Analysis of Noteworthy Issues in Illumination Processing for Face Recognition

Min YAO†a), Nonmember and Hiroshi NAGAHASHI††, Member

SUMMARY Face recognition under variable illumination conditions is a challenging task. Numbers of approaches have been developed for solving the illumination problem. In this paper, we summarize and analyze some noteworthy issues in illumination processing for face recognition by reviewing various representative approaches. These issues include a principle that associates various approaches with a commonly used reflectance model and the shared considerations like contribution of basic processing methods, processing domain, feature scale, and a common problem. We also address a more essential question—what to actually normalize. Through the discussion on these issues, we also provide suggestions on potential directions for future research. In addition, we conduct evaluation experiments on 1) contribution of fundamental illumination correction to illumination insensitive face recognition and 2) comparative performance of various approaches. Experimental results show that the approaches with fundamental illumination correction methods are more insensitive to extreme illumination than without them. Tan and Triggs’ method (TT) using $L_1$ norm achieves the best results among nine tested approaches.

key words: face recognition, illumination processing, noteworthy issues, analysis, evaluation

1. Introduction

For the past few decades, face recognition has been attracting much attention from both the scientific and business communities [1]–[4]. However, recognition performance will be largely degraded if faces under consideration are posed in severe illumination. To solve the illumination problem in face recognition, numerous approaches have been proposed. These approaches can be generally classified into two categories [4]: active approaches and passive approaches. Active approaches usually overcome the illumination problem by applying specific imaging techniques in order to capture face modalities (still face images or facial features) which are insensitive to illumination variances. Imaging techniques which capture 3D facial information [3], [5]–[7] and Infrared (IR) imagery including Thermal-IR [8]–[11] and Near-IR [12]–[14] belong to this category. By contrast, passive approaches deal with face images already contaminated by illumination. They try to 1) model the illumination or 2) correct the illumination in a basic way (e.g., HE [15]) or 3) sophisticatedly extract illumination insensitive features from the given images [16]–[26]. Surveys covering both of the categories can be found in [4], [27].

As can be observed from the available literature, most of the surveys on the illumination insensitive face recognition only gave an introduction to the methodological classification with an overview of existing algorithms [4], [27], or evaluations of several methods based on the recognition performance [28]–[30]. However, among the numerous approaches for processing the face illumination, some are based on the same model, some share common fundamentals, and some possess special trains of thinking. Moreover, there also exist unsolved problems in most of these approaches. Analysis of these issues seems both necessary and promising, which is also the purpose of this paper.

In this paper, we summarize and analyze several noteworthy issues in illumination processing of face images via reviewing various representative approaches. Specifically, we discuss the principle of the reflectance model and its implementations in various approaches. Then we consider the contributions of basic processing methods to the illumination normalization, assess the feature description domains and feature scales, and describe an unsettled common problem. We also discuss an essential question of what we should actually normalize. The active approach and illumination modeling are not included, because the former one is less pertaining to algorithms and the latter is usually computationally expensive or requiring much prior knowledge. The approaches discussed in this paper mainly belong to the illumination correction and illumination insensitive feature extraction. We focus on the viewpoints of interests for illumination processing rather than the specific approaches.

In the rest of this paper, we start with a brief introduction of the commonly used reflectance model in Sect. 2. Section 3 gives the analysis of how various approaches correspond to the reflectance model. Discussions on different considerations are shown in Sect. 4. An essential question is analyzed in Sect. 5. In Sect. 6, some evaluation experiments are presented. Sect. 7 concludes this paper.

2. Reflectance Model

The reflectance model is a widely used concept and foundation in illumination processing for face recognition. It expresses a given face image $I(x, y)$ under some illumination condition as a product [31]:

$$I(x, y) = R(x, y)L(x, y),$$

(1)
where \( R \) is the reflectance component and \( L \) is the luminance component at point \((x, y)\). Usually, the reflectance component \( R \) only depends on the intrinsic characteristics of the face; \( R \) is invariant to illumination. Therefore, the goal is to maintain the features of \( R \) and get rid of the effect of \( L \).

Figure 1 displays an example of attempting to separate the component \( R \) and \( L \) from a given face image. In most research works, the component \( L \) is commonly assumed to vary slowly in spatial level. This assumption is also verified by the human face model.

When it comes to the computer graphic application, a human face can be modeled as a combination of a series of small and flat facets [32, 33] such as CANDIDE-3 face model [34] which is composed of a sequence of triangular facets. The area of each facet \( W \) is small enough to be considered as a planar patch and thus the illumination is approximately fixed within the facet. Then the reflectance model can be also expressed as:

\[
I(x, y) = A \cdot R(x, y), \quad (x, y) \in W_{N \times N}, \tag{2}
\]

where \( A \) is a constant which denotes the illumination within squared small area \( W \), and \( N \) is the dimension of the area.

3. Correspondence to Reflectance Model

Except for the illumination correction methods such as histogram equalization, many approaches\(^1\) that cope with the illumination in face recognition are based on the reflectance model. Thus, it seems interesting to give an overview of representative approaches related to this model. Explanations of how they comply with this model are also included.

