

Retinex Combined with Total Variation for Image Illumination Normalization

Luigi Cinque and Gabriele Morrone

Dipartimento di Informatica, “Sapienza” Università di Roma
Via Salaria 113 00198 Roma, Italy
{cinque,morrone}@di.uniroma1.it

Abstract. This paper presents a method for the normalization of human facial images in arbitrary illumination conditions. The enhanced image is suitable to be used as an input to a face recognition system.

1 Introduction

Illumination normalization is an important task in the field of computer vision and pattern recognition. Face recognition is one of its most important applications and its accuracy depends heavily on how well the input images have been compensated for illumination, indeed the differences caused by varying illumination are more significant than the inherent differences between individuals.

Various methods have been proposed for face recognition under varying illumination: the Illumination Cone methods [3], a simplified Lambertian model [11,13,14], shape from shading [12] and quotient image based approaches [15]. However the performances of most of them are still far from ideal and many of these methods require either knowledge of the light source or a large number of training data, which are not suitable for most real world scenarios.

This paper presents an effective approach for illumination normalization based on the Land’s “Retinex” model [1], which estimated the reflectance R as the ratio of the image I using a low pass filter to remove the effect of the illuminance L . To improve the effect of the Retinex (its goodness to remove shadows and specularities is already known [4]) in illumination normalization, we have combined it with the Total Variation Quotient Image model ($TVQI$)[6]. With this approach applied on standard database we have raised the accuracy of the recognition from 68,89% to 97,46% using only a single training image for each individual. Then, applying the histogram matching before the $TVQI$ and Retinex, we were able to realize the 100% accuracy under variable lighting conditions.

In subsequent sections methodology and experiments are outlined.

2 Methodology

The method proposed uses the Histogram matching to bring the intensity distribution of an image to a specific one of a target image, the $TVQI$ to remove noise keeping the edges in the faces, and the SSR to remove shadows and specularities. The next sections explain the methods one by one.

2.1 Single Scale Retinex

Retinex is a human based image-processing algorithm, which provides color constancy and dynamic range compression. The color in the recorded images is strongly influenced by spectral shifts in the entire spectrum.

An image $I(x, y)$ is regarded as a product of reflectance R and the illuminance effect L . Assuming L changes slowly compared to R , the Retinex try to extract R by filtering the source image. Many variants of the Retinex have been published over the years. The last version is now referred to as the Single Scale Retinex (SSR) [2] is defined for a point (x, y) in an image as:

$$R_i(x, y) = \log I_i(x, y) - \log[F(x, y) \otimes I_i(x, y)] \quad (1)$$

where $R_i(x, y)$ is the Retinex output, $I_i(x, y)$ is the image distribution in the i spectral band, the symbol \otimes represents the convolution operator.

$F(x, y)$ is the Gaussian surround function given by Equation (2):

$$F(x, y) = Ke^{-(x^2+y^2)/\sigma^2} \quad (2)$$

where σ is the Gaussian surround constant, that represent the standard deviation and that is referred to as the scale of the SSR.

A small value provides very good dynamic range compression but at the cost of poorer color rendition, causing greying of the image in uniform areas of color. Conversely, a large scale provides better color rendition but at the cost of dynamic range compression [2].

Since face recognition is performed on grey-scale images we can ignore the greying effect, moreover we observed that the best illumination normalization is gained by small scales. Figure 1 illustrates the effect of Retinex processing on a facial image, I , for different values of σ

2.2 Total Variation Quotient Image

In order to improve the accuracy of the recognition, we applied the Total Variation Quotient Image before the SSR. As formalized in [6], an image f is modelled as the sum of large-scale output u and small-scale output v , where f , u and v are

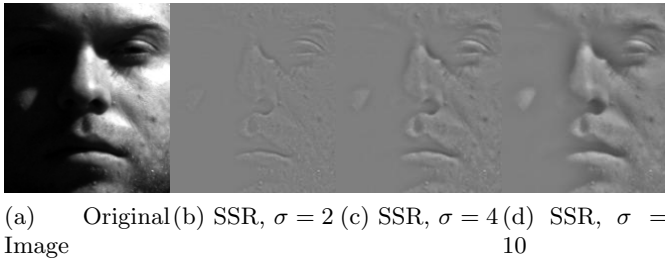


Fig. 1. SSR application



Fig. 2. Sample TV images

defined as functions in \mathbb{R}^2 . u contains background hues and important boundaries as sharp edges. The rest of the image, which is texture, is characterized by small-scale patterns. Since the level sets of large-scale signal u have simpler boundaries than those of small-scale signal v , we can obtain u from f by solving a variational regularization problem:

