that, in each figure, the upper row refers to the shadow-free version, while the lower row is shadow-retained; (e) is the final detected edge mask.

terms and perform the calculation in matrix space. Importantly, given an image of size $M \times N$, the Laplacian operator ∇^2, which acts essentially as a 2D convolution filter $[0, 1, 0; 1, -4, 1; 0, 1, 0]$, is represented by a sparse matrix Λ of size $MN \times MN$.

Let

$$D = \begin{bmatrix} -4 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & -4 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -4 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -4 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 & -4 & 1 \\ 0 & \cdots & 0 & 0 & 0 & 1 & -4 \end{bmatrix} \tag{23}$$

and I denotes an $M \times M$ unit matrix. We have

$$\Lambda = \begin{bmatrix} D & I & 0 & 0 & 0 & \cdots & 0 \\ I & D & I & 0 & 0 & \cdots & 0 \\ 0 & I & D & I & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & I & D & I & 0 \\ 0 & \cdots & 0 & 0 & I & D & I \\ 0 & \cdots & 0 & 0 & 0 & I & D \end{bmatrix} \tag{24}$$

Each row of Λ corresponds to a sparse full-size filter for one pixel, $\widehat{L}_i$ could be accordingly solved by a left division:

$$\widehat{L}_i = \Lambda \setminus \nu_i \tag{25}$$

After exponentiating $\widehat{L}_i$, a multiplicative scale factor per channel, which is computed by retaining the intensity of the brightest pixels in the raw image, will be finally applied to ensure that not only color but also intensity is properly recovered. See Fig. 6 for a demonstration of shadow-specific edge detection and color face recovery results.

6 Experimental Results

The effectiveness of the proposed method is first qualitatively assessed (subsection 6.2), and then quantitatively evaluated for face recognition in (subsections 6.3 and 6.4), using face images of two benchmarks, *i.e.*,CMU-PIE and FRGC, for their illumination variations (subsection 6.1).

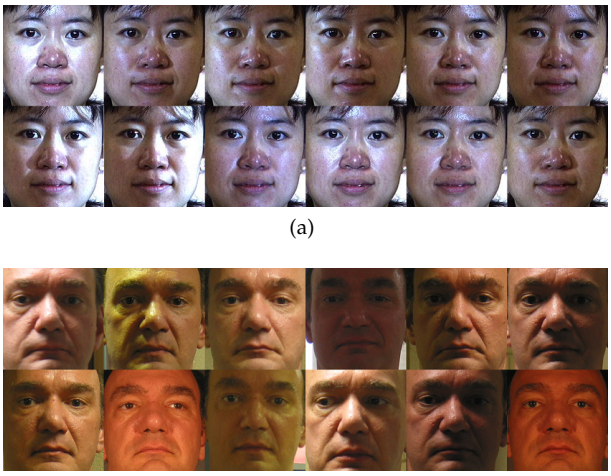


Fig. 7. Cropped face examples of the first subject in the (a): CMU-PIE database; (b): FRGC database.

TABLE 1
Overview of database division in our experiments

Database	Person	Target Set		Query Set	
		Lighting	Images	Lighting	Images
CMU-PIE	68	3	204	18	1,224
FRGC	466	controlled	16,028	uncontrolled	8,014

6.1 Databases and Experimental Settings

Databases. In light of the fact that our method aims to normalize and recover illumination in RGB color space, two criteria need to be fulfilled in selecting a database: 1) it includes face images taken with various lighting conditions; and 2) all images are provided with full color information. The two selected databases are CMU-PIE [36] and FRGC [37], and only lighting variations are considered.

Using the first subject of each database, Fig. 7 gives an illustration of some image samples across varying illumination environments. Note that all facial images are cropped and that resolution is 180×180. As can be visualized from these figures, the CMU-PIE database contains well-controlled illuminations and strictly unchanged pose for one subject, while the FRGC database contributes more to variations in illumination in combination with slight pose changes, thus bringing our evaluation closer to real-life application conditions.

Table 1 gives a detailed structure as well as an experimental protocol for each database. According to commonly used protocols, two different tasks are proposed for these two databases: 1-v-n face identification for CMU-PIE and 1-v-1 face verification for FRGC. These will be further detailed in the upcoming subsections.

Features. To evaluate performance robustness under different feature extraction algorithms, we have experimented with four popular descriptors in face recognition, including Local Binary Pattern (LBP), Local Phase Quantization (LPQ), Local Gabor Binary Pattern (LGBP), and deep CNN based face descriptor (VGG-Face). The parameter settings for each of them are detailed as follows:

- LBP [38]: For each image, a 59D uniform LBP his-

togram feature is extracted. We set the number of sample points as 8 and the radius as 2. Chi-square distance is computed between two LBP features to represent their dissimilarity.