3.1 Estimation of Luminance (\( L \))

As aforementioned, the goal of illumination processing in face recognition is to maintain the intrinsic facial properties represented by \( R \) and remove the illumination effect \( L \). As seen from the reflectance model, the illumination problem can be solved by first estimating \( L \) and then calculating the de-illuminated version \( I' \) of the input face image by

\[
I'(x, y) = R(x, y) = \frac{I(x, y)}{L(x, y)}. \tag{3}
\]

Sometimes the logarithmic form of the image is taken. In this case, \( I' \) can be obtained by first estimating the \( L \) or \( \ln(L) \) and then computing

\[
I'(x, y) = R(x, y) = \exp(f(x, y) - u(x, y)), \tag{4}
\]

where \( f = \ln(I) \), \( u = \ln(L) \), or

\[
I'(x, y) = v(x, y) = f(x, y) - u(x, y), \tag{5}
\]

where \( v = \ln(R) \).

Self quotient image (SQI). Based on the assumption that \( L \) varies little spatially, SQI [16], [28] employs a low pass filter \( G \) to estimate the low frequency component \( L \) by

\[
L(x, y) = G(x, y) \ast I(x, y), \tag{6}
\]

and uses Eq. (3) for the final measure of \( I' \).

Adaptive single scale retinex (ASR). ASR, proposed in [17], estimates \( L \) by iteratively convolving the input image with a \( 3 \times 3 \) weighted mask as

\[
L^{(n+1)}(x, y) = \max(L^{(n+1)}(x, y), L^{(0)}(x, y)), \quad 0 < t \leq T \tag{7}
\]

with

\[
L^{(0)}(x, y) = \frac{1}{N(x, y)} \sum_{(x, y) \in W_{3 \times 3}} L^{(0)}(x, y)u(x, y), \tag{8}
\]

\[
N(x, y) = \sum_{(x, y) \in W_{3 \times 3}} w(x, y), \tag{9}
\]

where \( W_{3 \times 3} \) denotes the local square mask, \( u(x, y) \) is the weight coefficient, and \( T \) is the total number of iterations. The coefficient is calculated via combining two measures of the illumination discontinuity at each pixel. This aims to preserve the shadow boundaries in \( L \). Finally, the measure of \( I' \) is obtained through Eq. (5).

Logarithmic total variation (LTV). In [18], the illumination effect and facial features were distinguished as large scale features and small scale features, respectively. Thus, the authors first operated decomposition in the logarithmic image using TV-L\(^1\) model [35] and used the decomposed large scale features as the estimation of \( \ln(L) \):

\[
u = \text{argmin}_u \int |\nabla u| + \lambda \|f - u\|_{L1}, \tag{10}
\]

where \( \int |\nabla u| \) is the total variation of \( u \), \( \lambda \) is a scalar threshold on the image scale, \( f = \ln(I) \), \( u = \ln(L) \). At last, the measure of \( I' \) can be calculated by Eq. (4). In the final result, the large scale features are excluded.

Discrete cosine transform-based method (DCT). DCT [19] is another classic method that decomposes the input face image into multiple scales (i.e., frequencies). Also based on the assumption that \( L \) is of low frequency, the authors defined the first \( n \) low frequency components as \( L \) estimation. They subtracted these components from the discrete cosine transformed image of the logarithmic form of the input image. To sum up, first, represent the logarithmic image \( f = \ln(I) \) by \( M \times N \) two dimensional cosine transformation:
\[ f(x, y) = \sum_{s=0}^{M-1} \sum_{t=0}^{N-1} D(s, t), \]  
\[ u(x, y) = \sum_{i=1}^{n} D(s_i, t_i), \]

where

\[ D(s, t) = \alpha(s)\alpha(t)C(s, t)\cos\left(\frac{\pi(2x+1)s}{2M}\right)\cos\left(\frac{\pi(2y+1)t}{2N}\right). \]

With the \( f(x, y) \) and \( u(x, y) \) described in Eqs. (9) and (10), Eq. (5) is eventually used to achieve the measure of \( I' \).

**Wavelet denoising-based method (WD).** WD is applied in [20] to generate a nonlinear estimation of \( \ln(L) \) using the wavelet coefficients of the logarithmic image. With the subbands of the wavelet transform labeled as \( HH_k, HL_k, LH_k, \) and \( LL_k \) (\( k \) denotes the decomposition level), the first three represent the fine details of the input image and the last one is the low resolution residual. By manipulating soft thresholding to the coefficients of \( HH_k, HL_k, LH_k, \) while keeping the coefficients of \( LL_k \) unaltered, \( u = \ln(L) \) is estimated:

\[ u(x, y) = IW\{th(T(Wf), Wf_{LL}), \]  
\[ \]  

where \( i \in \{HH_k, HL_k, LH_k\} \), \( Wf \) denotes the wavelet coefficient of logarithmic image \( f = \ln(I) \) and its subscript denotes the corresponding subband, \( th(T) \) is the thresholding function which is executed when \( Wf \) is the coefficient of the detail subbands, and \( IW(\cdot) \) is the inverse transform. Then measure of \( I' \) can be obtained via Eq. (5), where, \( LL_k \) components are removed.

**Non-local means normalization (NLM).** NLM [42] first transfers the given image \( I \) to the logarithmic form \( f = \ln(I) \) and then uses the non-local means algorithm to construct a smoothed image based on a weighted sum of locally similar patches comprising the whole image. The smoothed image is used as the estimation of \( u = \ln(L) \),

\[ u(x) = \sum_{z \in f(x)} w(z, x)f(x), \]

where \( x \), \( y \) stands for an arbitrary pixel location \((x, y)\). \( W(z, x) \) represents the weighting function that measures the similarity between the local neighborhoods at the pixel location \( z \) and \( x \). Then Eq. (4) is used to obtain \( I' \).

