Photometric Normalization Techniques for Illumination Invariance

AVTOR: Vitomir Štruc

Photometric Normalization Techniques for Illumination Invariance

ABSTRACT

Face recognition technology has come a long way since its beginnings in the previous century. Due to its countless application possibilities in both the private as well as the public sector, it has attracted the interest of research groups from universities and companies around the world. Thanks to this enormous research effort, the recognition rates achievable with the state-of-the-art face recognition technology are steadily growing, even though some issues still pose major challenges to the technology. Amongst these challenges, coping with illumination induced appearance variations is one of the biggest and still not satisfactorily solved. A number of techniques have been proposed in the literature to cope with the illumination induced appearance variations ranging from simple image enhancement techniques, such as histogram equalization or gamma intensity correction, to more elaborate methods, such as homomorphic filtering, anisotropic smoothing or the logarithmic total variation model. This chapter presents an overview of the most popular and efficient normalization techniques which try to solve the illumination variation problem at the preprocessing level. It assesses the techniques on the publicly available YaleB face database and explores their strengths and weaknesses from the theoretical and implementational point of view.

1. Introduction

Current face recognition technology has evolved to the point where its performance allows for its deployment in a wide variety of applications. These applications typically ensure (or impose) controlled conditions for the acquisition of facial images and, hence, minimize the variability in the appearance of different (facial) images of a given subject. Commonly controlled external factors in the image capturing process include ambient illumination, camera distance, pose and facial expression of the face, etc.

In these controlled conditions, state-of-the-art face recognition systems are capable of achieving the performance level which can match that of the more established biometric modalities, such as fingerprints, as shown in a recent survey (Gross et al., 2004; Phillips et al., 2007). However, the majority of the existing face recognition techniques employed in these systems deteriorate in their performance when employed in uncontrolled and unconstrained environments. Appearance variations caused by pose-, expression-and most of all illumination-changes pose challenging problems even to the most advanced face recognition approaches. In fact, it was

empirically shown that the illumination induced variability in facial images is often larger than the variability induced by the subject's identity (Adini et al., 1997), or, to put it differently, images of different faces appear more similar than images of the same face captured under severe illumination variations.

Due to this susceptibility to illumination variations of the existing face recognition techniques, numerous approaches to achieve illumination invariant face recognition have been proposed in the literature. As identified in a number of surveys (Heusch et al., 2005; Chen, W., et al., 2006; Zou et al., 2007), three main research directions have emerged with respect to this issue over the past decades. These directions tackle the problem of illumination variations at either:

- the pre-processing level,
- the feature extraction level, or
- the modeling and/or classification level.

When trying to achieve illumination invariant face recognition at the pre-processing level, the employed normalization techniques aim at rendering facial images in such a way that the processed images are free of illumination induced facial variations. Clearly, these approaches can be adopted for use with any face recognition technique, as they make no presumptions that could influence the choice of the feature extraction or classification procedures.

Approaches from the second group try to achieve illumination invariance by finding features or face representations that are stable under different illumination conditions. However, as different empirical studies have shown, there are no representations which would ensure illumination invariant face recognition in the presence of severe illumination changes, even though some representations, such as edge maps (Gao & Leung, 2002), gradient-based features (Wei & Lai, 2004), local binary patterns (Marcel et al., 2007) or Gabor wavelet based features (Liu, 2006; Štruc & Pavešić, 2009), are less sensitive to the influence of illumination. The unappropriatness of the feature extraction stage for compensating for the illumination induced appearance variations was also formally proven in (Chen et al., 2000).

The last research direction with respect to illumination invariant face recognition focuses on achieving illumination invariance at the modeling or classification level. Here, the techniques for compensating for the illumination changes are linked to the type of face model or classification technique employed in the face recognition system. Assumptions regarding the effects of illumination on the face model or classification procedure are made first and then based on these assumptions counter measures are taken to obtain illumination invariant face models or illumination insensitive classification procedures. Examples of these techniques include the famous illumination cones technique (Georghiades et al., 2001), the spherical

harmonics approach (Basri & Jacobs, 2003), the illumination light fields (Zhou & Chellapa, 2004), etc. While these techniques are amongst the most efficient ways of achieving illumination invariant face recognition, they usually require a large training set of facial images acquired under a number of lighting conditions and are, furthermore, also computationally expensive.

It has to be noted that all of the presented research directions represent valid efforts in solving the problem of illumination invariant face recognition. However, as we have pointed out, the effectiveness of tackling the illumination effects at the feature extraction level is questionable, while the modeling and/or classification stages impose often unrealistic requirements with respect to the size and characteristics of the training image set. In this chapter, we will, therefore, focus on the (in our opinion) most feasible way of achieving illumination invariance, namely, exploiting techniques which compensate for the illumination changes during the pre-processing stage. At this level computationally simple and simultaneously effective techniques for achieving illumination invariant face recognition can easily be devised.

In the remainder of the chapter we will first describe the mathematical models governing the normalization process at the preprocessing level and will then present the most popular normalization techniques proposed and presented in the literature. All techniques covered in the chapter will ultimately be assessed on facial images from the YaleB database.

2. Background

Before we can turn our attention to different preprocessing techniques proposed in the literature to achieve illumination invariant face recognition, we need to establish a mathematical model explaining the principles of image formation and/or scene perception, which is capable of forming the foundation for our pre-processing. To this end, researchers commonly turn to the so-called retinex theory developed and presented by Land and McCann in (Land & McCann, 1971), which states that the perceived sensation of color in a natural scene shows a strong correlation with reflectance, even though the amount of visible light reaching the eye depends on the product of reflectance and illumination (Land & McCann, 1971). This means that the human visual system is capable of correctly perceiving colors even in difficult illumination conditions by relying on the reflectance of the scene and in a way neglecting the scenes illumination. The theory suggests that the perception of a scene or an image of the scene can be modeled as follows:

$$I(x,y) = R(x,y)L(x,y), \tag{1}$$

where I(x,y) denotes an image of a natural scene, R(x,y) denotes the reflectance and L(x,y) stands for the illumination (or luminance) at each of the spatial positions (x,y).

Here, the reflectance R(x,y) is linked to the characteristics of the objects comprising the scene of an image and is dependant upon the reflectivity (or albedo) of the scenes surface (Short et al., 2004), or, in other words, it accounts for the illumination reflected by the objects in the scene (Delac et al., 2006). The luminance L(x,y), on the other hand, is determined by the illumination source and relates to the amount of illumination falling on the observed scene.

To produce illumination invariant versions of facial images, researchers commonly mimic the behavior of the human visual system and try to estimate the reflectance R(x,y) with the goal of adopting it as an illumination invariant representation of the face. Unfortunately, it is impossible to determine the reflectance of an image using Eq. (1), unless some assumptions are made regarding the nature of at least one of the factors of the presented imaging model. The most common assumptions adopted when estimating R(x,y) are (Park et al., 2008):

- edges in the image I(x,y) also represent edges in the reflectance R(x,y), and
- the luminance L(x,y) changes slowly with the spatial position of the image I(x,y).

Hence, the reflectance is assumed to be a high frequency phenomenon, while the luminance function is assumed to be a low-frequency phenomenon in nature. To determine the reflectance of an image, and thus, to obtain an illumination invariant image representation, the luminance L(x,y) of an image is commonly estimated first. This estimate L(x,y) is then exploited to compute the reflectance R(x,y) via the manipulation of the imaging model given by the expression (1). As we have already emphasized, the luminance is considered to vary slowly with the spatial position. It can, therefore, be estimated as a smoothed version of the original image I(x,y). Various smoothing filters and smoothing techniques have been proposed in the literature resulting in different normalization procedures that were successfully applied to the problem of face recognition under severe illumination changes. These (retinex-based) normalization techniques have common advantages as they exhibit relatively low computational complexity and impose no requirements for a special (generic or user-specific) training image set (Park et al., 2008).

Of course, other pre-processing techniques used for achieving illumination invariant face recognition can also be found in the literature. Some of these techniques are not based on the retinex theory, but can easily be linked to the imaging model in (1), while others are concerned with image characteristics which are not linked solely to illumination variations. While we will refer to the latter group of techniques as image enhancement techniques, we will use the term photometric normalization technique for any approach based on (or related to) the imaging model in (1).

3. Image enhancement techniques

Image enhancement techniques represent a group of pre-processing techniques which try to render the input images in such a way that the resulting enhanced images exhibit some predefined characteristics. These characteristics may include the dynamic range of the intensity values of the image, the shape of the images histogram, the contrast of the image, etc. While image enhancement techniques are often deployed to ensure at least some level of robustness to illumination changes, their primary goal is not illumination invariance. Nevertheless, various empirical studies have proven their usefulness for robust face recognition in the presence of illumination variations. Due to these characteristics, we will in this section briefly describe four different image enhancement techniques often used as preprocessing steps for face recognition, namely, the gamma intensity correction, the logarithmic transform, histogram equalization and histogram remapping.

3.1 Gamma intensity correction

Let us denote an arbitrary 8-bit input face image of size a×b pixels as I(x,y). The gamma intensity correction transforms the pixel intensity values at each spatial position (x,y), for x=1,2,…,a and y=1,2,…,b, in accordance with the following expression (Gonzalez & Woods, 2002):

$$I_{gam}(x,y) = I(x,y)^{\gamma}, \tag{2}$$

where $I_{gam}(x,y)$ denotes the gamma intensity corrected form of the face image I(x,y), and γ stands for the so-called gamma value which controls the type of the mapping. When γ is set to a value of γ>1, dark regions in the original image become lighter in the gamma corrected one; similarly, when γ is set to a value of γ<1 the light regions in the original image become darker in the gamma intensity corrected one. An example of the effect of different gamma values on the appearance of an image is presented in Fig. 1. Here, the left pair of images shows an input face image (on the left) and the gamma intensity corrected one for γ=0.5, while the right pair depicts a face image (on the left) and the gamma intensity corrected image for γ=4.