- LPQ [39]: We set the size of the local uniform window as 5 and the correlation coefficient ρ as 0.9. Accordingly, the α for the short-time Fourier transform equals the reciprocal of window size, *i.e.* $\alpha = 0.2$. With the decorrelation process, the output feature is a 256D normalized histogram of LPQ codewords. Chi-square distance is applied as a matching criterion.
- LGBP [40]: For each image, 4 wavelet scales and 6 filter orientations are considered to generate 24 Gabor kernels. Similarly to LBP, holistic LGBP features are extracted for test images, resulting in 1.416D feature vectors. A simple histogram-intersection-matching described in [40] is used as a similarity measurement.
- VGG-Face [3]: The VGG-Face descriptors are computed based on the VGG-Very-Deep-16 CNN architecture in [3], which achieves state-of-the-art performance on all popular FR benchmarks. Here we simply take the pre-trained model and replace the last Softmax layer by an identity module to extract 4,096D features for test images.

Methods. The main contributions of our method are to remove shadows and to recover illumination-normalized color face images instead of de-lighting in gray-scale like all other existing methods do. To better present the effectiveness and necessity of the proposed method, we implement it as a preprocessing followed by other gray-scale level lighting normalization techniques to test fusion performance compared with the results obtained without using our method. As an exception to the above, for the VGG-Face model which requires RGB images as input, we conduct the comparison only between original images and shadow-free recovered images with no gray-scale level lighting normalization.

For this comparative study, a bunch of gray-scale space-based approaches are covered, including basic methods such as Gaussian filter based normalization (DOG), Gradient faces based normalization (GRF) [13], wavelet-based normalization (WA) [41], wavelet-based denoising (WD) [6], single-scale and multi-scale retinex algorithms (SSR and MSR) [42], [43], and state-of-art methods such as logarithmic discrete cosine transform (DCT) [7], single-scale and multi-scale self-quotient image (SQI and MSQ) [10], single-scale and multi-scale Weberfaces normalization (WEB and MSW) [14]. Additionally, a well-known fusing preprocessing chain (TT) [15] is also experimented. Thankfully, an off-the-shelf implementation provided by Štruc and Pavešić [44], namely INface Toolbox, grants us the opportunity to achieve our target efficiently and accurately.

6.2 Visual Comparison and Discussion

Shadows. First, a comparison of shadow removal results on soft and hard shadows is conducted and depicted in Fig. 8. We can derive two observations from these results:

1) From a holistic viewpoint, our method satisfactorily handles the removal of both hard and soft edge shadows. In both cases, the lighting across the whole image is normalized and the shadow effects are eliminated.
2) Specified in dashed-red and dashed-blue rectangles, respectively, the two middle image patches reveal the differences while processing different shadows. Despite visually similar results, for face images on the left with a hard-edged shadow, shadow removal performance is actually more robust than for the image on the right with soft shadows. This is because more facial details are smoothed for soft shadows, where shadow edges are difficult to define. This drawback may also affect the performance of face recognition, a fact that will be detailed in the next subsection.

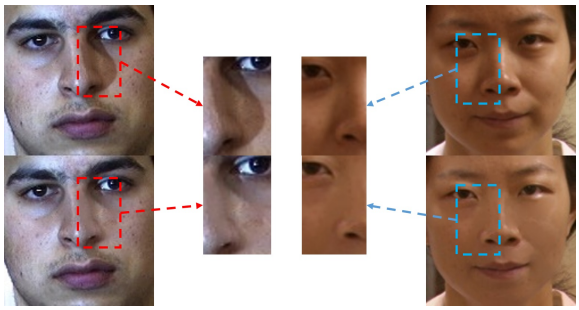
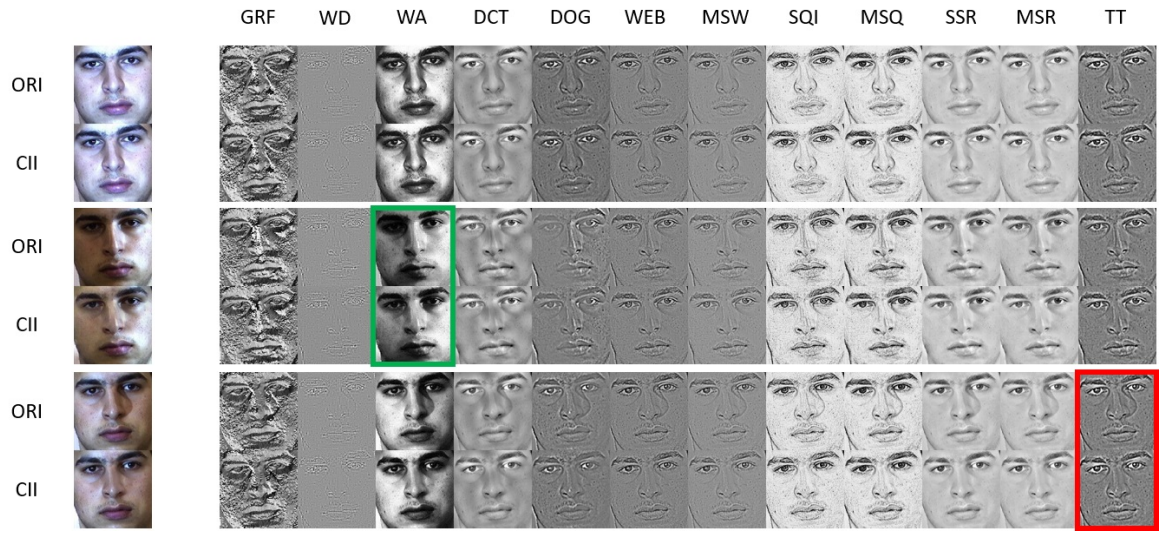


