Illumination Invariant Face Recognition by Non-Local Smoothing

Vitomir Štruc and Nikola Pavešić

Faculty of Electrical Engineering,
University of Ljubljana,
Tržaška 25, SI-1000 Ljubljana, Slovenia
{vitomir.struc,nikola.pavesic}@fe.uni-lj.si.com
http://luks.fe.uni-lj.si/

Abstract. Existing face recognition techniques struggle with their performance when identities have to be determined (recognized) based on image data captured under challenging illumination conditions. To overcome the susceptibility of the existing techniques to illumination variations numerous normalization techniques have been proposed in the literature. These normalization techniques, however, still exhibit some shortcomings and, thus, offer room for improvement. In this paper we identify the most important weaknesses of the commonly adopted illumination normalization techniques and presents two novel approaches which make use of the recently proposed non-local means algorithm. We assess the performance of the proposed techniques on the YaleB face database and report preliminary results.

Key words: Face recognition, retinex theory, non-local means, illumination invariance

1 Introduction

The performance of current face recognition technology with image data captured in controlled conditions has reached a level which allows for its deployment in a wide variety of applications. These applications typically ensure controlled conditions for the image acquisition procedure and, hence, minimize the variability in the appearance of different (facial) images of a given individual. However, when employed on facial images captured in uncontrolled and unconstrained environments the majority of existing face recognition techniques still exhibits a significant drop in their recognition performance.

The reason for the deterioration in the recognition (or verification) rates can be found in the appearance variations induced by various environmental factors, among which illumination is undoubtedly one of the most important. The importance of illumination was highlighted in several empirical studies where it was shown that the illumination induced variability in facial images is often larger than the variability induced to the facial images by the individual's identity [1]. Due to this susceptibility, numerous techniques have been proposed in the literature to cope with the problem of illumination. These techniques try to tackle the illumination induced appearance variations at one of the following three levels: *(i)* at the pre-processing level, *(ii)* at the feature extraction level, and *(iii)* at the modeling or/and classification level.

While techniques from the latter two levels represent valid efforts in solving the problem of illumination invariant face recognition, techniques operating at the pre-processing level exhibit some important advantages which make them a preferred choice when devising robust face recognition systems. One of their most essential advantages lies in the fact that they make no assumptions regarding the size and characteristics of the training set while offering a computationally simple and simultaneously effective way of achieving illumination invariant face recognition.

Examples of normalization techniques operating at the pre-processing level[1] include the single and multi scale retinex algorithms [2],[3], the self quotient image [4], anisotropic smoothing [5], etc. All of these techniques share a common theoretical foundation and exhibit some strengths as well as some weaknesses. In this paper we identify (in our opinion) the most important weaknesses of the existing normalization techniques and propose two novel techniques which try to overcome them. We assess the proposed techniques on the YaleB database and present encouraging preliminary results.

The rest of the paper is organized as follows. In Section 2 the theory underlying the majority of photometric normalization techniques is briefly reviewed and some weakness of existing techniques are pointed out. The novel normalization techniques are presented in Section 3 and experimentally evaluated in Section 4. The paper concludes with some final comments in Section 5.

2 Background and Related Work

The theoretical foundation of the majority of existing photometric normalization techniques can be linked to the Retinex theory developed and presented by Land and McCann in [6]. The theory tries to explain the basic principles governing the process of image formation and/or scene perception and states that an image $I(x,y)$ can be modeled as the product of the reflectance $R(x,y)$ and luminance $L(x,y)$ functions:

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

Here, the reflectance $R(x,y)$ relates to the characteristics of the objects comprising the scene of an image and is dependant on the reflectivity (or albedo) of the scenes surfaces [7], while the luminance $L(x,y)$ is determined by the illumination source and relates to the amount of illumination falling on the observed scene.

Since the reflectance $R(x,y)$ relates solely to the objects in an image, it is obvious that (when successfully estimated) it acts as an illumination invariant representation of the input image. Unfortunately, estimating the reflectance from

[1] We will refer to these techniques as photometric normalization techniques in the remainder of this paper.

the expression defined by (1) represents an ill-posed problem, i.e., it is impossible to compute the reflectance unless some assumptions regarding the nature of the illumination induced appearance variations are made. To this end, researchers introduced various assumptions regarding the luminance and reflectance functions, the most common, however, are that the luminance part of the model in (1) varies slowly with the spatial position and, hence, represents a low-frequency phenomenon, while the reflectance part represents a high-frequency phenomenon.