Figure 1: Two examples of gamma intensity corrected images for two different gamma values

3.2 Logarithmic transform

Let us adopt the same notation as in the case of gamma intensity correction and denote the input face image to be enhanced as I(x,y). The logarithmic transform non-linearly transforms the intensity values at each spatial position (x,y), for x=1,2,…,a and y=1,2,…,b, according to the following expression (Gonzalez & Woods, 2002):

$$I_{log}(x,y) = \log[I(x,y)], \qquad (3)$$

where I_{log}(x,y) denotes the logarithm transformed input image I(x,y).

The logarithmic transform improves the contrast of darker image regions by making them lighter and consequently ensures some level of robustness to illumination changes, as shown in Fig. 2.
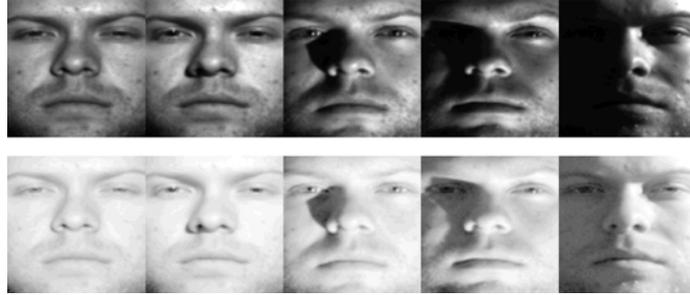

Figure 2: Impact of the logarithmic transform: original images (upper row), logarithm transformed images (lower row)

3.3 Histogram equalization

Let us again denote the input face image as I(x,y) and let us assume that the image is comprised of a total of *N* pixels with *k* grey-levels. Histogram equalization aims at transforming the distribution of the pixel intensity values of the input image I(x,y) into a uniform distribution and consequently at improving the image's global contrast (Gonzalez & Woods, 2002; Štruc et al., 2009).

Formally, histogram equalization can be defined as follows: given the probability $p(i)=n_i/N$ of an occurrence of a pixel with the grey level of *i*, where n_i denotes the number of pixels with the grey-level of *i* in the image, the mapping from the original intensity value *i* to the new transformed one i_{new} is given by (Heusch et al., 2005):

$$i_{new} = \sum_{j=0}^{i} \frac{n_j}{N} = \sum_{j=0}^{i} p(j). \qquad (4)$$

The above expression defines the mapping of the pixels' intensity values from their original range, for example the 8-bit interval 0-255, to the domain of [0,1]. Thus, to obtain pixel intensity values in the original range, the result has to be rescaled. A visual example of the effect of histogram equalization on the appearance of facial images is shown in Fig. 3.
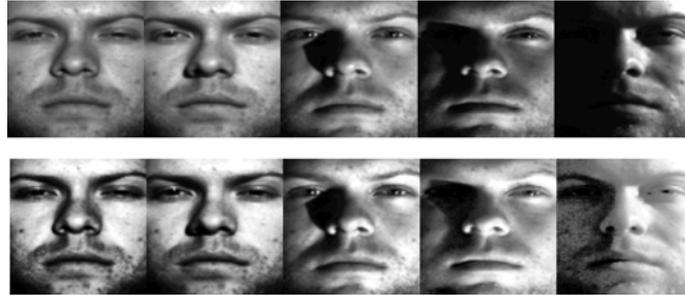
Figure 3: Impact of histogram equalization: original images (upper row), histogram equalized images (lower row)

Note how applying histogram equalization improves the contrast of the facial images, but as seen on the last image of the lower row also greatly enhances the background noise. The demonstrated property of contrast enhancement makes histogram equalization one of the most frequently employed image enhancement techniques in the field of face recognition.

3.4 Histogram remapping

More recently, Štruc et al. (Štruc et al., 2009) suggested that mapping a non-uniform distribution to facial images can also ensure robust face recognition in the presence of severe illumination changes. The authors' claim is based on the observation that while histogram equalization was empirically proven to provide enhanced face recognition performance when compared to unprocessed facial images, it is still only a useful heuristic, which represents a special case of the more general concept of histogram remapping techniques. With this class of techniques, the target distribution is not limited to the uniform distribution, but can rather represent an arbitrary one. Independent of the choice of the target distribution, the procedure of mapping an arbitrary distribution to the pixel intensity values of a facial image always follows the same procedure.

The first step common to all histogram remapping techniques is the transformation of the pixel intensity values using the rank transform. The rank transform is basically a histogram equalization procedure which renders the histogram of the given image in such a way that the resulting histogram approximates the uniform distribution. Here, each pixel value in the N-dimensional image I(x,y) is replaced with the index (or rank) R that the pixel would correspond to if the pixel intensity values were ordered in an ascending manner. For example, the most negative pixel-value is assigned the rank of 1, while the most positive pixel-value is assigned the ranking of N. The presented procedure is equivalent to histogram equalization; the only difference is in the domain the pixel intensity values are mapped to.

Once the rank R of each pixel is determined, the general mapping function to match the target distribution $f(x)$ may be calculated from (Gonzalez & Woods, 2002; Pelecanos & Sridharam, 2001; Štruc & Pavešić, 2009):

$$\frac{N-R+0.5}{N} = \int_{x=-\infty}^{t} f(x)dx \qquad (5)$$

and the goal is to find t. Obviously, the right hand side represents the cumulative distribution function (CDF) of the target distribution, while the left hand side represents a scalar value. If we denote the CDF of the target distribution as $F(x)$ and the scalar value on the left as u, then the mapped value t can be determined by computing the following expression:

$$t = F^{-1}(u), \qquad (6)$$

where F^{-1} denotes the inverse of the target distributions CDF.

To illustrate the effect of the histogram remapping technique on the appearance of a facial image, let us assume that our target distribution takes the following form:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \frac{\exp(-(\ln x - \mu)^2/2\sigma^2)}{x}, \qquad (7)$$

where the parameters μ and $\sigma>0$ define the shape of the presented lognormal distribution. Though the parameters of the distribution can be chosen arbitrary, we choose the mean value of $\mu=0$ and standard deviation of $\sigma=0.4$ here. These parameters result in (visually) properly normalized images as shown in Fig. 4
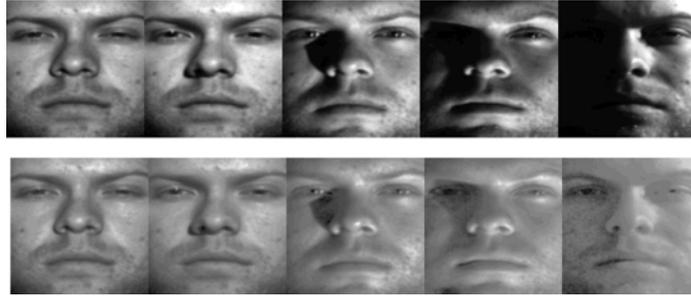

Figure 4: Impact of histogram remapping using a lognormal distribution: original images (upper row), images with a remapped histogram (lower row)

4. Photometric normalization techniques

In this section we introduce the most popular techniques based on the retinex theory. We present their strengths and weaknesses and provide visual examples of the effects they produce when applied to facial images.

4.1 The single scale retinex algorithm

The single scale retinex technique, originally named the center/surround retinex algorithm, is one of the straightforward approaches that can be derived from the imaging model defined by Eq. (1) under the common assumptions with regard to the characteristics of the reflectance and luminance functions of an image.

As proposed by Jobson et al. (Jobson et al., 1997a), the luminance and reflectance are first separated by taking the logarithm of the image, which results in the following expression:

$$\log I(x,y) = \log R(x,y) + \log L(x,y). \tag{8}$$

If we denote the logarithm of the reflectance as R' and consider that the luminance $L(x,y)$ can be estimated as a blurred version of the input image $I(x,y)$, we can rearrange Eq. (8) into the final form:

$$R'(x,y) = \log I(x,y) - \log\big[I(x,y) * K(x,y)\big], \tag{9}$$

where "*" denotes the convolution operator, $K(x,y)$ denotes a smoothing kernel and $R'(x,y)$ stands for the illumination invariant reflectance output of the single scale retinex algorithm.

It has to be noted that, when implementing the single scale retinex algorithm, we first have to choose an appropriate smoothing kernel. Several options were presented in the literature (see, for example, (Moore et al., 1991)), one of the most prominent, however, was introduced by Jobson et al. in (Jobson et al., 1997a). Here, the smoothing kernel takes the form of a Gaussian:

$$K(x,y) = k \exp(-\frac{r^2}{c^2}), \tag{10}$$

where $r = (x^2 + y^2)^{1/2}$, c denotes a free application dependant parameter, and k is selected in such a way that

$$\iint K(x,y)dxdy = 1. \tag{11}$$

Unfortunately, there are no rules on how to select the value of the parameter *c*. Instead, its value has to be determined through a trial and error procedure. Small values of the parameter *c* result in an extreme dynamic range compression of the intensity values of the images, while high values of the parameter produce only minimal changes in the reflectance when compared to the original image. Some examples of the effect of the single scale retinex algorithm are shown in Fig. 5.
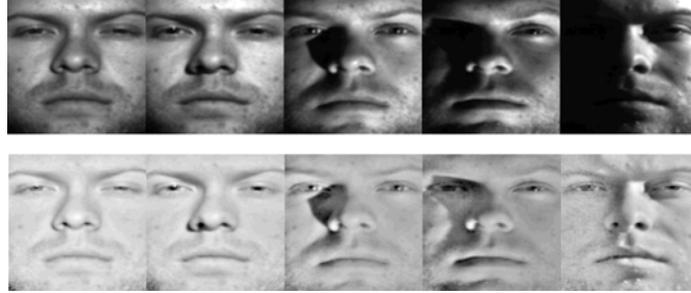

Figure 5: Impact of the single scale retinex technique: original images (upper row), processed images (lower row)

As the algorithm results in the compression of the images' dynamic range, the authors of the technique propose to clip the upper and lower parts of the histogram and rescale the remaining part of the histogram back to 8-bit interval.