$$\min \int_{\Omega} |\nabla u| + \lambda \|t(u, f)\|_B \tag{3}$$

where $\int_{\Omega} |\nabla u|$ is the total variation of u over its support Ω , $\|t(u, f)\|_B$ is some measure of the closeness between u and f , and λ is a scalar weight parameter. The choice of the measure $\|\cdot\|_B$ depends on applications. The first use of this model was for image denoising [7] with the measure $\|t(u, f)\|_B = \|f - u\|_{L^2}$. The L^2 term $\|f - u\|_{L^2}$ penalizes big $f(x) - u(x)$ values much more than small $f(x) - u(x)$ values, so the $TV+L^2$ model allows most small point-wise values (like most noise) in $f - u$. The use instead of the L^1 measure $\|f - u\|_{L^1}$ penalizes the difference between f and u in a linear way. The L^1 term lets $f - u$ contain nearly all the signals with scale $\leq 1/\lambda$ w.r.t. G-norm and with their original amplification. So the $TV+L^1$ model with an appropriate λ can successfully extract small-scale signals like the edges, lines, and corners of facial features from a face image, but it has some limitation if it is directly applied for normalizing illumination. The signal of the intrinsic structures in the image are not normalized, so before we use the SSR we applied the total variation based quotient image (TVQI) model. The intuition of the TVQI model is the Lambertian surface model. According to the Lambertian model, the intensity at the (x, y) position of an image f is defined as:

$$f_{x,y} = A_{x,y} \rho_{x,y} \cos \theta_{x,y} \tag{4}$$

where A is the strength of the light source, ρ , the albedo (texture) of the image, is the surface reflectance associated with each point in the image, and θ is the angle between the surface normal and the light source. The intensity of image point (x, y) is proportional to the strength A of the received light at this point. Therefore, the intensity and the variance of the small-scale signals are proportional to the intensity of the background in their vicinities. For recognition purpose,

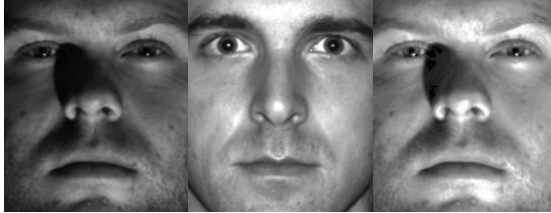
we approximate the normalized image by $v'_{x,y} = f_{x,y}/u_{x,y}$ for every point (x, y) using the output u of the TV+ L^1 model. So the TVQI model result as:

$$TVQI = \frac{f}{u} \quad (5)$$

To solve the variational regularization problem of TV we adopted the method proposed by Chambolle [8] using the L^1 term instead of L^2 and we chose λ based on the dimension of the face image [6]. See fig. 2 to observe the result of TV+ L^1 with $\lambda = 0.4$ and the sequent TVQI image.

2.3 Histogram Matching

It's useful for the process of recognition to normalize a poorly illuminated image via histogram fitting to a similar, well-illuminated image. The reference histogram is taken from a frontal illuminated facial image: the histogram of each image to enhance will fit to the one of reference(see fig. 3).



(a) Original im- (b) Reference im- (c) Image after
age age the histogram
matching

Fig. 3. Histogram matching

By defining the histogram of a digital image with levels of intensity in $[0, L-1]$ as a function $h(r_k)$:

$$h(r_k) = n_k \quad (6)$$

with r_k the k intensity value and n_k the number of pixel of the image with intensity r_k , and the normalized histogram as:

$$p(r_k) = \frac{r_k}{MN} \quad (7)$$

with MN the product of rows number and columns number, the procedure is as follows:

1. Get the histogram $p_r(r)$ from source image, and use it to find the function $T(r_k)$

$$s_k = T(r_k) = (L-1) \sum_{j=0}^k p_r(r_j) \quad (8)$$

2. Calculate the function G

$$G(z_q) = (L - 1) \sum_{i=0}^q p_z(z_i) \tag{9}$$

where $p_z(z_i)$ are the values of the reference histogram

3. Calculate the inverse function

$$z_q = G^{-1}(s_k) \tag{10}$$

4. With functions $T(r_k)$ and $G^{-1}(s_k)$, map all the values r_k on z_q

$$z_q = G^{-1}(T(r_k)) \tag{11}$$

5. Update the pixel of the image, changing the n_k pixel from value r_k to value z_q with k and q in $[0, L - 1]$

Figure 4 shows the application of the procedure on the image histogram of fig. 3 (a): center histogram come from image in fig. 3 (b), the rightmost is the result of the matching.