3.2 Representation Only Related to Reflectance (R)

Because of the presence of the common assumption that \( L \) spatially varies slowly, estimating \( L \) is a widely adopted step. However, finding an estimation of \( L \) is only a middle step of the whole procedure and it is very difficult to accurately estimate \( L \) just based on the assumption, especially for images under extreme illumination. It is unavoidable to incur errors in the \( L \) estimation, as discussed in [21].

Actually, the approaches that result in face representation only related to \( R \) can be regarded as illumination insensitive. They omit the step of estimating \( L \) and directly compute the \( R \)-related representation. In the following analysis, we will introduce several approaches of this kind and give the deduction about how they are only related to \( R \).

**Gradient face (GRF).** In [21], it was argued that direct face representation only related to illumination insensitive component \( R \) is more robust than measuring \( R \) by first estimating \( L \). Accordingly, they produced an illumination insensitive measure called gradient face as:

\[ GRF = \arctan\left(\frac{L_y}{L_x}\right), \quad GRF \in [0, 2\pi), \]

where \( L_x \) and \( L_y \) are the differentiations of image \( I \) along the \( x \) and \( y \) directions, respectively.

Proof: Based on the reflectance model, consider any pixel at \((x, y)\) and its neighbor at \((x + \Delta x, y)\):

\[ I(x + \Delta x, y) = R(x + \Delta x, y)L(x + \Delta x, y). \]

Subtract Eq. (1) from (14) and refer to the assumption that \( L \) is approximately smooth:

\[ I(x + \Delta x, y) - I(x, y) \approx R(x + \Delta x, y)L(x, y) - R(x, y)L(x, y) \approx (R(x + \Delta x, y) - R(x, y))L(x, y). \]

Transforming Eq. (15) into derivative form, we gain \( I_x \approx L(x, y)R_x \) and similarly \( I_y \approx L(x, y)R_y \). By dividing these two equations, one can obviously notice that

\[ GRF = \arctan\left(\frac{L_y}{L_x}\right) \approx \arctan\left(\frac{R_y}{R_x}\right), \]

where \( GRF \) is only related to \( R \).

**Weber face (WF).** In [22], based on the Weber’s law, another illumination insensitive measure only related to \( R \) was developed with the name of Weber face:

\[ WF = \arctan\left(\arctan\left(\frac{1}{\Delta x \Delta y} \sum_{\Delta x=-\Delta y}^{\Delta x=\Delta y} I(x, y) - I(x+\Delta x, y+\Delta y)\right), \]  
\[ \]  

where \( \Delta x \) is a parameter for adjusting the contrast within the neighborhood.

Proof: According to the reflectance model, we have

\[ I(x + \Delta x, y + \Delta y) = R(x + \Delta x, y + \Delta y)L(x + \Delta x, y + \Delta y). \]

Since \( L \) has slow spatial changes, it can be rewritten as

\[ I(x + \Delta x, y + \Delta y) \approx R(x + \Delta x, y + \Delta y)L(x, y). \]

Replace each of \( I(x, y) \) and \( I(x + \Delta x, y + \Delta y) \) in Eq. (16) by Eqs. (1) and (18):

\[ WF \approx \arctan\left(\arctan\left(\frac{1}{\Delta x \Delta y} \sum_{\Delta x=-\Delta y}^{\Delta x=\Delta y} \frac{R(x, y) - R(x+\Delta x, y+\Delta y)}{R(x, y)}\right), \]  
\[ \]  

which also depends on only \( R \).
**Local normalization (LN).** The underlying idea of LN [23] is that the finally processed image should be of local zero mean and with unit variance within a small area \(W\). The output representation is given as

\[
LN = \frac{I(x,y) - E(I(x,y))}{\text{Var}(I(x,y))}, \quad (x,y) \in W_{N \times N},
\]

where \(E(\cdot)\) and \(\text{Var}(\cdot)\) are, respectively, the mean and variance operation in the local small area \(W\) with dimension denoted by \(N\).

Proof: According to Eq. (2), Eq. (19) can be restated as

\[
LN = \frac{A \cdot R(x,y) - E(A \cdot R(x,y))}{\text{Var}(A \cdot R(x,y))}
\]

\[
= \frac{A \cdot R(x,y) - A \cdot E(R(x,y))}{A \cdot \text{Var}(R(x,y))} = \frac{R(x,y) - E(R(x,y))}{\text{Var}(R(x,y))}.
\]

We can also see that LN is merely related to \(R\).

### 3.3 Other Processing

Instead of estimating the luminance \(L\) in the first stage or building the representation related to \(R\), some other approaches aim to enhance the component \(R\) or suppress the effect of \(L\) in their own ways. They also share the assumption that \(L\) is probably of low frequency and \(R\) is relatively of high frequency.

**Homomorphic filtering-based method (HF).** HF was designed in [24] in order to reduce the contribution of \(L\). It applies a high pass filter to the logarithmic image \(f = \ln(I)\) for the purpose of excluding component \(L\) in the filtered output. The resultant representation can be expressed by

\[
I'(x,y) = \exp \left( \mathcal{F}^{-1} \{\mathcal{F}(f) \cdot H\} \right),
\]

where \(\mathcal{F}(\cdot)\) means Fourier transform and \(H\) denotes a Fourier transformed version of a high pass filter such as Butterworth filter [36] in [24].