4.2 The multi scale retinex algorithm

The multi scale retinex algorithm, proposed by Jobson et al. (Jobson et al., 1997b), is an extension of the single scale retinex approach presented in the previous section. While generally the single scale retinex technique produces good results with a properly selected Gaussian, there are still some shortcomings that limit a more extensive use of this normalization technique. The most significant issue not properly solved are halo affects[1], which are often visible at large illumination discontinuities. Such discontinuities are the result of strong shadows being casted over the face, which are in violation of one of the basic assumptions of the retinex based photometric normalization techniques, namely, that the luminance varies slowly with the spatial position.

To avoid the presented difficulties, the authors of the multi scale retinex technique proposed to use smoothing kernels of different sizes (i.e., different values of the parameter *c*) and basically combine the outputs of different single scale retinex implementations. Formally, the illumination invariant reflectance of an input face image I(x,y) using the multi scale retinex technique is computed as:

$$R'(x,y) = \sum_{i=1}^{M} w_i \left(\log I(x,y) - \log\left[I(x,y) * K_i(x,y) \right] \right), \tag{12}$$

---

[1] The term "halo effect" refers to the appearance of a halo in a certain region of the image.

where $K_i(x,y)$ denotes a Gaussian kernel at the *i*-th scale, and w_i stands for the weight associated with the *i*-th Gaussian kernel $K_i(x,y)$. Note that the Gaussian kernels used for the smoothing are defined by the expressions (10) and (11). Some visual examples of the multi scale retinex technique applied to facial images are shown in Fig. 6.
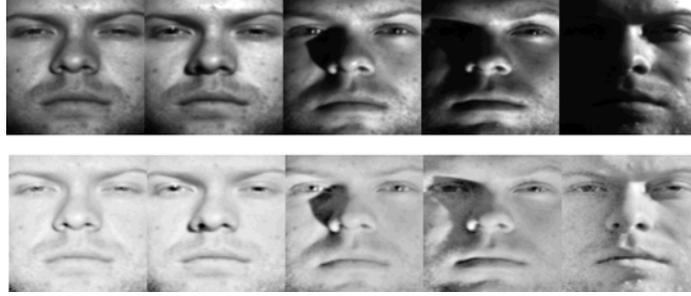

Figure 6: Impact of the multi scale retinex technique: original images (upper row), processed images (lower row)

While the presented multi scale retinex approach reduces the halo effect which can occur with the single scale retinex technique, the implementational issue of selecting the most appropriate kernel sizes for illumination invariant face recognition still remains.

4.3 The single scale retinex algorithm with adaptive smoothing

One of the latest modifications to the single scale retinex technique was presented by Park et al. in (Park et al., 2008). Here, the authors propose to tackle the halo effects often encountered with the original single scale retinex technique by incorporating an adaptive smoothing procedure with a discontinuity preserving filter into the single scale retinex algorithm with the goal of robustly estimating the images' luminance.

As stated by the authors, the key idea of adaptive smoothing is to iteratively convolve the input image I(x,y) with the 3×3 averaging mask *w(x,y)* whose coefficients reflect the discontinuity level of the input image at each of the spatial positions (x,y). Mathematically, the iterative smoothing procedure at the (*t+1*)-th iteration is given by

$$L^{(t+1)}(x,y) = \frac{1}{N^{(t)}(x,y)} \sum_{i=-1}^{1} \sum_{j=-1}^{1} L^{(t)}(x+i,y+j) w^{(t)}(x+i,y+j) \qquad (13)$$

and

$$L^{(t+1)}(x,y) = \max\left\{L^{(t+1)}(x,y), L^{(t)}(x,y)\right\}, \qquad (14)$$

where the normalizing factor $N^{(t)}(x,y)$ is defined as

$$N^{(t)}(x,y) = \sum_{i=-1}^{1}\sum_{j=-1}^{1} w^{(t)}(x+i,y+j).$$ (15)

The global and a local discontinuity measures in form of the gradient magnitude and local inhomogeneity, respectively are exploited to determine the values of the adaptive averaging mask $w(x,y)$ at each of the spatial positions (x,y). First, the gradient magnitude is computed based on the following expression:

$$|\nabla I(x,y)| = \sqrt{G_x^2(x,y) + G_y^2(x,y)},$$ (16)

where the partial derivatives are approximated as $G_x(x,y) = I(x+1,y) - I(x-1,y)$ and $G_y(x,y) = I(x,y+1) - I(x,y-1)$.

The second discontinuity measure, the local inhomogeneity $\tau(x,y)$, is defined as the average of local intensity differences at each spatial position (x,y):

$$\tau(x,y) = \frac{\sum\sum_{(m,n)\in\Omega} |I(x,y) - I(m,n)|}{|\Omega|}.$$ (17)

Here, Ω defines a 3×3 local neighborhood of the pixel location (x,y), and (m,n) indicates the locations of pixels in the local neighborhood Ω. Once the local inhomogeneity is determined, it is subjected to the following normalization procedure:

$$\tilde{\tau}(x,y) = \frac{\tau(x,y) - \tau_{min}}{\tau_{max} - \tau_{min}},$$ (18)

where τ_{max} and τ_{min} represent the maximal and minimal values of $\tau(x,y)$ across the entire face image. The final discontinuity measure is ultimately obtained by performing one last transformation which emphasizes the higher values of $\tilde{\tau}(x,y)$:

$$\hat{\tau}(x,y) = \sin(\frac{\pi}{2}\tilde{\tau}(x,y)).$$ (20)

Both the gradient magnitude as well as the normalized and transformed form of the local inhomogeneity are combined as $w(x,y) = \alpha(x,y)\beta(x,y)$ to produce the 3×3 averaging mask $w(x,y)$ for each spatial position (x,y). Here, $\alpha(x,y)$ and $\beta(x,y)$ denote the transformed discontinuity measures using the following conducting function:

$$g(d,k) = \frac{1}{1 + \sqrt{d/k}} \, .$$
(21)

Finally we have: $\alpha(x,y) = g(\hat{\tau}(x,y),h)$ and $\beta(x,y) = g(|\nabla I(x,y)|,S)$. Based on their empirical findings, the authors of the presented technique suggested to use between fifteen and twenty iterations to estimate the luminance function using equations (13) and (14) and to select the parameters of the conducting function somewhere in the range of 0≤h≤0.1 and 0≤S≤10. Similar to the single scale retinex technique, the final illumination invariant reflectance output is computed as:

$$R'(x,y) = \log I(x,y) - \log L(x,y) \, .$$
(22)

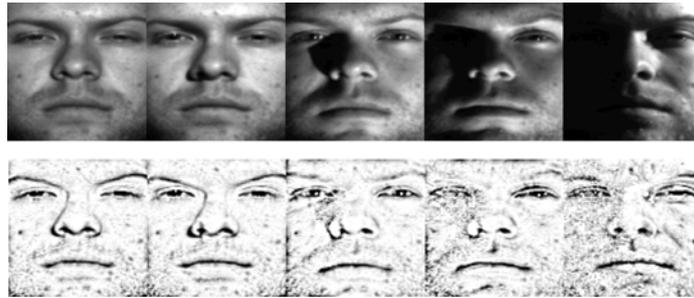Some examples of facial images processed with the presented technique are shown in Fig. 7.

Figure 7: Impact of the single scale retinex with adaptive smoothing technique: original images (upper row), processed images (lower row)

4.4 Homomorphic filtering

Homomorphic filtering is one of the few photometric normalization techniques operating in the frequency domain. Similar to other techniques based on the retinex theory, the homomorphic filtering technique first separates the reflectance and luminance functions of the image I(x,y) by taking the natural logarithm of the imaging model in (1). The result, similar to the one presented in (8), is then used at the starting point for the normalization procedure.

The second step of the technique represents the transformation of the image from the spatial to the frequency domain, which is easily achieved using the Fourier transform (Short et al., 2004):

$$\mathcal{F}\{\log I(x,y)\} = \mathcal{F}\{\log R(x,y)\} + \mathcal{F}\{\log L(x,y)\},$$
(23)

which can be written as the sum of two functions in the frequency domain. The first contains mainly high frequency components and is the frequency equivalent of the

spatial reflectance, the second function, on the other hand, is composed of mainly low frequency components and corresponds to the spatial luminance. The Fourier transform of the input image $Z(u,v)= \mathcal{F}\{\log I(x,y)\}$ can be filtered (in the frequency domain) with the filter or transfer function $H(u,v)$ that reduces the low frequencies and amplifies the high frequencies (Heusch et al., 2005; Short et al., 2004). The final illumination invariant image $I'(x,y)$ is ultimately obtained by finding the inverse transform of the filtered image and taking its exponential, i.e.,

$$I'(x,y)=\exp\{\mathcal{F}^{-1}[H(u,v).Z(u,v)]\}, \qquad (24)$$

where "." denotes the element-wise multiplication. Note that the result of this normalization procedure is a normalized image $I'(x,y)$ rather than the reflectance of the image since no direct subtraction of the luminance was performed. Nevertheless, the result still approximates the reflectance since the effect of the luminance was reduced and that of the reflectance emphasized through the filtering operation in the frequency domain. Some visual examples of the effect of homomorphic filtering are shown in Fig. 8.
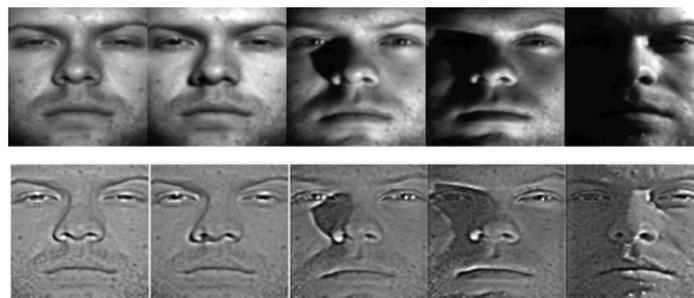

Figure 8: Impact of the homomorphic filtering technique: original images (upper row), processed images (lower row)

It has to be noted that even though the filtering operation in the presented technique is performed in the frequency rather than the spatial domain, the basic issue of filter design still remains. To achieve a suitable performance of illumination invariant face recognition the filter parameters must be determined empirically.