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 via the manipulation of the image model given by the expression (1), i.e.:

$$\ln R(x, y) = \ln I(x, y) - \ln L(x, y) \quad \text{or} \quad R(x, y) = I(x, y)/L(x, y), \qquad (2)$$

where the right hand side equation of (2) denotes an element-wise division of the input image $I(x, y)$ with the estimated luminance $L(x, y)$. We will refer to the reflectance computed with the left hand side equation of (2) as the logarithmic reflectance and to the reflectance computed with the right hand side equation of (2) as the quotient reflectance in the rest of this paper.

As already emphasized, the luminance is considered to vary slowly with the spatial position [8] and 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 photometric normalization procedures that were successfully applied to the problem of face recognition under severe illumination changes.

The single scale retinex algorithm [2], for example, computes the estimate of the luminance function $L(x, y)$ by simply smoothing the input image $I(x, y)$ with a Gaussian smoothing filter. The illumination invariant image representation is then computed using the expression for the logarithmic reflectance. While such an approach generally produces good results with a properly selected Gaussian, its broader use in robust face recognition systems is still limited by an important weakness: at large illumination discontinuities caused by strong shadows that are casted over the face halo effects are often visible in the computed reflectance [8]. To avoid this problem the authors of the algorithm extended their normalization technique to a multi scale form [3], where Gaussians with different widths are used and basically outputs of different implementations of the single scale retinex algorithm are combined to compute the final illumination invariant face representation.

Another solution to the problem of halo effects was presented by Wang et al. [4] in form of the self quotient image technique. Here, the authors approach the problem of luminance estimation by introducing an anisotropic smoothing filter. Once the anisotropic smoothing operation produces an estimate of the luminance $L(x, y)$, the quotient reflectance $R(x, y)$ is computed in accordance with the right hand side equation of (2). However, due to the anisotropic nature of the employed smoothing filter flat zones in the images are not smoothed properly.

Gross and Brajovic [5] presented a solution to the problem of reliable luminance estimation by adopting an anisotropic diffusion based smoothing technique. In their method the amount of smoothing at each pixel location is controlled by the images local contrast. Adopting the local contrast as means to control the smoothing process results in flat image regions being smoothed properly while still preserving image edges and, thus, avoiding halo effects. Despite the success of the normalization technique in effectively determining the quotient reflectance, one could still voice some misgivings. An known issue with anisotropic diffusion based smoothing is that it smoothes the image only in the direction orthogonal to the images gradient [9]. Thus, it effectively preserves only straight edges, but struggles at edge points with high curvature (e.g., at corners). In these situations an approach that better preserves edges would be preferable. To this end, we present in the next section two novel algorithms which make use of the recently proposed non-local means algorithm.

3 Non-Local Means for Luminance Estimation

3.1 The Non-Local Means Algorithm

The non-local means (NL means) algorithm [9] is a recently proposed image denoising technique, which, unlike existing denoising methods, considers pixel values from the entire image for the task of noise reduction. The algorithm is based on the fact that for every small window of the image several similar windows can be found in the image as well, and, moreover, that all of these windows can be exploited to denoise the image.

Let us denote an image contaminated with noise as $I_n(\mathbf{x}) \in \mathcal{R}^{a\times b}$, where a and b are image dimensions in pixels, and let $\mathbf{x}$ stand for an arbitrary pixel location $\mathbf{x} = (x, y)$ within the noisy image. The NL means algorithm constructs the denoised image $I_d(\mathbf{x})$ by computing each pixel value of $I_d(\mathbf{x})$ as a weighted average of pixels comprising $I_n(\mathbf{x})$, i.e. [9]:

$$I_d(\mathbf{x}) = \sum_{\mathbf{x}\in I_n(\mathbf{x})} w(\mathbf{z},\mathbf{x}) I_n(\mathbf{x}), \tag{3}$$