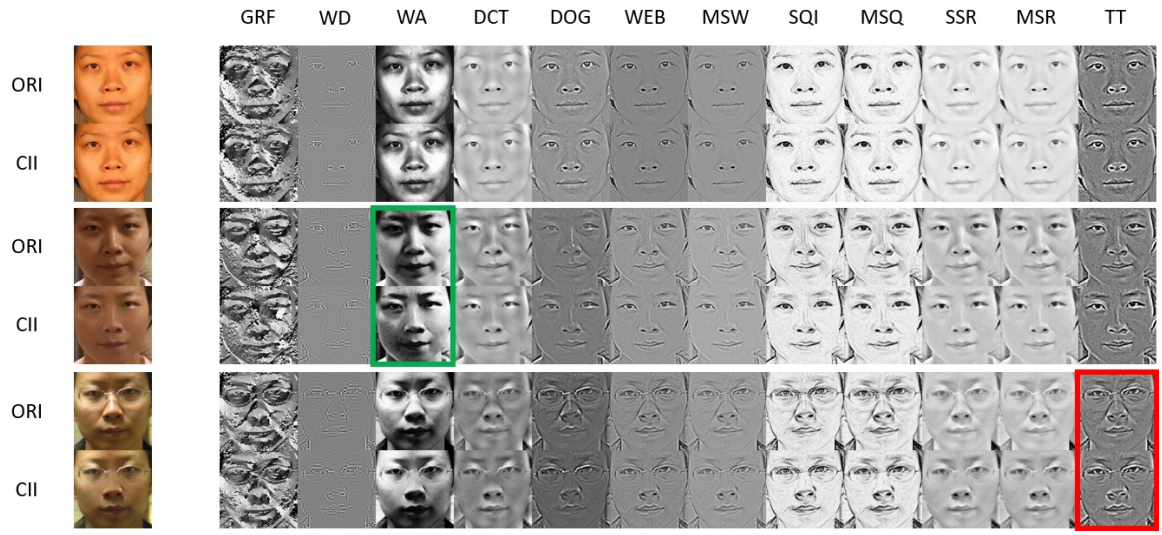
Fig. 8. Holistic and local shadow removal results on hard-edged shadows (left) and soft shadows (right).

Fusions. To visually evaluate the effectiveness of the proposed method before its quantitative evaluation, we consider some image samples selected from both databases as well as the corresponding results after different lighting normalization methods in Fig. 9. Three gradually varying illumination scenarios are considered in our illustration, including uniformly distributed frontal lighting, a side lighting causing soft shadows, and another side lighting causing some hard-edged shadows. This setting aims to evaluate the robustness of the proposed method against a wide variety of illumination environments. From the visual inspection of Fig. 9, it can be seen that:

1) In the first scenarios of both Figs. 9a and 9b, we observe hardly any differences between the original images and the recovered images. This is due to the homogeneous distribution of lighting, which tends to assign a zero value to most elements of the shadow-specific edge mask $M(x, y)$. In this case, our algorithm considers that very few changes are required to hold this homogeneous distribution.
2) The two middle rows in Fig. 9a depict a face with soft shadows mainly located on its left half. Before applying additional lighting normalization methods, the two leftmost images show that the recovered color image successfully normalizes holistic lighting intensity while retaining texture details. This property can also be evidenced by contrast after fusion with a diverse range of lighting normalization methods. Note that most of these techniques could handle perfectly the removal of soft shadows such as DCT, SQI, SSR and TT. For these techniques,