**Wavelet-based method (WA).** WA [25] also relies on the wavelet transform and processes the coefficients in a specific way as WD does. It increases the coefficients corresponding to the image details, i.e., \(HH, HL, \text{and} LH\) subbands, and does histogram equalization with the \(LL\) subband. These operations imply the accentuation of \(R\) and suppression of \(L\), which are also considered capable for illumination insensitive representation. The processing can be shown by

\[
I'(x,y) = IW(\mu \cdot (WI), \text{HE}(WI_{LL})),
\]

where \(i \in \{HH, HL, LH\}\), \(WI\) denotes the wavelet coefficient and its subscript denotes the corresponding subband, and \(\mu\) is a scalar factor (>1) with which the multiplication is applied when \(WI\) is the coefficient of the detail subbands. \(\text{HE}(\cdot)\) denotes the operation of histogram equalization and \(IW(\cdot)\) denotes the inverse transform with the information of wavelet coefficients.

**Tan and Triggs’ method (TT).** TT [50] is a three-step chain which orderly consists of Gamma Correction, DoG (Difference of Gaussian) Filtering and Contrast Equalization. The first and third steps are used to do local and overall enhancement, respectively. DoG achieves bandpass behaviour to remove shading effects. The inner (smaller) Gaussian is typically set quite narrow in order to maintain facial details that correspond to the component \(R\). The outer Gaussian is certainly wider, depending on the spatial frequency at which the information of low frequency becomes residual rather than informative; this is to suppress \(L\) to some extent.

### 3.4 Summary

From the above discussion of various approaches, it is more clear that the key goal of the illumination processing in face recognition is to maintain the intrinsic-feature-dependant component \(R\) and attenuate the illumination component \(L\). It can be easily drawn that \(L\) is commonly associated to the following terms:

- Low resolution, Slow spatial changes,
- Low frequency, Large scale.

By contrast, \(R\) is associated to

- High resolution, Rapid spatial changes,
- High frequency, Small scale.

Under these considerations, \(R\) may be approximated and \(L\) may be suppressed by means of filtering, denoising, decomposing, etc. These techniques can be further promoted via multi-scale analysis, addition of variable weights, or integration with other approaches to make it more sophisticated. This is a potential direction for future research.

One more notable thing is that the idea of local operation is shared by many approaches. Besides ASR, GRF, WF, LN, and NLM mentioned above, the famous LBP [37], [38], its improved versions [39], [40], the recently proposed LRM [41], and OLHE [43] also adopt the local operation. They all describe the relations between neighboring pixels for illumination insensitive representation. There may be two explanations for the wide usage of the local operation: a) illumination has different effects in different regions, and b) component \(L\) varies slowly within local small region. Therefore, one can take advantage of the local operation which tends to be robust to the change of \(L\) and achieve reliable face representation.

### 4. Considerations for Illumination Processing

In this section, apart from the reflectance model, we analyze some other aspects from the viewpoints of contribution of fundamental illumination correction, processing domain, feature scale, and a common problem.

#### 4.1 Fundamental Illumination Correction

This subsection presents four widely used illumination cor-
rection methods. They were proposed for denoising or image enhancement. They can correct the illumination to some extent and are taken as one part or preprocessing step in age enhancement. They can correct the illumination to some . It can be deemed as a preprocessing step, which would facilitate the subsequent illumination normalization. Under this transformation, the reflectance model can be rewritten as

\[
\ln(I(x, y)) = \ln(R(x, y)) + \ln(L(x, y)).
\]  
\[(22)\]

It changes the multiplicative form of the reflectance model to additive one, which is likely to remove noise from the image and make it possible to promote useful signals.

We have already given examples of approaches using this transformation in Sect. 2. Besides, more approaches making use of LT can be referred to the recently proposed methods using neighboring wavelet coefficients [45], non-subsampled contourlet transform [46], total variation for preprocessed decomposition [47], [48], etc. As seen from Eq. (5), the final representation sometimes skips the inverse transformation from the logarithmic form, which actually justifies the significance of LT.

Histogram equalization (HE). HE [15] maps the histogram of an input image to a new uniform distribution, so that the intensity probabilities are equal and the image is fairly enhanced. Consider an input image \(I(x, y)\) with \(n\) pixels and a total number of \(L\) grey levels. The probability of the occurrence of one pixel with \(r_q\) grey level is denoted by \(p_r(r_q) = n_q/n\), where \(n_q (q = 0, 1, \cdots, L - 1)\) is the number of pixels with \(r_q\) grey level. HE transforms a given intensity value \(r_q\) to a new one \(r_q'\) by

\[
r_q' = \sum_{i=0}^{q} n_i/n = \sum_{i=0}^{q} p_r(r_i) = \text{CDF}(r_q),
\]  
\[(23)\]

where CDF(\(\cdot\)) represents the cumulative distribution function. This equation redistributes the intensity to the domain of [0, 1]. The target values \(r_q'\) need to be rescaled to obtain pixel values in the original domain, e.g., [0, 255].

In the implementation of WA [25], HE is applied to make the available dynamic range of the original image be sufficiently used. LN [23] combines HE with the proposed method to get better result. OLHE [43] also takes HE as a useful instrument by applying oriented HE locally.

Gamma correction (GC). GC nonlinearly transforms the original gray-level image \(I\) to \(I'\), where \(\gamma \in [0, 1]\) is a user-defined parameter. By setting different values of \(\gamma\), it can accordingly enlarge the local dynamic intensity range in dark regions while compressing that in bright regions, which meets the requirement of human visual conception under common illumination conditions (not pitch black or blindingly bright) [49]. This capability in illumination balance of GC allows for its application to illumination processing in face recognition. It is applied in Tan and Triggs’ method (TT) [50] as one of the three steps of the proposed pipeline. In [51], [52], GC was used for illumination correction by solving an optimization problem.