4.5 The self-quotient image

Similar to the single scale retinex technique with adaptive smoothing, the self-quotient image also represents a more recent addition to the group of photometric normalization techniques. Even though, originally not derived from the retinex theory, the self-quotient image can, nonetheless, be linked to other retinex-based approaches, especially to the multi scale retinex approach. We will not discuss the underlying theory of the self quotient image at this time, but rather point out that it is based on the so-called Lambertian model and the concept of the quotient image.

The reader should refer to (Basri & Jacobs, 2003) and (Shashua & Riklin-Raviv, 2001) for a more detailed description of these concepts. Here, we will only focus on the basic characteristics of the technique and the similarities with the retinex based methods.

The self-quotient image, proposed and presented by Wang et al. in (Wang et al., 2004), is based on a mathematical model similar to the one presented in Eq. (1). Therefore, the illumination invariant image representation Q(x,y) can be derived in the form of the following quotient:

$$Q(x,y) = \frac{I(x,y)}{K(x,y)*I(x,y)},$$
(25)

where the nominator of the above expression denotes the original input face image and the denominator represents the smoothed version of the input image. Like with the single and multi scale retinex approaches, K(x,y) again represents a smoothing kernel. Clearly, the technique takes its name from the fact that the illumination invariant image representation is derived based on the quotient of the original and smoothed version of the same input image.

The key element of the self-quotient image, which distinguishes the procedure from other similar techniques presented in the literature, is the structure of the smoothing kernel K(x,y). For each convolution region, the kernel is modified using a weighting function, which is constructed as follows: first the mean value τ of the convolution region is computed, and based on this value, two non-overlapping regions (denoted as M_1 and M_2) are constructed (Du & Ward, 2005; Heusch et al., 2005). Each pixel from the convolution region is then assigned to one of the two sub-regions M_1 and M_2 based on the following criterion (Wang et al., 2004):

$$I(x,y) \in \begin{cases} M_1, & \text{if } I(x,y) \leq \tau \\ M_2, & \text{if } I(x,y) > \tau \end{cases}.$$
(26)

Assuming that there are more pixels in M_1 than in M_2 then the weights of the weighting function for the Gaussian smoothing filter G(x,y) are given by:

$$W(i,j) = \begin{cases} 1, & \text{if } I(x,y) \in M_1 \\ 0, & \text{if } I(x,y) \in M_2 \end{cases},$$
(27)

and the final smoothing kernel K(x,y) is subject to the following condition:

$$\frac{1}{M} \sum_{\Omega} W(x,y).G(x,y) = \sum_{\Omega} K(x,y) = 1.$$
(28)

15

In the above expressions *M* denotes a normalizing factor, G(x,y) represents the original Gaussian kernel and Ω stands for the convolution region. As stated by the authors of the self-quotient image technique, the essence of the anisotropic filter lies in the fact that it smoothes only the main part of the convolution region and, therefore, effectively preserves discontinuities present in the image.

Due to similar reasons as the authors of the multi scale retinex technique, Wang et al. (Wang et al., 2004) also proposed to use different filter scales to produce the final illumination invariant image representation. This multi scale form of the self quotient image is obtained by a simple summation of self-quotient images derived with different filter scales. As the final processing step, a non-linear (logarithm, sigmoid function, …) mapping is applied to the self-quotient image to compress the dynamic range. Some examples of the deployment of the self-quotient image technique are shown in Fig. 9.
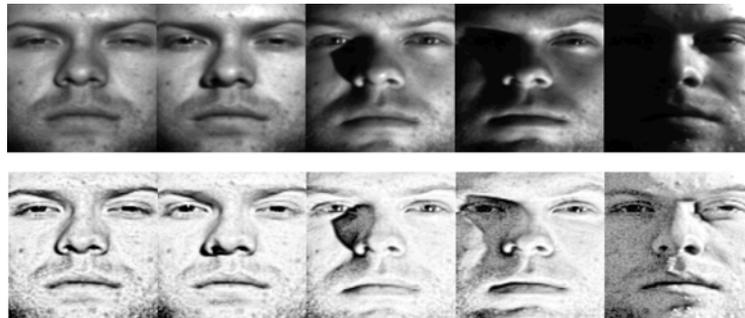

Figure 9: Impact of the self-quotient image technique: original images (upper row), processed images (lower row)

4.6 DCT-based normalization

The discrete-cosine-transform-based (DCT) photometric normalization technique is a recently proposed normalization technique introduced by Chen et al. in (Chen, W., at al., 2006). Like the majority of approaches already presented in this chapter, the technique relies on the retinex theory and its accompanying imaging model and, hence, makes similar assumptions about the characteristics of the illumination induced appearance variations.

The technique presumes that the illumination variations are related to the low frequency coefficients of the DCT transform and suggests discarding these low frequency coefficients before transforming the image back to the spatial domain via the inverse DCT in order to obtain an illumination invariant face representation (Chen, W., at al., 2006). To achieve illumination invariance, the DCT-based normalization takes the following steps: first the technique takes the logarithm of the input image I(x,y) to separate the reflectance and luminance. Next, the entire image is transformed to the frequency domain via the DCT transform, where the

manipulation of the DCT coefficients, with the goal of achieving illumination invariance, takes place. Here, the first DCT coefficient C(0,0) is set to:

$$C(0,0) = \log \mu \cdot \sqrt{MN} \,,\qquad\qquad (29)$$

where M and N denote the dimensions of the input image I(x,y) and μ is chosen near the mean value of I(x,y).

A predefined number of DCT coefficients (drawn from the DCT coefficient matrix in a zigzag manner) encoding the low-frequency information of the image is then set to zero. As the final step, the modified matrix of DCT coefficients is transformed back to the spatial domain via the inverse DCT to produce the illumination invariant representation of the facial image. Some examples of applying the described procedure to face images are presented in Fig. 10.
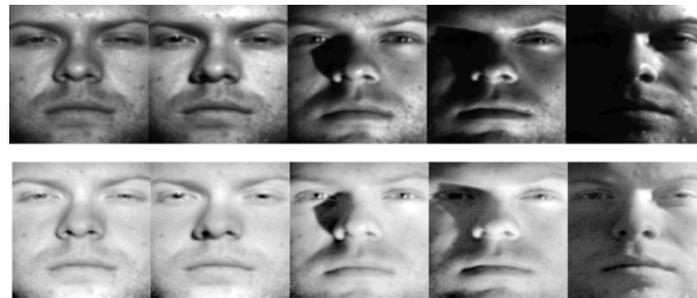
Figure 10: Impact of the DCT-based normalization technique: original images (upper row), processed images (lower row)

It has to be noted that the authors of the technique suggest discarding between 18 and 25 DCT coefficients for the best recognition performance.

4.7 Wavelet-based normalization

Du and Ward (Du & Ward, 2005) presented a photometric normalization technique which is based on the wavelet transform. The authors proposed to decompose the image via the 2D discrete wavelet transform (DWT) as to obtain the following four sub-bands: the low-low sub-band generated by the approximation coefficients and the low-high, the high-low and high-high sub-bands generated by the detail coefficients. The level one decomposition of an input face image into the four sub-bands using the DWT is presented in Fig. 11. Note that the detail-coefficient-images (i.e., the images resembling the gradient magnitude of the input image) are scaled to the 8-bit interval for visualization purposes.
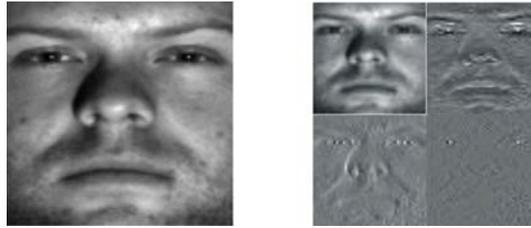
Figure 11: An sample image from the YaleB database (left) and its wavelet decomposition (right)

After the decomposition, the four sub-bands are subjected to the photometric normalization procedure. First, histogram equalization is applied to the approximation coefficients to increase the dynamic range of the image and to enhance the image's contrast. Then, the detail coefficients are multiplied by a scalar value higher than 1 to enhance edges present in the input image. Once all the sub-bands have been modified, the illumination invariant face representation is obtained using the inverse DWT. The described procedure resembles the image enhancement technique histogram equalization, with the distinction that it also enhances the high frequency information contained in the input image. Some examples of the effect of the presented technique on the appearance of the facial images and a comparison with histogram equalization is shown in Fig. 12.