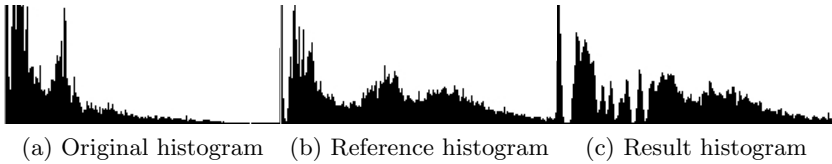


Fig. 4. Histogram matching: the histograms

3 Experiments

The feasibility and the validity of the proposed scheme is confirmed by the experimental results in this section. We have tested our algorithms using the extended Yale face database B cropped [9]. Five subjects, each with 64 different illumination frontal face images, have been chosen. We picked the frontal illumination images, one for each subject, as training set, we left the others for testing set (315 images, 192x168 pixel).

Support Vector Machines (SVM) were used for the face recognition experiments. An SVM with a linear kernel was trained for each set of experiments using the libSVM [10].

In order to normalize all the images in the database, three different experiments have been performed. First we have tested the SSR without any other enhancement to find the best value for the parameter σ obtaining the best accuracy in the face recognition (68,89%) with $\sigma = 4$ (see table 1). In the second experiment, we have normalized the database with the TVQI, $\lambda = 0.4$, and SSR with $\sigma = 4$. In this case the accuracy was 97,46% (only 8 images misclassified).

Finally, we applied all the proposed illumination correction methods (Histogram matching, TVQI and SSR sequencely), obtaining an accuracy of 100% . Performed test with relative accuracy are summarized in table 2.

Table 1. SSR results varying σ

SSR				
σ	ok	%	err	%
1	215	68,25	100	31,75
2	215	68,25	100	31,75
3	215	68,25	100	31,75
4	217	68,89	98	31,11
5	213	67,62	102	32,38
6	213	67,62	102	32,38
7	212	67,30	103	32,70
8	210	66,67	105	33,33
9	206	65,39	109	34,61
10	205	65,08	110	34,92

In fig. 5 are presented the results of applying the different methods, either in isolation (TVQI fig. 5 (b) and (e)), or together (fig. 5 (c) and (f)).



Fig. 5. Sample images

Table 2. Average recognition Rate (%) comparison

Methodology	SSR	TVQI+SSR	Hist. match.+TVQI+SSR
Average	68,88	97,46	100

4 Conclusion

In this paper we have proposed a new methodology to normalize the facial images in order to improve the performance of a face recognition system. With the Histogram Matching we were able to compensate the difference in the intensity distribution between a poorly illuminated image and a well-illuminated one. The TVQI has enhanced the edges that are on the face and has removed most of the noise. At the end, with the SSR we have removed almost all the shadows in the image obtaining also a good dynamic compression. With the combined use of these three methods we were able to realize the 100% accuracy under variable lighting conditions in the process of face recognition.

Acknowledgments. This work was partially supported by the Italian Ministry of University and Scientific Research within the framework of the project “Ambient Intelligence: event analysis, sensor reconfiguration and multimodal interfaces” (PRIN - 2006-2008).

References

1. Land, E.H., McCann, J.J.: Lightness and Retinex Theory. *J. Opt. Soc. Am.* 61(1) (1971)
2. Jobson, D.J., Rahman, Z., Woodell, G.A.: A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes. *IEEE Transactions on Image Processing* 6(3), 965–976 (1997)
3. Belhumeur, P.N., Kriegman, D.J.: What is the set of images of an object under all possible lighting conditions? *IEEE International Conference on Computer Vision and Pattern Recognition* (1996)
4. Bhattacharyya, J.: Detecting Removing Specularities and Shadows in Images. Masters Thesis, Department of Electrical and Computer Engineering, McGill University (June 2004)
5. Rammamorthi, R., Hanrahan, P.: A Signal-Processing Framework for Inverse Rendering. In: *ACM SIGGRAPH* (2001)
6. Chen, T., Yin, W., Zhou, X.S., Comaniciu, D., Huang, T.S.: Illumination normalization for face recognition and uneven background correction using total variation based image models. In: *Proceedings of the IEEE Computer Society CVPR*, pp. 532–539 (2005)
7. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* 60, 259–268 (1992)
8. Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* 20, 89–97 (2004)
9. Lee, K.C., Ho, J., Kriegman, D.: Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence* 27(5), 684–698 (2005)
10. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001), <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
11. Hayakawa, H.: Photometric stereo under light source with arbitrary motion. *Journal of Optical Society of America A*, 11 (1994)
12. Horn, B., Brooks, M. (eds.): *Shape from Shading*. MIT Press, Cambridge (1989)
13. Shashua, A.: On photometric issues in 3d visual recognition from a single 2D image. *International Journal of Computer Vision* 21, 99–122 (1997)
14. Yuille, A.L., Snow, D., Epstein, R., Belhumeur, P.N.: Determining generative models of objects under varying illumination: Shape and albedo from multiple images using svd and integrability. *International Journal of Computer Vision* 35, 203–222 (1999)
15. Wang, H., Li, S.Z., Wang, Y.: Generalized quotient image. *IEEE International Conference on Computer Vision and Pattern Recognition* (2004)