Gaussian filtering (GF). Gaussian kernel function with standard deviation \(\sigma\) is defined as

\[
G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right).
\]  
\[(24)\]

It is also frequently used as a preprocessing technique for the purpose of excluding noises and specific illumination from an input image. Especially, the techniques which extract gradient-related features prefer GF for computational ease with derivative of Gaussian kernel function such as GRF [21]. In [50], Difference of Gaussian (DoG) was used to discard high spatial frequencies that were present in the given face image such as shading effects. A 2D Gaussian illumination model was also used in [53] to compensate the illumination variances caused by shadows.

4.2 Processing Domain

It is important to articulate how the face recognition is affected by taking into account the processing domain. Some classic face recognition algorithms are only implemented within the spatial domain, which ignore the information concealed in the relationship between neighboring pixels. More sophisticated facial feature representation methods considering frequency information had been proposed later such as Garbor feature based method [54] and Wavelet feature based method [55]. The features extracted from the frequency domain seems more discriminant than those from the spatial domain.

It can be found that some illumination processing approaches explicitly transform an original image into frequency domain (e.g., DCT, WD, WA, and NWC [45]). Thus we also suppose that frequency domain provides more useful information which can aid to distinguish between facial features and illumination. This is easy to understand if one recalls the reflectance model and summary in Sect. 3.4. Some approaches utilize the local operation to obtain neighborhood relations, which also contain frequency information of faces. By contrast, early work such as LT or HE handles the illumination problem with the original image. It spatially normalizes the whole given image without distinguishing between the illumination and facial features. Hence, methods of this kind may result in unreliable performance especially in uneven illumination.

4.3 Feature Scale

We have mentioned that some terms can be usually associated to the intrinsic facial feature component \(R\) and illumination component \(L\). The small scale (high resolution or high
frequency) features refer to $R$, and the large scale (low resolution or low frequency) features refer to $L$. Therefore the concept of feature scale is important for extracting $R$ from illuminated face images $I(= R \cdot L)$.

To our observation, the contribution and essentiality of large scale features and small scale features had been rarely evaluated and analyzed before [47], [56]. In [47], Xie et al. claimed that the small scale features represents the small intrinsic facial features, but the large scale features not only contain the illumination component but also the large intrinsic facial features. However, many existing approaches regard the whole large scale features as $L$ and discard them completely (e.g., LTV, DCT). As a result, they neglect the useful large intrinsic facial features. In order to utilize both of the small and large intrinsic features of a face, the method developed in [47] first decomposes the given face image into two images of small and large scale and then executes smoothing on the small scale image and normalizing on the large scale image, respectively. Final reconstruction is accomplished by fusing the two processed images of different scales. It makes use of the normalized large scale image rather than completely discarding it. The procedure can be illustrated in Fig. 2. It was proved that this method significantly outperformed LTV which also adopts the decomposition but only reserves the small scale image while discarding the whole large scale image.

More recently, in [48], Matsukawa et al. also kept the large intrinsic features by normalizing the large scale features rather than discarding them. Only the normalization method is different from that in [47]. They experimentally confirmed that their method could lead to better visual quality and more robust performance under varying illumination, compared with other approaches.

In summary, to create a more discriminating feature representation, one can consider along this direction of maintaining small scale intrinsic features together with large scale intrinsic ones. Nevertheless, to our knowledge, it is not easy to distinguish which is useful (intrinsic features) or not (illumination) in the large scale features. It is a familiar phenomenon that illumination and facial features tend to be removed or maintained simultaneously. This is what we want to talk about in the following analysis.

4.4 Common Problem

When surveying the majority of existing illumination processing approaches, some specific problem remains difficult to overcome.

One obvious problem is the negative effect of shadow boundaries. Figure 3 shows several face images of the same subject under varying illumination and their corresponding face representations using different illumination processing methods. The red blocks demonstrate the shadow boundaries. We can see that these methods are able to remove most effects caused by the illumination even within shadows. However, the artifacts of shadow boundaries are still extrusive in the processed images. The reason is that the luminance component $L$ is assumed to vary very slowly, which contradicts with the fact that shadow boundaries are of fast spatial changes. This negative effect could be very strong to impair the recognition effectiveness. In the literature, we found two famous approaches concerning about the processing of shadow boundaries. They are ASR and LTV. As introduced in Sect. 2, ASR [17] takes the illumination discontinuity into account and measures them during each iterated smoothing. LTV [18] uses TV-$L^1$ model for decomposition, which is able to preserve the strong edges in the illumination component. Figure 4 displays the face representations by ASR and LTV. It can be noticed that the shadow boundaries are largely suppressed by these two approaches. Despite their ability to counter the influence of shadow boundaries, they tend to lose more facial information at the same time, which is also notable in Fig. 4. This is one case of the more general problem which can be stated as: simultaneous increase or decrease of the facial features and the
illumination. This so called “synchronization” is usually controlled by specific parameters. In LTV, the parameter $\lambda$ in Eq. (8) is used to adjust the decomposition scale; when $\lambda$ becomes large, illumination and facial features will be removed to a large extent simultaneously. In ASR, the maximal iteration number for termination denoted by $T$ needs to be set in Eq. (7); if $T$ is too small, facial features will be hardly distinguished although shadows will be removed effectively. In LN, the block size $N$ in Eq. (19) is proportional to the extent of both facial feature maintenance and illumination redundancy. In WF, the parameter $\alpha$ in Eq. (16) controls the contrast in the neighborhood, affecting facial features and illumination at the same time. Mostly, compromised parameter values are selected based on the recognition performance with little theoretical analysis. This seems open problem which is worthy of further study.