where $w(\mathbf{z},\mathbf{x})$ represents the weighting function that measures the similarity between the local neighborhoods of the pixel at the spatial locations $\mathbf{z}$ and $\mathbf{x}$. Here, the weighting function is defined as follows:

$$w(\mathbf{z},\mathbf{x}) = \frac{1}{Z(\mathbf{z})}\, e^{-\frac{G_\sigma \| I_n(\Omega_\mathbf{x}) - I_n(\Omega_\mathbf{z})\|_2^2}{h^2}} \text{ and } Z(\mathbf{z}) = \sum_{\mathbf{x}\in I_n(\mathbf{x})} e^{-\frac{G_\sigma \| I_n(\Omega_\mathbf{x}) - I_n(\Omega_\mathbf{z})\|_2^2}{h^2}}. \tag{4}$$

In the above expressions G_σ denotes a Gaussian kernel with the standard deviation σ, $\Omega_\mathbf{x}$ and $\Omega_\mathbf{z}$ denote the local neighborhoods of the pixels at the locations $\mathbf{x}$ and $\mathbf{z}$, respectively, h stands for the parameter that controls the decay of the exponential function, and $Z(\mathbf{z})$ represents a normalizing factor.

From the presented equations it is clear that if the local neighborhoods of a given pair of pixel locations $\mathbf{z}$ and $\mathbf{x}$ display a high degree of similarity, the pixels at $\mathbf{z}$ and $\mathbf{x}$ will be assigned relatively large weights when computing their denoised estimates. Some examples of image windows used by the algorithm are presented in Fig. 1. Here, similar image windows are marked white, while

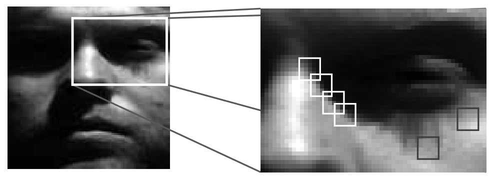

Fig. 1. The principle of the NL means algorithm: an input image (left), similar and dissimilar image windows (right).

dissimilar image windows are marked black. When computing the denoised value of the center pixel of each of the white windowed image regions, center pixels of the similar windows will be assigned relatively large weights, the center pixels of the dissimilar windows, on the other hand, will be assigned relatively low weights.

Whit a proper selection of the decay parameter h, the presented algorithm results in a smoothed image whit preserved edges. Hence, it can be used to estimate the luminance of an input image and, consequently, to compute the (logarithmic) reflectance. An example of the deployment of the NL means algorithm (for a 5×5 local neighborhood and $h = 10$) for estimation of the logarithmic reflectance is shown in Fig. 2 (left triplet).

Fig. 2. Two sample images processed with the NL means (left triplet) and adaptive NL means (right triplet) algorithms. Order of images in each triplet (from left to right): the input image, the estimated luminance, the logarithmic reflectance.

3.2 The Adaptive Non-Local Means Algorithm

The NL means algorithm assigns different weights to each of the pixel values in the noisy image $I_n(\mathbf{x})$ when estimating the denoised image $I_d(\mathbf{x})$. As we have shown in the previous section, this weight assignment is based on the similarity of the local neighborhoods of arbitrary pixel pairs and is controlled by the decay parameter h. Large values of h result in a slow decay of the Gaussian weighted Euclidian distance[2] and, hence, more neighborhoods are considered similar and

[2] Recall that the Euclidian distance serves as the similarity measure between two local neighborhoods.

are assigned relatively large weights. Small values of h, on the other hand, result in a fast decay of the Euclidian similarity measure and consequently only a small number of pixels is assigned a large weight for the estimation of the denoised pixel values.

Rather than using the original NL means algorithm for estimation of the luminance of an image, we propose in this paper to exploit an adaptive version of the algorithm, where the decay parameter h is a function of local contrast and not a fixed and preselected value. At regions of low contrast, which represent homogeneous areas, the image should be smoothed more (i.e., more pixels should be considered for the estimation of the denoised pixel value), while in regions of high contrast the image should be smoothed less, (i.e., less pixels should be considered for the estimation of the denoised pixel value).