(a)



(b)

Fig. 9. Illustration of illumination normalization performance of two samples in (a) CMU-PIE and (b) FRGC database. For each sample, three lighting conditions are considered, *i.e.* from top to bottom, the image with frontal lighting, the image with soft shadows, and the image with hard-edged shadows.The columns represent different lighting normalization techniques to be fused with the original image or the CII recovered image.

visually indistinguishable results are obtained on both the original images and the recovered images. On the other hand, for techniques which are less robust to soft shadows such as WA (visualized in green boxes), taking the recovered image as input enables a globally normalized lighting intensity where dark regions, especially the area around eyes, are brightened. Compared with the original image, this process yields better visualization results. Unlike the first subject in CMU-PIE, we choose a female face from FRGC with complicated illumination conditions where shadows are more scattered. Even though certain shadows still remain around the mouth region with our method, we can nevertheless perceive the improvement of shadow suppression on the upper half of the face.

3) The two bottom rows in Fig. 9a and 9b focus on hard-edged shadows caused by occlusion by the nose and glasses against the lighting direction, respectively. In this scenario, the resulting images generated by adopting the proposed recovery method as preprocessing show distinct advantages over those generated from the original image. This kind of shadow edge is difficult to remove for existing lighting normalization methods, including the state-of-art algorithm TT (visualized in red boxes), since these methods can barely distinguish shadow edges from the intrinsic texture.

To summarize, according to the visual comparison results, our shadow-free color face recovery algorithm could (1) provide original images with intuitively identical results when illumination is homogeneously distributed everywhere; (2) normalize holistic lighting in color space when soft shadows occur; (3) be performed as a supplementary measure specifically to remove hard-edged shadows before being fused with other gray-scale level lighting processing.

Fig. 10. Faces in the wild before (top) and after (bottom) shadow removal. From left to right we choose images with a gradual decrease (left: strong, middle two: moderate, right: weak) in shadow intensity.

Faces in the wild. To further analyze the effectiveness and limitation of our approach, we conduct additional experiments on natural face images in the wild with a far wider range of lighting conditions. The first row of Fig. 10 illustrates four face images with a gradual decrease in shadow intensity. As can be seen on the bottom row images after shadow removal, our method can effectively handle faces under moderate lighting conditions (middle two images) quite well. However, it will fail when holistic lighting is poor with intense shadows (first image), or when holistic lighting is too bright with soft shadows (last image). In both cases, lighting conditions are saturated (pixel values are limited by either 0 or 255) and, accordingly, our assumption of linearity in chromaticity space becomes much weaker.

6.3 Identification Results on CMU-PIE

A rank-1 face identification task is generally described as a 1-to-n matching system, where n refers to the number of recordings in the target set. In this scenario, closed-set identification is performed on various recognition algorithms to evaluate the effectiveness of our method.

Table 2 tabulates the identification rate for different features. For each feature and each gray-scale lighting normalization method, we compare the results before and after taking the CII recovery algorithm as preprocessing. In particular, we adopt a state-of-the-art reflectance recovery algorithm SIRFS [32], which defined an optimization problem based on a series of priors on shape, reflectance and illumination, as another preprocessing method for comparative study. The highest accuracy is highlighted. Several observations can be derived from these results:

1) Generally, fusing our method in the preprocessing chain helps improve performance on this identification task with different gray-scale approaches and features. This is because our method emphasizes shadow edge removal while all other methods suffer from retaining such unwanted extrinsic features.
2) Without any gray-scale methods (N/A in the Table) or even with gray-scale methods such as WA, which are relatively less robust to lighting variations, our recovery method helps boost performance significantly. This observation implies that, besides the effect of shadow removal, our method also provides us with holistic lighting normalization.
3) For SQI and MSQ, our method causes slight yet unpleasant side effects with LBP and LPQ features. This is due to the phenomenon previously observed, *i.e.* that our method will smooth the detected shadow edges and thus that SQI/MSQ may become more sensitive to this unrealistic smoothness as images would be further divided by their smoothed version. Nevertheless, with LGBP features, we still achieves better results with SQI/MSQ as introduction of 24 Gabor filters helps alleviate the effect of the smoothed region.
4) Fusion of CII and TT failed to result in performance improvement. As a preprocessing sequence itself, TT has been carefully adjusted to the utmost extent, making it hard to combine with another method.
5) In compliance with state-of-the-art FR methods, the deep learning-based VGG-Face model largely outperforms the other features. Such a performance has been made possible for 2 major reasons: a) the adoption of deep CNNs; b) the availability of very large-scale labeled training datasets. VGG-Face [3] has been trained using a deep CNN of 18 layers and a dataset of 2.6M face images over 2.6K people. The requirement of large-scale training data highlights the limit of the current massively data-driven machine learning approach, *e.g.*,deep learning, which increasingly faces severe data starvation as disturbing factors, *e.g.*, lighting variations, are multiplied. In the case of the CMU-PIE where lighting is strictly controlled, the training data of 2.6M face images used by VGG-Face more or less cover these lighting variations, thereby enabling the VGG-Face features to achieve an accuracy rate as high as 99.7%. However, using the proposed CII recovery, we still witnessed a 0.3% improvement.
6) The reflectance from shading methods always leads to weaker performance, showing the limitation of using parsimony and smoothness as strong assumptions. Our method retains useful facial details by adopting such assumptions only for shadow detection, while SIRFS tends to average similar pixels and obtains reflectances which are difficult to recognize.