5. What to Normalize

In some researches it is called “illumination normalization” to cover the whole passive approaches of illumination processing. However, this term may bring about confusion in this paper since we would like to distinguish between objects to be normalized. Some of the passive approaches aim to normalize the illumination of given images and others try to normalize the intra-class difference between images of one subject. In the following, we prefer to denote these two normalizations as “illumination normalization” and “intra-class difference normalization”, separately. It seems constructive to discuss this point because different viewpoints will lead to completely different research directions.

5.1 Illumination Normalization

Almost all of the existing approaches try to compensate the illumination in each of given images separately. They fundamentally correct the illumination or represent the faces with extracted illumination insensitive features, by focusing on the current image one by one. They need to analyze the property of illumination and construct a new face representation in an illumination insensitive way. These methods have been exploited extensively for recognition applications. However, it is still difficult to thoroughly separate the illumination and faces. For instance, the shadow boundaries are difficult to remove, even with the commonly adopted assumption. One more negative example is the illumination normalization using HE. HE copes with the illumination in different images separately and suppresses the illumination largely with regard to each given image. However, since input images are illuminated in different ways, the resultant images will be far different from each other even they are the same subject. Thus, images processed by HE may not be suitable for the subsequent recognition. Figure 5 shows some visual samples of illumination normalization by HE.

Consider that the ultimate goal is to accomplish face recognition (classification) under varying illumination with high accuracy. One can figure out that producing more similar cues for the same subject seems more important than trying to get rid of the illumination. The illumination only needs to be suppressed under the condition of not disturbing the classification. Actually, GRF is not fully removing illumination. It sacrifices the visual quality for closer representations of same subjects and obtains relatively better recognition performance.

5.2 Intra-Class Difference Normalization

As mentioned above, it is difficult to result in uniform face images, and thus “illumination processing in face recognition” can be considered in another way. That is not to normalize the illumination in each face image, but to obtain approximately identical face representation for same subjects. This can be deemed as intra-class difference normalization.

One example embracing this idea is maximizing a correlation (MAC)[52]. It attempts to maximize the intra-individual correlations rather than separately processing the illumination for each face image. With a gallery face image $I_G$ and a probe $I_P$, it represents the optimized results by

$$\{U'_G, I'_P\} = \{f(I_G, \xi'_G), f(I_P, \xi'_P)\}, \quad (25)$$

where $f(\cdot)$ is a transform function for illumination processing. $\xi'_G$ and $\xi'_P$ are parameters of $f$, which are estimated by solving the optimization problem of

$$\{\xi'_G, \xi'_P\} = \arg\max_{\xi_G, \xi_P} \text{corr}(f(I_G, \xi_G), f(I_P, \xi_P)), \quad (26)$$

where $\text{corr}\{}$ calculates the correlation between two face images. MAC processes images pair by pair between galleries and probes. It adjusts the illumination rather than normalizing it and aims to generate the illumination as similar as possible for the same subject. In other words, it normalizes the difference within the same class. MAC actually provides a nice scheme which needs to incorporate a proper existing or new illumination processing method as denoted by $f(\cdot)$ in Eq. (25). In [52], GC and LTV were tried in this scheme. Other methods can be also properly implemented in this scheme in the future work.

The intra-class difference normalization is likely to be more suitable for the face recognition task than the separate illumination normalization. The reason is that the former one can assist the final classification by diminishing the intra-class difference while prolonging the inter-class distance. For example, MAC helped achieve much better performance than only using an illumination processing.
method which focuses on each single input image in [52]. In this sense, a more potential way to inherit this idea may be a joint normalization of intra-class difference, which processes the illumination in all face images jointly.

6. Experiments

In this section, we do summative evaluations from two perspectives. One is the contribution of fundamental illumination correction methods to illumination insensitive face recognition. The other is the comparative performance of various illumination processing methods. Two databases were used for each assessment.

The first database, cropped Extended Yale B [57], consists of 38 subjects with 9 poses and 64 illumination conditions. We choose the frontal faces in our experiments to focus only on the illumination problem and there are totally 2432 face images. The size of each image is 192×168. We divide this database into five subsets according to the lighting angles. The five subsets are: subset 1 (0°∼12°, 266 images), subset 2 (13°∼25°, 456 images), subset 3 (26°∼50°, 456 images), subset 4 (51°∼77°, 532 images), and subset 5 (above 78°, 722 images). On this database, two experiments are conducted. Experiment 1 uses relatively more gallery images, that is, all the images from subset 1 are used as the gallery images (266 images), and the other images from subset 2 to 5 as the probe images. Experiment 2 uses fewer gallery images, only 1 image for each subject from subset 1 as the galleries (38 images), the rest images in subset 1 together with subsets from 2 to 5 as the probes. These two situations test the effectiveness of various methods with different amounts of reference information during recognition. Figure 6 displays 5 sample images of one subject from different subsets from Extended Yale B database.

The second database is CMU-PIE database [58], which contains 68 subjects under large variations in illumination, pose and expression, totally 41368 images. The illumination subset including 68 subjects under 21 different illumination directions (21 images per subject) is used in our experiments. All the images are properly cropped to the size of 161×161 with only the face region. To get fair evaluation, one image per subject (totally 68 images) is chosen as the galleries each time (totally 21 times) and the others are used as the probes. The results are averaged finally for each method. Figure 7 shows 21 face images with various illumination conditions of one subject from CMU-PIE database.