Figure 12: Impact of the wavelet-based normalization technique: original images (first pair), histogram equalized images (second pair), and normalized images using the presented technique (third pair)

Note that the presented wavelet-based normalization technique could be extended to the multi-resolution case (which is a common practice with the DWT) with an arbitrary wavelet. The authors, however, did not focus on this issue and adopted only the level one decomposition of the images using Daubechies wavelets for their normalization procedure**.**

4.8 Wavelet-based image denoising

Another wavelet-based photometric normalization technique was proposed by Zhang et al. in (Zhang et al., 2009). Here, the wavelet-based image denoising approach is exploited to obtain an illumination invariant representation of the facial image.

The technique starts with the modified imaging model of the retinex theory given by (8). Under the assumption that the key facial features are high frequency phenomena equivalent to "noise" in the image denoising model, the authors propose to estimate the luminance L'(x,y)=log L(x,y) by the wavelet denoising model and then to extract the illumination invariant reflectance R'(x,y)=log R(x,y) in accordance with Eq. (22).

Let us denote the wavelet coefficient of the input image I'(x,y)=log I(x,y) as X(x,y)=W(I'(x,y)), where W stands for the 2D DWT operator; and, similarly, let Y(x,y)=W(L'(x,y)) denote the matrix of wavelet coefficients of the luminance L'(x,y). The estimate of the luminance in the wavelet domain Y(x,y) is then obtained by modifying the detail coefficients of X(x,y) using the so-called soft thresholding technique and keeping the approximation coefficients unaltered. Here, the soft thresholding procedure for each location (x,y) is defined as:

$$Y_s(x,y) = \begin{cases} X_s(x,y) - T, & \text{if } X_s(x,y) \geq T \\ X_s(x,y) + T, & \text{if } X_s(x,y) \leq T \\ 0, & \text{if } |X_s(x,y)| < T \end{cases} , \tag{30}$$

where $X_s(x,y)$ denotes one of the three sub-bands generated by the detail DWT coefficients (either the low-high (LH), the high-low (HL) or the high-high (HH) sub-bands, i.e., $s \in \{LH, HL, HH\}$), $Y_s(x,y)$ stands for the corresponding soft thresholded sub-band and T represent a predefined threshold.

It is clear that for an efficient rendering of the facial images, an appropriate threshold has to be defined. The authors propose to compute the threshold T as follows:

$$T = \frac{\sigma^2}{\sigma_X}, \tag{31}$$

where the standard deviations σ and σ_X are robustly estimated from:

$$\sigma = \frac{mad(|X_{HH}(x,y)|)}{\lambda}, \quad \sigma_Y^2 = var(X_s(x,y)) \quad \text{and} \quad \sigma_x = \sqrt{\max(\sigma_Y^2 - \sigma^2, 0)}. \tag{32}$$

In the above expressions "mad" denotes the mean absolute deviation and "var" denotes the variance. Note that the noise variance σ^2 is estimated from the HH sub-band, while the signal standard deviation σ_X is computed based on the estimate of the variance of the processed sub-band $X_s(x,y)$ for $s \in \{LH, HL, HH\}$. For an optimal implementation of the presented denoising procedure, the authors suggest using a value of λ somewhere in the range from 0.01 to 0.30.

Once, all three detail coefficient sub-bands have been thresholded, they are combined with the unaltered approximate coefficient sub-band to form the denoised wavelet coefficient matrix Y(x,y). The estimate of the luminance in the spatial domain is ultimately obtained by applying the inverse DWT to the wavelet coefficients in Y(x,y), and can be used to compute the illumination invariant reflectance. Some visual examples of the effect of the presented technique are shown Fig. 13.

Figure 13: Impact of the wavelet-based image denoising technique: original images (upper row), processed images (lower row)

4.9 Isotropic smoothing

Isotropic smoothing follows a similar principle as the single or multi scale retinex algorithm already presented in this chapter. It tries to estimate the luminance L(x,y) of the imaging model in (1) as a blurred version of the original input image I(x,y). However, it does not apply a simple smoothing filter to the image to produce the blurred output, but rather constructs the luminance function L(x,y) by minimizing the following energy-based cost function (Short et al., 2004):

$$J(L(x,y)) = \int_x\int_y (L(x,y) - I(x,y))^2\, dxdy + \lambda \int_x\int_y (L_x^2(x,y) + L_y^2(x,y))dxdy \,, \qquad (33)$$

where the first term forces the luminance L(x,y) to be close to the original input image I(x,y), and the second term imposes a smoothing constraint on L(x,y), and the parameter λ controls the relative importance of the smoothing constraint [16].

As stated in (Heusch et al., 2005), the problem in (33) can be solved by a discretized version of the Euler-Lagrange diffusion process:

$$I(x,y) = L(x,y) + \lambda\big[(L(x,y) - L(x,y-1)) + (L(x,y) - L(x,y+1)) + \\ + (L(x,y) - L(x-1,y)) + (L(x,y) - L(x+1,y))\big]. \qquad (34)$$

Once, the above equation is set up for all possible pixel locations (x,y) it forms a large sparse linear system of equations that can be rearranged into the following matrix form:

$$\mathbf{A} \cdot L_v = I_v \,, \tag{35}$$

where I_v represents the vector form of the input image I(x,y), L_v stands for the vector form of the luminance we are trying to compute and $\mathbf{A}$ denotes the so-called differential operator. It has to be noted that the dimensionality of the operator $\mathbf{A}$ is enormous as it represents a N×N matrix with N being the number of pixels in I(x,y). The equation given by (35), therefore, cannot be solved efficiently by a direct inversion of the square matrix $\mathbf{A}$, rather multigrid methods have to be exploited to produce an estimate of the luminance L(x,y). The description of appropriate multigrid methods to solve the problem given by (35) is unfortunately beyond the scope of this chapter. The reader should refer to (Heusch et al., 2005) for a more detailed description of these methods.

Let us assume that the multigrid method has been deployed to produce an estimate of the luminance L(x,y). Then, the illumination invariant reflectance can ultimately be computed by simply rearranging Eq. (1). Some examples of the normalized images using isotropic smoothing are shown in Fig. 14.

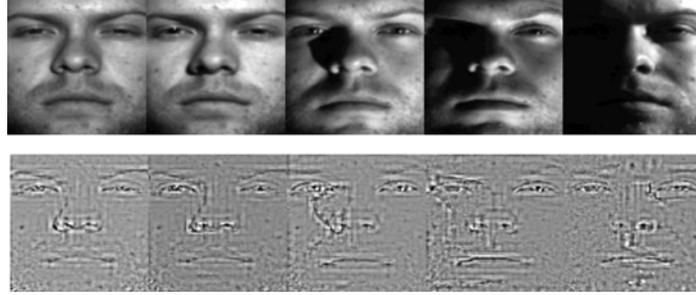Figure 14: Impact of the isotropic smoothing technique: original images (upper row), processed images (lower row)

4.10 Anisotropic smoothing

Photometric normalization using anisotropic smoothing is in its nature very similar to the isotropic smoothing technique presented in the previous section. The technique, proposed by Gross and Brajovic in (Gross & Brajovic, 2003), generalizes upon the energy-based cost function in (33) by introducing an additional weight function ρ(x,y) featuring the anisotropic diffusion coefficients, which enable the penalization of the fit between the original image I(x,y) and the luminance L(x,y). Or, in other words, the weight function ρ(x,y) ensures that the fit between the input image I(x,y) and the

luminance L(x,y) can be influenced by an additional criterion. Thus, anisotropic smoothing is based on the following cost function:

$$J(L(x,y)) = \int_x \int_y \rho(x,y)(L(x,y) - I(x,y))^2 \, dxdy + \lambda \int_x \int_y (L_x^2(x,y) + L_y^2(x,y))dxdy \, . \quad (36)$$

Again, the above expression is solved by a discretized version of the Euler-Lagrange diffusion process, which for the anisotropic case takes the following form:

$$I(x,y) = L(x,y) + \lambda \left[\frac{1}{\rho(x,y-1)}(L(x,y) - L(x,y-1)) + \frac{1}{\rho(x,y+1)}(L(x,y) - L(x,y+1)) + \right.$$
$$\left. + \frac{1}{\rho(x-1,y)}(L(x,y) - L(x-1,y)) + \frac{1}{\rho(x+1,y)}(L(x,y) - L(x+1,y)) \right]. \quad (37)$$

Gross and Brajovic suggested weighing the smoothing procedure with the inverse of the local image contrast. Thus, the goodness of fit between the input image and luminance is penalized by the inverse of this local contrast. They suggested two different contrast measures between pixel locations a and b, namely, the Weber and Michelson contrasts, which are defined as:

$$\rho(b) = \frac{|I(a) - I(b)|}{\min(I(a), I(b))} \qquad \text{and} \qquad \rho(b) = \frac{|I(a) - I(b)|}{|I(a) + I(b)|}, \quad (38)$$

respectively. Note that the local contrast is actually a function of two pixel locations, i.e., $\rho(b) = \rho(a,b)$; however, to be consistent with the notation in expression (37) we made a modification of the definitions of the contrast.