6.4 Verification Results on FRGC

Notwithstanding its one-to-one characteristic, face verification on FRGC v2.0 is always considered as a highly challenging task. This is because a large number of face images in FRGC are captured in uncontrolled and thus complicated illumination environments, with sensor or photon noise as well. For each preprocessing combination and each feature, we conduct a 16.028×8.014 pair matching and then compute the verification rate (VR) based on this similarity matrix. The experimental results are evaluated by Receiving Operator Characteristics (ROC), which represents the VR varying with the False Acceptance Rate (FAR).

Similarly, we list the performance of different methods on the ROC value for FAR at 0.1% in Table 3. Moreover, ROC curves for each gray-scale method are illustrated in Fig 11. We derive our observations from these results:

TABLE 2
Rank-1 Recognition Rates (Percent) of Different Methods on CMU-PIE Database

Feature	Preprocessing	Gray-Scale Lighting Normalization Methods												
		N/A	GRF	WD	WA	DCT	DOG	WEB	MSW	SQI	MSQ	SSR	MSR	TT
LBP	Original	44.0	32.8	20.8	39.2	72.6	65.4	59.2	58.9	**61.4**	**71.8**	66.8	67.7	**67.7**
	SIRFS [32]	41.8	32.2	20.0	37.6	70.0	63.2	55.8	55.6	58.4	69.9	66.6	67.2	64.2
	CII Recovery	**48.3**	**34.1**	**23.0**	**45.6**	**75.3**	**66.0**	**59.8**	**61.0**	61.3	70.9	**68.8**	**69.8**	65.6
LPQ	Original	58.0	34.7	**37.9**	49.6	85.2	79.1	73.9	**77.6**	**74.0**	**80.2**	84.4	84.8	**82.4**
	SIRFS [32]	56.4	32.8	35.7	51.0	84.8	77.4	74.1	73.9	72.0	76.3	82.3	83.7	80.0
	CII Recovery	**62.6**	**35.1**	35.5	**55.3**	**86.6**	**81.0**	**75.7**	75.1	73.1	77.1	**88.3**	**87.4**	82.3
LGBP	Original	75.5	67.8	84.6	67.2	97.4	91.3	99.0	99.4	99.5	99.2	98.0	97.8	**97.9**
	SIRFS [32]	74.2	66.6	83.9	68.6	97.0	89.7	99.1	99.2	98.8	98.0	97.4	97.6	94.7
	CII Recovery	**77.6**	**74.6**	**84.8**	**72.1**	**98.0**	**93.6**	**99.4**	**99.7**	**99.6**	**99.4**	**99.5**	**98.4**	96.7
VGG-Face	Original	99.7	-	-	-	-	-	-	-	-	-	-	-	-
	SIRFS [32]	99.1	-	-	-	-	-	-	-	-	-	-	-	-
	CII Recovery	**100**	-	-	-	-	-	-	-	-	-	-	-	-

TABLE 3
Verification Rate (Percent) at FAR = 0.1% Using Different Methods on FRGC V2.0 Exp.4