Following the work of Gross and Brajovic [5], we define the local contrast between neighboring pixel locations $\mathbf{a}$ and $\mathbf{b}$ as: $\rho_{\mathbf{a},\mathbf{b}} =\mid I_n(\mathbf{a}) - I_n(\mathbf{b}) \mid / \mid I_n(\mathbf{a}) + I_n(\mathbf{b}) \mid$. Assuming that $\mathbf{a}$ is an arbitrary pixel location within $I_n(\mathbf{x})$ and $\mathbf{b}$ stands for a neighboring pixel location above, below, left or right from $\mathbf{a}$, we can construct four contrast images encoding the local contrast in one of the possible four directions. The final contrast image $I_c(\mathbf{x})$ is ultimately computed as the average of the four (directional) contrast images. To link the decay parameter h to the contrast image we first compute the logarithm of the inverse of the (8-bit grey-scale) contrast image $I_{ic}(\mathbf{x}) = \log[\mathbf{1}/I_c(\mathbf{x})]$, where $\mathbf{1}$ denotes a matrix of all ones and the operator "/" stands for the element-wise division. Next, we linearly map the values of our inverted contrast image $I_{ic}(\mathbf{x})$ to values of the decay parameter h, which now becomes a function of the spatial location: $h(\mathbf{x}) = [(I_{ic}(\mathbf{x}) - I_{ic_{min}})/(I_{ic_{max}} - I_{ic_{min}})] * h_{max} + h_{min}$, where $I_{ic_{max}}$ and $I_{ic_{min}}$ denote the maximum and minimum value of the inverted contrast image $I_{ic}(\mathbf{x})$, respectively, and h_{max} and h_{min} stand for the target maximum and minimum values of the decay parameter h. An example of the deployment of the presented algorithm is shown in Fig. 2 (right triplet).

4 Experiments

To assess the presented two photometric normalization techniques we made use of the YaleB face database [10]. The database contains images of ten distinct subjects each photographed under 576 different viewing conditions (9 poses 64 illumination conditions). Thus, a total of 5760 images is featured in the database. However, as we are interested only in testing our photometric normalization techniques, we make use of a subset of 640 images with frontal pose in our experiments. We partition the 640 images into five image set according to the extremity in illumination under which they were taken and employ the first image set for training and the remaining ones for testing.

In the experiments we use principal component analysis as the feature extraction technique and the nearest neighbor (to the mean) classifier in conjunction with the cosine similarity measure as the classifier. The number of features is set to its maximal value in all experiments.

In our first series of recognition experiments we assess the performance of the NL means (NLM) and adaptive NL means (ANL) algorithms for varying values of their parameters, i.e., the decay parameter h for the NLM algorithm and h_{max} for the ANL algorithm. It has to be noted that the parameter h_{min} of the ANL algorithm was fixed at the value of $h_{min} = 0.01$ and the local neighborhood of 5×5 pixels was chosen for the NLM and ANL algorithms in all experiments. The results of the experiments in terms of the rank one recognition rates for the individual image sets as well as its average value over the entire database are presented in Table 1. We can see that the best performing implementations of

Table 1. The rank one recognition rates (in %) for the NLM and ANL algorithms.

Algorithm	ANL - parameter h_{max}					NLM - parameter h			
Parameter value	40	80	120	160	200	10	30	60	120
Image set no. 2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Image set no. 3	98.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Image set no. 4	90.7	94.3	94.3	92.1	92.9	91.4	95.0	97.1	95.7
Image set no. 5	87.9	97.4	92.6	84.7	82.1	96.3	99.5	92.6	85.3
Average	94.2	**97.9**	96.7	94.2	93.8	96.9	**98.6**	97.4	95.3

the NLM and ANL algorithm feature parameter values of $h = 30$ and $h_{max} = 80$, respectively.