Like with the isotropic case, setting up Eq. (37) for all possible pixel locations results in a large and sparse linear system of equation (similar to the one presented in (35)), which is efficiently solved by a properly selected multigrid method. Some examples of the effect of anisotropic smoothing are shown in Fig. 15. Here, the value of the parameter λ of $\lambda=7$ was used and Michelson's contrasts was exploited as the local contrast measure.
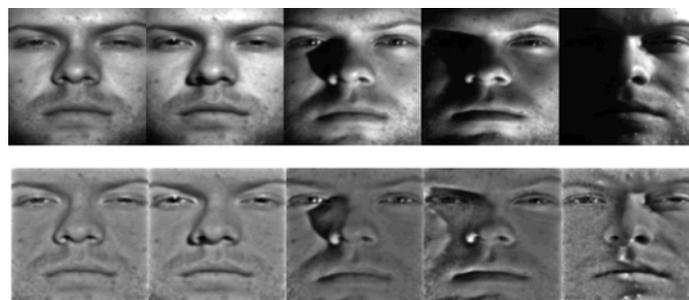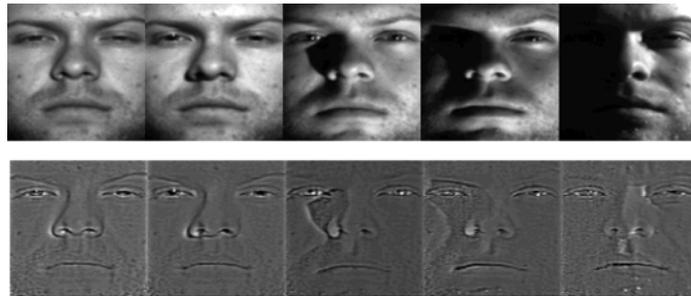
Figure 15: Impact of the anisotropic smoothing technique: original images (upper row), processed images (lower row)

It has to be noted that the usefulness of the anisotropic smoothing procedure heavily depends on the right choice of the parameter λ (the same goes for the isotropic smoothing). Unfortunately, there is no explicit rule on how to select the parameter to achieve the best possible recognition results in the presence of illumination induced appearance variations.

4.11 The logarithmic total variation model

The last photometric normalization technique covered in this chapter is the logarithmic total variation model proposed by Chen et al. in (Chen, T. et al., 2006). Like the self-quotient image, the logarithmic total variation model relies on the Lambertian model of images rather than the imaging model derived from the retinex theory.

The logarithmic total variation model is based on a variant of the total variation model commonly abbreviated as TV-L^1 (see (Chen, T.F. et al., 2004) for a more detail description of the model). This TV-L^1 model is capable of separating or decomposing an input image I(x,y) into two distinct outputs, the first, denoted as u(x,y), containing large scale facial components, and the second, denoted as v(x,y), containing mainly small scale facial components. To be consistent with the notation adopted in previous sections, we will still presume that an input image can be written as a product (I(x,y)=R(x,y)L(x,y)) of two components; however, as already indicated, these components are linked to the Lambertian model and, hence, represent the surface albedos and light received at each location (x,y).

The photometric normalization procedure starts by taking the logarithm of input face image I'(x,y) = log I(x,y), and then solves the variational problem of the following form to estimate the large scale component u(x,y):

$$u(x,y) = \arg\min_u \int_\Omega |\nabla u(x,y)| + \lambda |I'(x,y) - u(x,y)| dx , \qquad (39)$$

where $|\nabla u(x,y)|$ denotes the total variation of u(x,y) and λ stands for a scalar threshold on scale.

As noted in (Chen, T. et al., 2006), the expression given by (39) can be solved by a number of techniques, for example, PDE-based gradient descent techniques, interior-point second order cone programs or network flow methods. The description of these methods is beyond the scope of this chapter; however, the

reader should refer to (Alizadeh & Goldfarb, 2003) for more information on some of the listed techniques.

Once the large scale component u(x,y) is computed, it can be used to determine the illumination invariant small scale component v(x,y) as:

$$v(x,y) = I'(x,y) - u(x,y). \qquad (40)$$

It was shown by the authors of the technique that the logarithmic total variation model exhibits desirable properties, such as edge preservation and multi-scale decomposition (Chen, T. et al., 2006). The authors, furthermore, suggest using the following values for the scalar threshold λ: λ=0.7 to 0.8 for images of 100 × 100 pixels, λ=0.35 to 0.4 for images of 200 × 200 pixels, and λ=0.175 to 0.2 for images of 400 × 400 pixels. Some examples of normalized images using the presented technique are presented in Fig. 16.
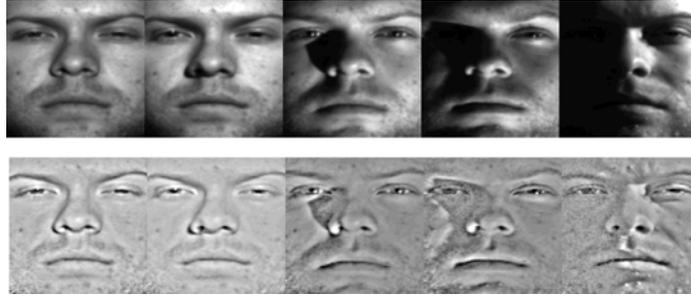
Figure 16: Impact of the logarithmic total variation model: original images (upper row), processed images (lower row)

5. Experimental assessment

5.1 Databases and experimental setup

To assess the performance of the presented photometric normalization techniques, we conduct face recognition experiments using the publicly available YaleB and XM2VTS face databases (Georhiades et al., 2001; Messer et al., 1999).

The first, the YaleB face database, was recorded at the Yale University and contains images of 10 distinct subjects taken under 576 different viewing conditions (9 poses × 64 illumination conditions). Thus, a total of 5760 images is featured in the database and can be used for assessing the robustness of different algorithms to facial illumination and pose changes. However, as this chapter is only interested in the illumination induced appearance variations and their impact on the recognition performance, we employ a subset of 640 facial images with frontal pose in our experiments.

Following the suggestion of the authors of the database, we divide the selected 640 experimental images into five image sets according to the extremity of illumination present during the image acquisition (the reader is referred to (Georhiades et al.,

2001) for a detailed description of these subsets). We can see from Fig. 17 that the first image set (S1) comprises images captured in "excellent" illumination conditions, while the conditions get more extreme in the image sets two (S2) to five (S5). Due to these characteristics, we employ the first image set (featuring 7 images per subject) for training and the remaining ones for testing. Such an experimental setup results in highly miss-matched conditions for the recognition stage and poses a great challenge to the photometric normalization techniques (Štruc & Pavešić, 2009). Furthermore, it is also in accordance with real life settings, as the training and enrollment stages are commonly supervised and, hence, the training and/or enrollment images are usually of good quality. The actual operational conditions, on the other hand, are typically unknown in advance and often induce severe illumination variations.
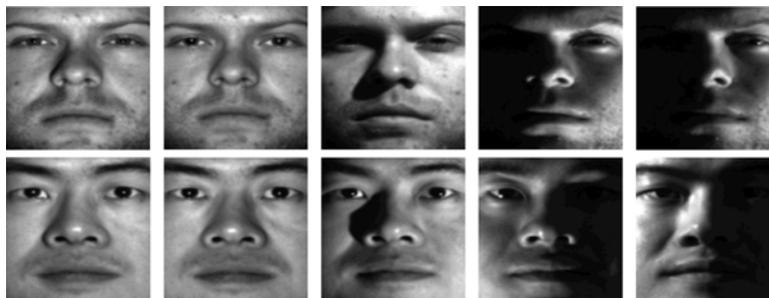
Fig. 17: Sample images of two subjects from the YaleB face database drawn from (from left to right): image set 1 (S1), image set 2 (S2), image set 3 (S3), image set 4 (S4) and image set 5 (S5).

The second, the XM2VTS database, features 2360 facial images of 295 subjects, with the images being captured under controlled conditions. For the experiments presented in the next sections, the database is divided into client and impostor groups and these groups are further partitioned into training, evaluation and test sets, as defined by the first configuration of the experimental protocol associated with the database (Messer et al., 1999). As the XM2VTS database features only images captured in controlled conditions, it is employed as a reference database for measuring the effect that any given photometric normalization technique has on images not affected by illumination changes. Some sample images from the database are shown in Fig. 18.

Fig. 18: Sample images from the XM2VTS database

In all experiments we use Principal Component Analysis (PCA) as the feature extraction technique, and the cosine similarity measure in conjunction with the nearest neighbor classifier for the classification. To make the experimental setup more demanding we retain merely 15% of the PCA coefficients, while the remaining 85% are discarded. This setup results in 10 PCA coefficients being employed for the experiments on the YaleB database and 90 PCA coefficients being used for the experiments on the XM2VTS database. Note that the feature vector length is not the primary concern of our experiments and could also be set differently.

Prior to feature extraction, we first apply a pre-processing procedure to all experimental images to remove artifacts not related to the face and to align the facial landmarks. The images are then cropped to the standard size of 128×128 pixels and stored as 8-bit grey-scale images for further processing. Note that manually marked eye coordinates are adopted for the presented pre-processing stage to ensure that localization errors do not interfere with the recognition results.

5.2 Assessing the image enhancement techniques

In our first series of face recognition experiments, we assess the performance of the image enhancement techniques, namely, histogram equalization (HQ), gamma intensity correction (GC) with $\gamma=4$, logarithmic transformation (LT) and remapping of the image histogram to a lognormal distribution (HL) for $\mu=0$ and $\sigma=0.4$. For baseline comparisons, experiments on unprocessed grey scale images (GR) are conducted as well. Note that the listed techniques are tested separately from other photometric normalization techniques due to their intrinsic characteristics, which make them more of a complement than a real substitute for other, more elaborate, photometric normalization techniques - an observation supported by the following facts:

- image enhancement techniques do not result in dynamic range compression as is the case with the photometric normalization techniques; rather, they cause a remapping of the pixel intensity distribution of the processed facial images,
- image enhancement techniques do not rely on the mathematical model presented in Section 2, but are linked to basic image properties, such as, for example, image contrast, and
- image enhancement techniques (histogram equalization in particular) were shown to improve the recognition performance of various face recognition techniques even when the facial images exhibited no illumination induced appearance variations; hence, they cannot be considered as being solely photometric normalization techniques.

The results of the assessment on the YaleB database in terms of the rank one recognition rate (in %) for all tested image sets are shown in Table 1. Here, the

26

numbers in the brackets next to the image set labels S1 to S5 denote the number of images in each of the test subsets. We can see that on the experimental set S5 all image enhancement techniques resulted in an increase of the recognition rate, while surprisingly only the HQ and HL improved upon the baseline recognition rate on the test sets S3 and S4.