Feature	Preprocessing	Gray-Scale Lighting Normalization Methods												
		N/A	GRF	WD	WA	DCT	DOG	WEB	MSW	SQI	MSQ	SSR	MSR	TT
LBP	Original	1.0	12.8	3.5	1.1	6.0	14.5	18.5	17.7	18.5	12.3	3.8	3.9	15.7
	CII Recovery	**1.3**	**14.8**	**5.3**	**1.5**	**6.2**	**18.8**	**23.3**	**23.1**	**25.6**	**18.0**	**5.3**	**5.9**	**20.4**
LPQ	Original	1.4	14.2	7.4	2.0	6.6	**15.3**	18.3	18.8	13.4	12.0	6.2	7.5	**21.4**
	CII Recovery	**2.0**	**17.6**	**7.5**	**2.5**	**6.7**	14.9	**19.1**	**19.7**	**16.8**	**15.2**	**7.3**	**8.1**	20.2
LGBP	Original	13.0	31.0	18.8	12.7	28.2	37.0	37.9	35.7	29.1	30.9	27.2	28.2	38.8
	CII Recovery	**16.7**	**33.2**	**25.9**	**14.3**	**29.6**	**42.4**	**38.4**	**37.0**	**31.0**	**33.1**	**29.4**	**29.9**	**44.4**
VGG-Face	Original	92.5	-	-	-	-	-	-	-	-	-	-	-	-
	CII Recovery	**93.6**	-	-	-	-	-	-	-	-	-	-	-	-

1) Using the recovered color image is generally an effective way to improve performance on this verification task with different gray-scale methods and features. Compared with the identification task on CMU-PIE, this effectiveness is enhanced since CII helps improve the verification rate at FAR = 0.1% for almost all gray-scale methods with different features, thus validating the superiority of our method.
2) When dealing with the FRGC database where a large number of face images are captured in unconstrained conditions, thereby presenting far more complicated illumination variations, the VGG-Face feature sees its verification rate decrease to 92.5% at FAR=0.1%. This can be explained by the fact that its 2.6M training data fail to cover the whole spectrum of lighting variations depicted in FRGC. Once more, the proposed CII recovery shows its superiority in displaying a higher verification rate of 93.6% in comparison with the VGG-Face directly applied to the raw face images.
3) The performance variance for different gray-scale methods is not totally consistent with our previous observation on CMU-PIE. Unlike the results on CMU-PIE, GRF, DOG and WEB achieve better results than DCT and SSR, which suggests that these methods are more robust when dealing with more uncontrolled lighting conditions.

7 Conclusion

In this paper we have presented a novel pipeline in chromaticity space for improving performance on illumination-normalized face analysis. Our main contributions consist in: (1) introducing the concept of chromaticity space in FR as a remedy to illumination variations, (2) achieving an intrinsic face extraction processing and (3) achieving a photo-realistic full color face recovery after shadow removal. Overall, the proposed approach explores physical interpretations for skin color formation and has proven to be effective by improving performance for FR across illumination variations on different databases. Meanwhile, it shows promising potential in practical applications for its photo-realism and extensibility. Further efforts in developing this work will include synthesizing face images under different illumination conditions and combining other techniques in order to address face analysis problems in the wild.

Acknowledgments

This work was partially supported by the French Research Agency, Agence Nationale de Recherche (ANR) through the Jemime Project under Grant ANR-13-CORD-0004-02, the Biofence project under Grant ANR-13-INSE-0004-02, the National Nature Science Foundation of China under Grant 91746111, 61303121, the Partner University Fund (PUF) through the 4D Vision project, and Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University.