With regard to the recognition protocol, we use the raw pixel values of the resultant face representations of different methods to do template matching and evaluate each method by using the one nearest neighbor rule with three distance measures \( L_1 \) norm, \( L_2 \) norm and \( \chi^2 \) distance measure are separately defined by

\[
L_1(X, Y) = \sum_{i,j} |X_{i,j} - Y_{i,j}|, \\
L_2(X, Y) = \sqrt{\sum_{i,j} (X_{i,j} - Y_{i,j})^2}, \\
\chi^2(X, Y) = \sum_{i,j} \frac{(X_{i,j} - Y_{i,j})^2}{2(X_{i,j} + Y_{i,j})^2}.
\]

Recognition rate in our experiments means the rate between the number of correct recognitions and the total number of images used for test each time. Since different methods achieved the highest recognition rate using different distance measures, we choose the best result among the three measures for each method as its result for comparison.

6.1 Contribution of Fundamental Illumination Correction

We first explore the contribution of fundamental illumination correction methods to the illumination insensitive face recognition. For this purpose, we select four methods DCT [19], WA [25], TT [50] and GRF [21], each of which were originally developed by using one corresponding fundamental illumination correction method (i.e., LT, HE, GC, and GF, respectively). We test these methods before and after removing the corresponding fundamental illumination corrections, that is to compare the performance of DCT vs DCT without LT, WA vs WA without HE, TT vs TT without GC, and GRF vs GRF without GF.

Tables 1 and 2 show the comparative results of the two experiments on Extended Yale B database using different numbers of gallery images, respectively. The distance measure used by each method to achieve the best performance is denoted as well. The face recognition rates achieved by using only the fundamental illumination correction and the raw image with no processing are also presented. From Table 2, some methods perform better in subset 2 than 1. This is because these methods are likely to blur facial features as
Table 1: Recognition rates (%) on Extended Yale B database using the whole subset 1 as the galleries.

<table>
<thead>
<tr>
<th>Method</th>
<th>ORI</th>
<th>LT</th>
<th>DCT_no</th>
<th>DCT</th>
<th>HE</th>
<th>WA_no</th>
<th>WA</th>
<th>GC</th>
<th>TT_no</th>
<th>TT</th>
<th>GF</th>
<th>GRF_no</th>
<th>GRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subset2</td>
<td>90.13</td>
<td>96.05</td>
<td>98.90</td>
<td>94.74</td>
<td>99.12</td>
<td>92.11</td>
<td>88.82</td>
<td>96.05</td>
<td>99.56</td>
<td>99.56</td>
<td>96.93</td>
<td>98.90</td>
<td>99.34</td>
</tr>
<tr>
<td>Subset3</td>
<td>41.89</td>
<td>74.34</td>
<td>94.85</td>
<td>74.01</td>
<td>91.01</td>
<td>72.04</td>
<td>71.05</td>
<td>76.54</td>
<td>99.67</td>
<td>99.78</td>
<td>74.56</td>
<td>99.34</td>
<td>99.78</td>
</tr>
<tr>
<td>Subset4</td>
<td>5.45</td>
<td>51.45</td>
<td>70.50</td>
<td>58.59</td>
<td>65.03</td>
<td>47.37</td>
<td>48.13</td>
<td>53.39</td>
<td>97.58</td>
<td>98.50</td>
<td>49.17</td>
<td>90.98</td>
<td>85.90</td>
</tr>
<tr>
<td>Subset5</td>
<td>2.63</td>
<td>7.62</td>
<td>4.85</td>
<td>33.10</td>
<td>34.90</td>
<td>2.49</td>
<td>9.97</td>
<td>9.97</td>
<td>83.10</td>
<td>97.92</td>
<td>3.74</td>
<td>51.52</td>
<td>59.14</td>
</tr>
<tr>
<td>Avg.</td>
<td>30.01</td>
<td>51.05</td>
<td>59.72</td>
<td>60.95</td>
<td>67.63</td>
<td>47.02</td>
<td>48.80</td>
<td>52.77</td>
<td>93.61</td>
<td>98.80</td>
<td>49.43</td>
<td>81.26</td>
<td>82.73</td>
</tr>
</tbody>
</table>

Table 2: Recognition rates (%) on Extended Yale B database using one image for each subject from subset 1 as the galleries.

<table>
<thead>
<tr>
<th>Method</th>
<th>ORI</th>
<th>LT</th>
<th>DCT_no</th>
<th>DCT</th>
<th>HE</th>
<th>WA_no</th>
<th>WA</th>
<th>GC</th>
<th>TT_no</th>
<th>TT</th>
<th>GF</th>
<th>GRF_no</th>
<th>GRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subset2</td>
<td>95.2</td>
<td>96.49</td>
<td>90.79</td>
<td>86.84</td>
<td>98.68</td>
<td>86.40</td>
<td>81.58</td>
<td>96.49</td>
<td>82.89</td>
<td>84.65</td>
<td>96.93</td>
<td>85.09</td>
<td>85.96</td>
</tr>
<tr>
<td>Subset3</td>
<td>91.45</td>
<td>96.05</td>
<td>93.20</td>
<td>75.66</td>
<td>94.30</td>
<td>87.06</td>
<td>84.43</td>
<td>95.61</td>
<td>100.00</td>
<td>100.00</td>
<td>95.39</td>
<td>98.68</td>
<td>99.34</td>
</tr>
<tr>
<td>Subset4</td>
<td>21.49</td>
<td>36.84</td>
<td>57.46</td>
<td>45.83</td>
<td>44.74</td>
<td>33.33</td>
<td>35.31</td>
<td>36.84</td>
<td>82.24</td>
<td>83.55</td>
<td>28.95</td>
<td>82.46</td>
<td>84.43</td>
</tr>
<tr>
<td>Subset5</td>
<td>4.70</td>
<td>9.77</td>
<td>22.93</td>
<td>31.20</td>
<td>17.11</td>
<td>5.26</td>
<td>10.34</td>
<td>11.28</td>
<td>82.52</td>
<td>90.98</td>
<td>5.08</td>
<td>64.10</td>
<td>67.67</td>
</tr>
<tr>
<td>Avg.</td>
<td>32.46</td>
<td>38.39</td>
<td>43.65</td>
<td>44.53</td>
<td>48.29</td>
<td>33.25</td>
<td>34.71</td>
<td>38.76</td>
<td>79.49</td>
<td>89.39</td>
<td>35.01</td>
<td>64.79</td>
<td>67.29</td>
</tr>
</tbody>
</table>