Image set	GR	HQ	GC	LT	HL
S2 (120)	100	100	100	100	100
S3 (120)	93.3	99.2	73.3	67.5	95.8
S4 (140)	42.1	53.6	35.7	33.6	75.0
S5 (190)	13.7	53.2	35.8	35.3	75.0

Table 1: Rank one recognition rates obtained on the YaleB database with the tested image enhancement techniques

The results of the assessment on the XM2VTS database are presented in Table 2. Note that differently from the YaleB database, images from the XM2VTS database are used in face verification rather than identification experiments. Hence, the performance of the photometric normalization techniques is assessed in terms of the false rejection error rate (FRR), the false acceptance error rate (FAR), and the half total error rate (HTER) defined as the mean of the FAR and FRR. All errors are computed based on the threshold that ensures equal error rates on the evaluation sets.

Image set	Error (%)	GR	HQ	GC	LT	HL
Evaluation set (ES)	FRR	9.3	8.2	10.2	11.3	8.8
	FAR	9.3	8.2	10.2	11.3	8.8
	HTER	9.3	8.2	10.2	11.3	8.8
Test set (TS)	FRR	6.5	4.8	7.3	7.8	6.3
	FAR	9.7	8.4	11.3	12.5	9.2
	HTER	8.1	6.6	9.3	10.2	7.8

Table 2: The verification performance of the tested image enhancement techniques on the XM2VTS database

Similar to the results presented in Table 1, we can see that only the HQ and HL techniques ensured better performance than the unprocessed grey-scale images, while the GC and LT did not improve upon the baseline error rates obtained with the unprocessed images.

It was suggested by Short et al. in (Short et al., 2004) that HQ should be used with photometric normalization techniques to further enhance their performance. However, based on the presented results, we can conclude that mapping a lognormal distribution to the pixels intensity distribution of the images might be useful as well. This issue will be further investigated in the remainder of this chapter.

5.3 Assessing the photometric normalization techniques

Our second series of face recognition experiments aimed at evaluating the performance of the photometric normalization techniques presented in Section 4. In Tables 3 and 4, where the results of the assessment are presented, the following abbreviations are used for the tested techniques: SR for the single scale retinex, MR for the multiscale retinex, SRA for the single scale retinex with adaptive smoothing, HO for the homomorphic filtering, SQ for the self quotient image, DCT for the discrete cosine transformation based normalization, WA for the wavelet based normalization, WD for the wavelet denoising, IS for the isotropic diffusion, AN for the anisotropic diffusion and LTV for the logarithmic total variation model.

Sets	SR	MR	SRA	HO	SQ	DCT	WA	WD	IS	AN	LTV
S2 (120)	100	100	100	100	100	100	100	100	100	94.2	100
S3 (120)	99.2	94.2	100	100	100	95.0	100	100	94.2	97.5	100
S4 (140)	82.9	71.4	98.6	84.3	98.6	59.3	55.0	98.6	84.3	80.1	99.3
S5 (190)	81.1	66.8	99.5	81.1	100	42.6	52.1	99.5	76.8	87.4	99.5

Table 3: Rank one recognition rates obtained with the tested photometric normalization techniques

Sets	Err. (%)	SR	MR	SRA	HO	SQ	DCT	WA	WD	IS	AN	LTV
ES	FRR	12.7	13.0	10.0	18.7	11.8	12.0	8.3	20.8	15.8	21.3	17.8
	FAR	12.7	13.0	10.0	18.7	11.8	12.0	8.3	20.8	15.8	21.3	17.8
	HTER	12.7	13.0	10.0	18.7	11.8	12.0	8.3	20.8	15.8	21.3	17.8
TS	FRR	8.0	8.0	7.0	12.3	7.0	9.3	4.8	13.0	7.8	19.3	11.8
	FAR	13.3	13.6	10.1	18.2	12.2	13.2	8.5	20.5	15.7	21.4	17.3
	HTER	10.7	10.8	8.6	15.3	9.6	11.3	6.7	16.8	11.7	20.4	14.6

Table 4: The verification performance of the tested photometric normalization techniques on the XM2VTS database

The results of the experiments suggest that while the majority of photometric normalization techniques ensure significant improvements upon the baseline performance of the unprocessed images on the YaleB database, they result in a deterioration (in most cases) in performance on images captured in controlled

conditions, i.e., on the XM2VTS database. Such a result can be linked to the fact that the photometric normalization techniques remove the low frequency information, which is susceptible to illumination changes, even though it is important for the recognition task.

Considering the performance of the tested techniques on both databases, we can conclude that the SRA and SQ techniques resulted in the best performance with nearly 100% rank one recognition rates on the YaleB database and the HTER almost identical to that obtained with unprocessed grey-scale images on the XM2VTS database.

It has to be noted, however, that the performance of all tested techniques heavily depends on their implementation, and even more on the proper selection of their parameters. Since there is no theoretical foundation for determining these parameters, they have to be set empirically. Having said that, we can quickly notice that differently from the results obtained by other researchers (Heusch et al., 2005, Short et al., 2004), our implementation of the diffusion based techniques IS and AN did not result in a competitive performance. This can mainly be linked to the selection of the parameter λ (see Section 4.10 for more details) which would need to be adjusted for each image separately to achieve optimal performance.

Based on the presented results, we can find that pure photometric normalization techniques without any pre- or post-processing of facial images cannot ensure satisfactory recognition results on images captured in controlled as well as varying illumination conditions. To this end, we assess a number of options to further improve their performance in the next section.

5.4 Combining image enhancement and photometric normalization techniques

Several empirical research studies have suggested that histogram equalization successfully improves the performance of a number of photometric normalization techniques. However, no specifics as to whether the histogram should be equalized before or after the actual deployment of the photometric normalization procedure were given, nor was a justification of why histogram equalization improves the performance of the photometric normalization techniques ever suggested.

If we recall that the photometric normalization techniques perform a compression of the dynamic range of the intensity values of facial images and thus reduce the variability in the image, it becomes evident that histogram equalization applied after the photometric normalization increases the dynamic range of the normalized image by improving the image's contrast and redistributing the pixel intensities equally across the whole 8-bit dynamic range. This procedure adds to images variability needed to successfully discriminate between different subjects. As histogram equalization is notorious for its background noise enhancing property, applying

histogram equalization prior to the photometric normalization would result in the retention (or worse - emphasis) of the background noise (which is a high frequency phenomenon) in the photometrically normalized images. Clearly, the right way to go is to use histogram equalization as a post-processing step for the photometric normalization, which is also consistent with the work presented in (Heusch et al., 2005).

Before we turn our attention to the experiments, let us examine the possibility of replacing the (post-processing) image enhancement technique HQ with HL, as proposed in Section 5.2. We have suggested that HQ improves upon the performance achieved with "pure" photometric normalization techniques because of its capability to increase the image's dynamic range. For a technique to be suitable to take the place of HQ, it should exhibit similar characteristics as HQ. As we have shown during the introduction of the histogram remapping techniques, HL remaps the histogram of the input image to approximate the lognormal distribution with a selected mean and variance, thus increasing the dynamic range of the processed images to the full 8-bit interval. Furthermore, as it does not force the target distribution to be uniform, it does not increase the background noise. Without a doubt HQ and HL can both be adopted as post-processing techniques to photometric normalization.

Using HQ and HL for post-processing, we perform two types of experiments. The results of these experiments are presented in Tables 5 and 6 for the YaleB database and Tables 7 and 8 for the XMVTS database.

Sets	SR	MR	SRA	HO	SQ	DCT	WA	WD	IS	AN	LTV
S2 (120)	100	100	100	100	100	100	100	100	100	100	100
S3 (120)	99.2	98.3	99.2	100	100	95.8	100	100	100	100	100
S4 (140)	91.4	87.1	97.1	88.6	93.6	72.9	55.7	99.3	86.4	95.7	99.3
S5 (190)	98.4	91.1	100	99.5	100	77.4	57.9	100	98.4	93.2	100

Table 5: Rank one recognition rates on all image sets of the YaleB database using histogram equalization as the post-processing step to the normalization

Sets	SR	MR	SRA	HO	SQ	DCT	WA	WD	IS	AN	LTV
S2 (120)	100	100	100	100	100	100	100	100	100	100	100
S3 (120)	100	100	100	100	100	97.5	99.2	100	100	100	100
S4 (140)	92.1	91.4	97.9	100	96.4	75.7	75.0	99.3	91.4	97.1	100
S5 (190)	97.4	96.3	99.5	100	100	83.2	85.3	100	95.3	100	100

Table 6: Rank one recognition rates on all image sets of the YaleB database using remapping of the histogram to a lognormal distribution as the post-processing step to the normalization

Sets	Err. (%)	SR	MR	SRA	HO	SQ	DCT	WA	WD	IS	AN	LTV
ES	FRR	9.2	9.0	7.0	13.2	8.2	8.2	8.00	17.3	12.3	12.5	12.2
	FAR	9.2	9.0	7.0	13.2	8.2	8.2	8.00	17.3	12.3	12.5	12.2
	HTER	9.2	9.0	7.0	13.2	8.2	8.2	8.00	17.3	12.3	12.5	12.2
TS	FRR	5.8	6.3	6.0	7.3	5.5	4.3	4.5	8.8	6.5	10.0	7.5
	FAR	8.9	8.9	7.0	13.3	8.4	8.4	8.3	17.0	12.2	12.4	11.8
	HTER	7.4	7.6	6.5	10.3	7.0	6.4	6.4	12.9	9.4	11.2	9.7