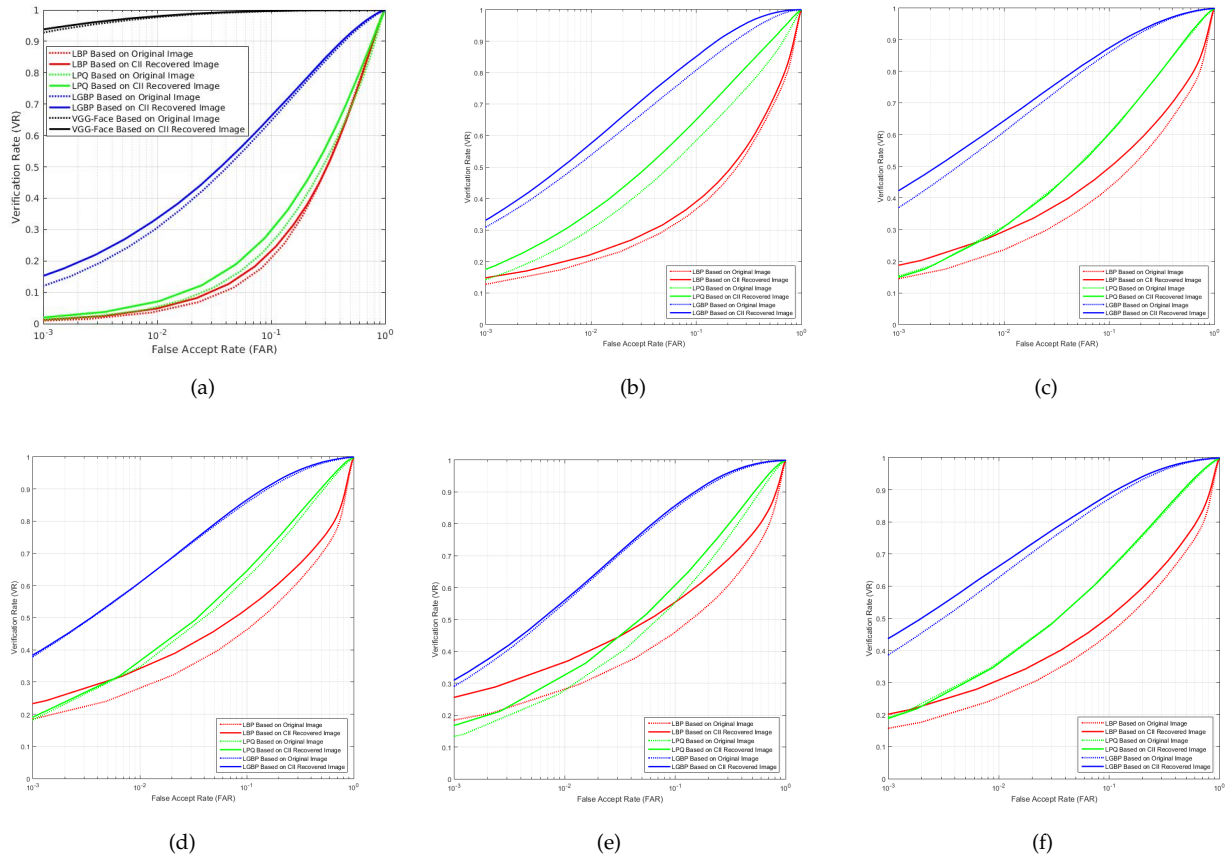


Fig. 11. Several ROC curves for different gray-scale methods. (a) No gray-scale method, (b) GRF, (c) DOG, (d) WEB, (e) SQI, (f) TT. Note that only (a) contains ROC curves for the VGG-Face model as it requires RGB images as model input.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[2] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 721–732, 1997.

[3] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[4] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[5] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," in *Analysis and Modeling of Faces and Gestures, IEEE International Workshop on*. IEEE, 2003, pp. 157–164.

[6] T. Zhang, B. Fang, Y. Yuan, Y. Yan Tang, Z. Shang, D. Li, and F. Lang, "Multiscale facial structure representation for face recognition under varying illumination," *Pattern Recognition*, vol. 42, no. 2, pp. 251–258, 2009.

[7] W. Chen, M. J. Er, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 36, no. 2, pp. 458–466, 2006.

[8] E. H. Land and J. J. McCann, "Lightness and retinex theory," *JOSA*, vol. 61, no. 1, pp. 1–11, 1971.

[9] T. Riklin-Raviv and A. Shashua, "The quotient image: class based recognition and synthesis under varying illumination conditions," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on.*, vol. 2. IEEE, 1999.

[10] H. Wang, S. Z. Li, and Y. Wang, "Generalized quotient image," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 2. IEEE, 2004, pp. II–498.

[11] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 9, pp. 1519–1524, 2006.

[12] X. Xie and K.-M. Lam, "An efficient illumination normalization method for face recognition," *Pattern Recognition Letters*, vol. 27, no. 6, pp. 609–617, 2006.

[13] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *Image Processing, IEEE Transactions on*, vol. 18, no. 11, pp. 2599–2606, 2009.

[14] B. Wang, W. Li, W. Yang, and Q. Liao, "Illumination normalization based on weber's law with application to face recognition," *Signal Processing Letters, IEEE*, vol. 18, no. 8, pp. 462–465, 2011.

[15] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

[16] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 2, pp. 218–233, 2003.

[17] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1063–1074, 2003.

[18] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3d face model for pose and illumination invariant face recognition," in *Advanced video and signal based surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*. IEEE, 2009, pp. 296–301.

[19] 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," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 11, pp. 1968–1984, 2009.