Table 3: Average recognition rates (%) on CMU-PIE database.

<table>
<thead>
<tr>
<th>Method</th>
<th>ORI</th>
<th>LT</th>
<th>DCT_no</th>
<th>DCT</th>
<th>HE</th>
<th>WA_no</th>
<th>WA</th>
<th>GC</th>
<th>TT_no</th>
<th>TT</th>
<th>GF</th>
<th>GRF_no</th>
<th>GRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg.</td>
<td>28.47</td>
<td>34.48</td>
<td>40.70</td>
<td>44.50</td>
<td>54.77</td>
<td>31.04</td>
<td>35.59</td>
<td>34.39</td>
<td>88.32</td>
<td>98.39</td>
<td>38.47</td>
<td>84.17</td>
<td>85.18</td>
</tr>
</tbody>
</table>

6.2 Comparative Performance of Various Approaches

In this part, we present recognition results of comparative study on nine illumination insensitive approaches including DCT [19], WA [25], SQI [16], ASR [17], TT [50], GRF [21], WF [22], LTV [18], and NLM [42]. Figure 8 shows the average comparative results of the two experiments on Extended Yale B database. We can see that the highest three recognition rates are achieved by TT, ASR, and WF in both of these two experiments. NLM ranks the third and GRF ranks the fourth. It should be pointed out that a high recognition rate is often achieved by using a specific distance measure according to different methods. For example, TT using $L_1$ norm yields better result than ASR using $L_1$ norm; but TT using $L_2$ norm performs worse than ASR using $L_2$ norm. The distance measure used by each method to achieve the best performance is denoted. In addition, it can be found that SQI drops the most from experiment 1 to 2. This unstable result demonstrates that SQI is less reliable for illumination insensitive face recognition with small number of gallery images. This is because SQI does not consider about tackling shadow boundaries and thus its discriminative ability between facial features and illumination effects is not satisfactory enough and is sensitive to reference information. When less reference information is given, the illumination tends to be confused with the inter-class difference and thus more false recognition will be yielded. Figure 9 gives the average recognition rates of different methods on CMU-PIE database. It shows similar rank of the tested methods, in which TT, ASR, and WF generate the best three recognition results using $L_1$, $L_2$, and $L_1$ norm, respectively. With

![Fig. 8 Comparative results of various methods on Extended Yale B database.](image-url)
regard to the good performance of TT, several points should be noticed. Motivated by bottom-up human perception, it tries to estimate the intrinsic facial features directly without passing through the Luminance estimation $L$ or $u$ and thus inaccuracy due to Luminance estimation is not produced. Besides, via DoG filtering, not only most of the facial details can be preserved, but informative low frequency information can also remain. What’s more, with both local and global considerations, it adopts a reasonable combination of the three operations in the chain: balancing the regional dynamic range, processing the intensity gradients, and normalizing the overall contrast.

7. Conclusion

In this paper, some noteworthy issues in illumination processing for face recognition have been analyzed. Focusing on the special topic of face illumination processing, we explained the correspondence of several well-known approaches to the reflectance model. We then discussed various considerations including contribution of fundamental illumination correction methods, processing domain, feature scale, and a common problem. Furthermore, we discussed an essential question of what to actually normalize. From these explanations and discussions, potential directions for future work were also analyzed. Finally, the experimental results on the contribution of fundamental illumination correction to the illumination insensitive face recognition and comparative performance of various methods were given as well. It has been shown that the fundamental illumination correction is often important in increasing the insensitivity to extreme illumination and Tan and Triggs’ method (TT) using $L_1$ norm achieved the highest recognition rates among nine tested methods.

References

YAO and NAGAHASHI: ANALYSIS OF NOTEWORTHY ISSUES IN ILLUMINATION PROCESSING FOR FACE RECOGNITION


[40] Available at: http://en.wikipedia.org/wiki/Min_Yao

[41] Available at: http://en.wikipedia.org/wiki/Hiroshi_Nagahashi

[42] Copyright © 2013. IEEE. All rights reserved.

[43] Hiroshi Nagahashi received B.S. and Dr.Eng. degrees from Tokyo Institute of Technology in 1975 and 1980, respectively. Since 1990, he has been with Imaging Science and Engineering Laboratory, Tokyo Institute of Technology, where he is currently a professor. His research interests include pattern recognition, computer graphics, image processing, and computer vision.

[44] Min Yao received B.E. degree from Southeast University, China in 2009 and Ph.D. degree from Tokyo Institute of Technology, Japan in 2014. Currently, She is a Lecturer with College of Information Engineering, Shanghai Maritime University. Her research interests include pattern recognition, image processing, and face detection & recognition in difficult conditions.