Table 7: The verification performance on the XM2VTS database using histogram equalization as the post-processing step to the normalization

Sets	Err. (%)	SR	MR	SRA	HO	SQ	DCT	WA	WD	IS	AN	LTV
ES	FRR	11.3	11.3	8.2	13.7	10.2	10.0	8.8	16.8	14.2	15.7	15.3
	FAR	11.3	11.3	8.2	13.7	10.2	10.0	8.8	16.8	14.2	15.7	15.3
	HTER	11.3	11.3	8.2	13.7	10.2	10.0	8.8	16.8	14.2	15.7	15.3
TS	FRR	6.5	6.8	6.8	9.8	6.0	6.8	6.5	10.3	8.3	11.0	8.8
	FAR	11.4	11.5	8.3	13.8	10.1	10.4	8.9	16.2	14.1	15.7	14.9
	HTER	9.0	9.2	7.6	11.8	8.1	8.6	7.7	13.3	11.2	13.4	11.9

Table 8: The verification performance on the XM2VTS database remapping of the histogram to a lognormal distribution as the post-processing step to the normalization

The presented results support our suggestion that the increase of the dynamic range of the processed images governs the improvements in the recognition rates, as both HQ and HL improve upon the vast majority of the results presented in Tables 3 and 4. Furthermore, we notice that HL ensures better performance than HQ on the YaleB database, while HQ outperforms HL on the XM2VTS database. While the relative ranking of the assessed techniques remained similar to the one presented in the previous series of face recognition experiments, the differences in the rank one recognition rates and half total error rates are smaller in absolute terms. This again suggests that decent performance gains can be achieved when photometric normalization techniques are combined with image enhancement approaches and, furthermore, that photometric normalization techniques need to be combined with image enhancement methods to ensure proper performance on both images captured in controlled as well as uncontrolled conditions.

6. Conclusion

The chapter presented an overview of the most popular photometric normalization techniques used to achieve illumination invariance for robust face recognition. A number of techniques were presented and later on assessed on the YaleB and XM2VTS databases. The experimental results suggest that several photometric normalization techniques are capable of achieving state-of-the-art recognition results and that the performance of a vast majority of these techniques can be further improved when combined with image enhancement techniques, such as histogram equalization or histogram remapping.

It has to be noted that many of the photometric normalization techniques result in an extreme compression of the dynamic range of the intensities values of images and that a small dynamic range usually implies that the photometric normalization procedure has removed most of the variability from the image, albeit induced by illumination or some other factor, and that the pixel intensity distribution thus exhibits a strong peak around a specific pixel value. In such cases, recognition algorithms cannot model the variations of faces due to intrinsic factors, such as facial expression or pose, even though shadows are removed perfectly (Park et al., 2008).

Future research with respect to photometric normalization techniques will, therefore, undoubtedly focus on the development of pre-processing techniques capable of efficiently removing the influence of illumination changes present during image acquisition, while still preserving the image's dynamic range. These techniques will be deployable on larger scale databases and will provide illumination invariance without the loss of subject specific information contained in the low-frequency part of the images.

To promote the future development of photometric normalization techniques we make most of the Matlab source code used in our experiments freely available. To obtain a copy of the code, the reader should follow the links given at: http://luks.fe.uni-lj.si/en/staff/vitomir/index.html.

References

Adini, Y., Moses, Y., and Ullman, S. (1997). Face Recognition: The Problem of Compensating for Changes in Illumination Direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19 (7)*, 721-732.

Alizadeh, F., and Goldfarb, D. (2003). Second-Order Cone Programming. *Mathematical Programming*, *95(1)*, 3-51.

Basri, R., and Jacobs, D. (2003). Lambertian Reflectance and Linear Subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25(2)*, 218-233.

Chen, H., Belhumeur, P., and Jacobs, D. (2000). In Search of Illumination Invariants. In: *Proc. of the International Conference on Computer Vision and Pattern Recognition (ICVPR'00),* pp. 254-261.

Chen, T.F., and Esedoglu, S. (2004). Aspects of Total Variation Regularized L^1 Function Approximation. *CAM Report 04-07*, Univ. of California, Los Angeles.

Chen, T., Yin, W., Zhou, X.S., Comaniciu, D., and Huang, T.S. (2006). Total Variation Models for Variable Lighting Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(9)*, 1519-1524.

Chen, W., Er, M.J., and Wu, S. (2006). Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithmic Domain. *IEEE Transactions on Systems, Man and Cybernetics – part B*, *36(2)*, 458-466.

Delac, K., Grgic, M. and Kos, T. (2006). Sub-Image Homomorphic Filtering Technique for Improving Facial Identification under Difficult Illumination Conditions. In: *Proc. of the International Conference on Systems, Signals and Image Processing (IWSSIP'06)*, pp. 95-98.

Du, S., and Ward, R. (2005). Wavelet-based Illumination Normalization for Face Recognition. In: *Proc. of the IEEE International Conference on Image Processing (ICIP'05)*.

Gao, Y., and Leung, M.K.H. (2002). Face Recognition Using Line Edge Map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24(6)*, 764-779.

Georghiades, A.G., Belhumeur, P.N., and Kriegman, D.J. (2001). From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23(6)*, 643-660.

Gonzalez, R.C., and Woods, R.E. (2002). Digital Image Processing, 2nd Edition, Prentice Hall.

Gross, R., and Brajovic V. (2003). An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. In: *Proc. of the 4th International Conference on Audio- and Video-Based Biometric Personal Authentication (AVPBA'03), pp.* 10-18.

Gross, R., Baker, S., Matthews, I., and Kanade T. (2004). Face Recognition Across Pose and Illumination. In: Li, S.Z., and Jain A.K. (Eds.): *Handbook of Face Recognition*. Spriger-Verlag.

Heusch, G., Cardinaux, F., and Marcel S. (2005, March). Lighting Normalization Algorithms for Face Verification. *IDIAP-com 05-03*.

Jobson, D.J., Rahman, Z., and Woodell, G.A. (1997a). Properties and Performance of a Center/Surround Retinex. IEEE Transactions on Image Processing, 6(3), 451-462.

Jabson, D.J., Rahmann, Z., and Woodell, G.A. (1997b). A Multiscale Retinex for Bridging the Gap Between Color Images and the human Observations of Scenes. *IEEE Transactions on Image Processing, 6(7),* 897-1056.

Land, E.H., and McCann, J.J. (1971). Lightness and Retinex Theory. *Journal of the Optical Society of America*, *61(1)*, 1-11.

Liu, C. (2006). Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28(5)*, 725-737.

Marcel, S., Rodriguez, Y., and Heusch, G. (2007). On the Recent Use of Local Binary Patterns for Face Authentication. *International Journal of Image and Video Processing*, Special Issue on Facial Image Processing.

Messer, K., Matas, J., Kittler, J., Luettin, J., and Maitre, G. (1999). XM2VTSDB: the extended M2VTS database*. In: Proc. of the International Conference on Audio- and Video-Based Biometric Personal Authentication (AVBPA'99)*, pp. 72-77.

Moore, A., Allman, J., and Goodman, R.M. (1991). A Real Time Neural System for Color Consistency. *IEEE Transactions on Neural Networks, 2*, 237-247.

Park, Y.K., Park, S.L., and Kim, J.K. (2008). Retinex Method Based on Adaptive smoothing for Illumination Invariant Face Recognition. *Signal Processing, 88(8)*, 1929-1945.

Pelecanos, J., and Sridharam, S. (2001). Feature Warping for Robust Speaker Verification. In: *Proc. of the Speaker Recognition Workshop Odyssey*, pp. 213–-218.

Phillips, P. J., Scruggs, W. T., O'Toole, A. J. , Flynn, P. J. , Bowyer, K. W. , Schott, C. L., and Sharpe M. (2007, March). FRVT 2006 and ICE 2006 Large-Scale Results. *NISTIR 7408*.

Shashua, A., and Riklin-Raviv, T. (2001). The Quotient Image: Class Based Re-rendering and Recognition with Varying Illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2)*, 129-139.

Short, J., Kittler, J., and Messer, K. (2004). A Comparison of Photometric Normalization Algorithms for Face Verification. In: *Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition (AFGR'04)*, pp. 254- 259.

Štruc, V. and Pavešić, N. (2009). Gabor-Based Kernel Partial-Least-Squares Discrimination Features for Face Recognition. *Informatica, 20(1)*, 115-138.

Štruc, V., Žibert, J., and Pavešić, N. (2009). *Histogram Remapping as a Preprocessing Step for Robust Face Recognition*. *WSEAS Transactions on Information Science and Applications, 6(3)*, 520-529.

Wang, H., Li, S.Z., Wang, Y., and Zhang, J. (2004). Self Quotient Image for Face Recognition. In: *Proc. of the International Conference on Pattern Recognition (ICPR'04),* pp. 1397- 1400.

Wei, S.D., and Lai, S.H. (2004). Robust Face Recognition under Lighting Variations. In: *Proc. of the 17ᵗʰ International Conference on Pattern Recognition (ICPR'04)*, pp. 354-357.

Zhang, T., Fang, B., Yuan, Y., Tang, Y.Y., Shang, Z., Li, D., and Lang, F. (2009). Multiscale Facial Structure Representation for Face Recognition Under Varying Illumination. *Pattern Recognition, 42(2)*, 252-258.

Zhou, S.K., and Chellapa, R. (2004). Illumination Light Fields: Image-based Face Recognition across Illumination and Pose. In: *Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition (AFGR'04)*, pp. 229–234.

Zou, X., Kittler, J., and Messer, K. (2007). Illumination Invariant Face Recognition: A Survey. In: *Proc. of the first IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS'07)*, pp. 1-8.