In our second series of recognition experiments we compare the performance of the two proposed algorithms (for $h = 30$ and $h_{max} = 80$) and several popular photometric normalization techniques. Specifically, the following techniques were implemented for comparison: the logarithm transform (LN), histogram equalization (HQ), the single scale retinex (SR) technique and the adaptive retinex normalization approach (AR) presented in [8]. For baseline comparisons, experiments on unprocessed grey scale images (GR) are conducted as well. It should be noted that the presented recognition rates are only indicative of the general performance of the tested techniques, as the YaleB database represent a rather small database, where it is possible to easily devise a normalization technique that effectively discriminates among different images of the small number of subjects. Several techniques were presented in the literature that normalize the facial images by extremely compressing the dynamic range of the images, resulting in the suppression of most of the images variability, albeit induced by illumination or the subjects identity. The question of how to scale up these techniques for use with larger numbers of subjects, however, still remains unanswered. To get an impression of the scalability of the tested techniques we present also recognition rates obtained with the estimated logarithmic luminance functions (where applicable). These results provide an estimate of how much of the useful information was removed from the facial image during the normalization. For the experiments with the logarithmic luminance functions logarithm transformed images from the first image set were employed for training.

The presented results show the competitiveness of the proposed techniques. Similar to the best performing AR technique, they achieve an average recognition

Table 2. Comparison of the rank one recognition rates (in %) for various algorithms.

Representation	Normalized image					Log. luminance - $\ln L(x,y)$				
Image sets	No. 2	No. 3	No. 4	No. 5	Avg.	No. 2	No. 3	No. 4	No. 5	Avg.
GR	100.0	100.0	57.9	16.3	68.6	n/a	n/a	n/a	n/a	n/a
HQ	100.0	100.0	58.6	60.0	79.7	n/a	n/a	n/a	n/a	n/a
LN	100.0	98.3	58.6	52.6	77.4	n/a	n/a	n/a	n/a	n/a
SR	100.0	100.0	92.1	84.2	94.1	100.0	90.8	46.4	41.1	69.6
AR	100.0	100.0	97.1	98.4	**98.9**	100.0	95.0	49.3	44.3	72.1
NLM	100.0	100.0	95.0	99.5	98.6	100.0	86.7	39.3	26.3	63.1
ANL	100.0	100.0	94.3	97.4	97.9	100.0	65.8	36.4	26.8	**57.3**

rate of approximately 98%, but remove less of the useful information as shown by the results obtained on the luminance estimates. The results suggest that the proposed normalization techniques will perform well on larger databases as well.

5 Conclusion and Future Work

In this paper we have presented two novel image normalization techniques, which try to compensate for the illumination induced appearance variations of facial images at the preprocessing level. The feasibility of the presented techniques was successfully demonstrated on the YaleB database were encouraging results were achieved. Our future work with respect to the normalization techniques will be focused on their evaluation on larger and more challenging databases.

References

1. Heusch, G., Cardinaux, F., Marcel, S.: Lighting Normalization Algorithms for Face Verification. IDIAP-com 05-03, March (2005)
2. Jobson, D.J., Rahman, Z., Woodell, G.A.: Properties and Performance of a Center/Surround Retinex. IEEE Transactions on Image Processing 6(3), 451–462 (1997)
3. Jobson, D.J., Rahman, Z., Woodell, G.A.: 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 (1997)
4. Wang, H., Li, S.Z., Wang, Y., Zhang, J.: Self Quotient Image for Face Recognition. In: Proc. of the Int. Conference on Pattern Recognition, pp. 1397–1400 (2004)
5. Gross, R., Brajovic, V.: An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. In: Proc. of AVPBA'03, pp. 10–18 (2003).
6. Land, E.H., McCann, J.J.: Lightness and Retinex Theory. Journal of the Optical Society of America 61(1), 1–11 (1971)
7. Short, J., Kittler, J., Messer, K.: Photometric Normalisation for Face Verification, In: Proc. of AVBPA'05, pp. 617–626 (2005)
8. Park, Y.K., Park, S.L., Kim, J.K.: Retinex Method Based on Adaptive smoothing for Illumination Invariant Face Recognition. Signal Processing 88(8), 1929–1945 (2008)
9. Buades, A., Coll, B., Morel, J.M.: On Image Denoising Methods. Prepublication, avaiable from: http://www.cmla.ens-cachan.fr
10. Georghiades, A.G., Belhumeur, P.N., Kriegman, D.J.: From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. IEEE TPAMI 23(6), 643–660 (2001)