[20] X. Zhao, G. Evangelopoulos, D. Chu, S. Shah, and I. Kakadiaris, "Minimizing illumination differences for 3d to 2d face recognition

using lighting maps." *IEEE transactions on cybernetics*, vol. 44, no. 5, pp. 725–736, 2014.
[21] B. T. Phong, "Illumination for computer generated pictures," *Communications of the ACM*, vol. 18, no. 6, pp. 311–317, 1975.
[22] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions?" *International Journal of Computer Vision*, vol. 28, no. 3, pp. 245–260, 1998.
[23] L. Zhang, S. Wang, and D. Samaras, "Face synthesis and recognition from a single image under arbitrary unknown lighting using a spherical harmonic basis morphable model," in *Computer Vision and Pattern Recognition.*, vol. 2. IEEE, 2005, pp. 209–216.
[24] S. C. Kee, K. M. Lee, and S. U. Lee, "Illumination invariant face recognition using photometric stereo," *IEICE TRANSACTIONS on Information and Systems*, vol. 83, no. 7, pp. 1466–1474, 2000.
[25] A. Madooei and M. S. Drew, "Detecting specular highlights in dermatological images," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4357–4360.
[26] Z. L. Stan and K. J. Anil, "Handbook of face recognition," *Springer*, 2005.
[27] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *The Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.
[28] G. Wyszecki and W. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, ser. Wiley Series in Pure and Applied Optics. Wiley, 2000.
[29] G. D. Finlayson, M. S. Drew, and C. Lu, "Entropy minimization for shadow removal," *International Journal of Computer Vision*, vol. 85, no. 1, pp. 35–57, 2009.
[30] D. I. MacLeod and R. M. Boynton, "Chromaticity diagram showing cone excitation by stimuli of equal luminance," *JOSA*, vol. 69, no. 8, pp. 1183–1186, 1979.
[31] I. Omer and M. Werman, "Color lines: image specific color representation," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2004, pp. II–946–II–953 Vol.2.
[32] J. T. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 8, pp. 1670–1687, 2015.
[33] D. Freedman and P. Diaconis, "On the histogram as a density estimator: L 2 theory," *Probability theory and related fields*, vol. 57, no. 4, pp. 453–476, 1981.
[34] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 1, pp. 59–68, 2006.
[35] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 1–14.
[36] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1615–1618, 2003.
[37] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, vol. 1. IEEE, 2005, pp. 947–954.
[38] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.
[39] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.
[40] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1. IEEE, 2005, pp. 786–791.
[41] S. Du and R. Ward, "Wavelet-based illumination normalization for face recognition," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 2. IEEE, 2005, pp. II–954.
[42] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *Image Processing, IEEE Transactions on*, vol. 6, no. 3, pp. 451–462, 1997.
[43] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *Image Processing, IEEE Transactions on*, vol. 6, no. 7, pp. 965–976, 1997.
[44] V. Štruc and N. Pavešic, "Photometric normalization techniques for illumination invariance," *Advances in Face Image Analysis: Techniques and Technologies, IGI Global*, pp. 279–300, 2011.

Wuming Zhang was awarded his M.S. degree in computer science from Beihang University, Beijing, China, in 2013 and his Ph.D. degree in computer vision from the Laboratoire dInfoRmatique en Image et Systeme dinformation (CNRS UMR 5205), Ecole Centrale de Lyon, University of Lyon, Ecully, France, in 2017. His current research interests include 2D/3D face recognition, face detection, emotion recognition, image processing and deep learning.

Xi Zhao was awarded his Ph.D. degree in computer science from the Ecole Centrale de Lyon, Ecully, France, in 2010. He conducted research in the fields of biometrics and pattern recognition as a Research Assistant Professor with the Department of Computer Science, University of Houston, USA. He is currently an Associate Professor with the School of Management, Xian Jiaotong University, Xian, China. His current research interests include biometrics, social computing and mobile computing. Dr. Zhao serves as a reviewer of the IEEE Transactions on Image Processing, the IEEE Transactions on Cybernetics, etc.

Jean-Marie Morvan was awarded his Ph.D. from the University Paul Sabatier, Toulouse, France. He is a Professor of mathematics with the University Claude Bernard Lyon 1, Lyon, France, and a Visiting Professor at the King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. His research interests include differential geometry, in particular Riemannian and symplectic geometry, geometric measure theory, and application of geometry to fields such as geology and geophysics.

Liming Chen was awarded his Ph.D. degrees in computer science from the University of Paris 6, Paris, France, in 1989. He is a full Professor since 1998 at Ecole Centrale de Lyon, University of Lyon where he leads an advanced research group on computer vision and machine learning. Liming has over 250 publications and successfully supervised over 35 PhD students. He has been a grant holder for a number of research grants from EU FP program, French research funding bodies and local government departments. Liming has so far guest-edited 3 journal special issues. He is an associate editor for Eurasip Journal on Image and Video Processing and a senior IEEE member. His current research interests include machine learning, image and video analysis and categorization, face analysis and recognition, and affective computing.