FACE RECOGNITION UNDER VARYING LIGHTING BASED ON DERIVATES OF LOG IMAGE

Laiyun Qing[1,2], *Shiguang Shan*[2], *Wen Gao*[1,2]

[1]Graduate School, CAS, Beijing, 100039, China
[2]ICT-ISVISION Joint R&D Laboratory for Face Recognition, CAS, Beijing 100080, China

ABSTRACT

This paper considers the problem of recognizing faces under varying illuminations. First, we investigate the statistics of the derivative of the irradiance images (log) of human face and find that the distribution is very sparse. Based on this observation, we propose an illumination insensitive similarity measure based on the *min* operator of the derivatives of two images. Our experiments on the CMU-PIE database have shown that the proposed method improves the performance of a face recognition system when the probes are collected under varying lighting conditions.

1. INTRODUCTION

Face recognition has attracted much attention in the past decades for its wide potential applications in commerce and law enforcement, such as mug-shot database matching, identity authentication, access control, information security, and surveillance. Much progress has been made in the past few years [21].

However, face recognition remains a difficult, unsolved problem in general due to several bottlenecks, among which illumination change is one of the most challenging problems. It has been argued that the variations between the images of the same face due to illumination and viewing direction are almost always larger than the image variations due to the change in face identity [11]. These observations have been further verified by the evaluation of the state-of-the-art systems as well. The FERET tests [12] and the recent FRVT 2002 [13] revealed that the recognition performances of the best systems degraded significantly when the illuminations changed.

There has been much work dealing with illumination variation in face recognition. Generally, these approaches can be categorized into three fundamental categories: the invariant features based approaches, the statistics-based approaches and the model-based approaches [14].

The invariant features based approaches seek to utilize features that are invariant to the changes in appearance. Examples of such representation considered by early researchers are edge maps, image intensity derivatives, and images convolved with 2D Gabor-like filers. However, Adini's empirical study [1] had shown that "None of the representations considered is sufficient by itself to overcome image variations because of a change in the direction of illumination". Chen et al [4, 6] had drawn the conclusion that even for diffuse object, there were no discriminative functions that are invariant to illumination. But they provided two illumination insensitive measures based probabilistic information: the ratio image [4] and the distribution of the image gradient [6].

The statistics-based approaches learn the distribution of the images collected under varying lighting conditions in some suitable subspace/manifold by statistical analysis of a large training set. Recognition is then conducted by choosing the subspace/manifold closest to the novel image. Eigenface [14], Fisherface [3], and Bayesian method [10] are the typical methods belonging to this category. In this context, one of the most important observations is that the images of a convex Lambertian surface caused by varying illuminations lie in a low-dimensional subspace embedded in the image subspace. Especially, Hallinan [9] reported that 5-D subspace would suffice to represent most of the image variations due to illumination changes including extreme cases.

The model-based approaches are based on the reflectance equation and model the functional role of the extrinsic imaging parameters (such as lighting and view direction) explicitly. These methods build a generative model of the variations of the face image. 3D Linear subspace [17], Quotient Image method [18], Illumination Cones [8] and 3D Morphable model [5] are the successful methods belonged to this kind of approaches. Recently, the 9-D linear subspace [2, 16] using spherical harmonics was proposed, which facilitated the modeling of the more generic illuminations.

All the above approaches have their advantages and disadvantages. The invariant features need little *prior* knowledge and they are suitable for preprocessing, though the performances are not the best. The performances of the statistics-based approaches depend on the images used for training. If the illuminations of the images for training are sampled well enough, the performance is satisfying. The model-based methods are generative but need several images per object or the class in-

formation. And the assumption of Lambertian surface is not strictly satisfied in real applications.

In this paper, we propose an illumination insensitive similarity measure for face recognition under varying lighting. The measure is based on the observation that the partial derivatives of the log irradiances are very sparse.

The rest of the paper is organized as follows. The statistics of the derivates of the log of the face irradiances and the measure of the *min* operator based on the statistics are described in section 2. We show the experimental results of the proposed measure in section 3, followed by some concluding remarks in section 4.

2. THE ILLUMINATION INSENTIVE MEASURE

Assuming the face is Lambertian surface, denoting by $I(x,y)$ the input face image, $R(x,y)$ the reflectance image and $E(x,y)$ the irradiance image, the three images are related by

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

Rewriting Eq. (1) in log domain, it is

$$\log I(x,y) = \log R(x,y) + \log E(x,y)\,. \tag{2}$$

If we apply a derivative filter to Eq. (2), we can get

$$\nabla \log I(x,y) = \nabla \log R(x,y) + \nabla \log E(x,y)\,. \tag{3}$$

That is, the resulting filter output of a face image is the sum of that of the reflectance image and the irradiance images. We then check the statistical properties of the outputs of the filter of the irradiance images in the next subsection.

2.1. The statistics of the derivates of face irradiances

Assuming the human surface is a convex Lambertian surface and according the spherical harmonic images model proposed in [2, 16], the irradiance images is determined by the low frequency components of the illumination environment and the shape of the face, i.e.

$$E(x,y) = \sum_{l=0}^{2}\sum_{m=-l}^{l} A_l L_{lm} Y_{lm}(\alpha,\beta)\,, \tag{4}$$

where $A_l (A_0 = \pi, A_1 = 2\pi/3, A_2 = \pi/4)$ [2, 16] are the spherical harmonic coefficients of Lambertian reflectance, L_{lm} are the coefficients of the incident light, and Y_{lm} are the spherical harmonic functions. (α,β) is the normal of the point (x,y) in the input image.

We analyze the distribution of the derivates of the irradiance maps from a training set. Human faces can be assumed rationally to have the similar shapes. Therefore the training sets are constructed by the irradiance images of an average face shape under varying illuminations.

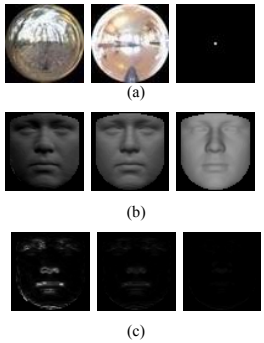

Figure 1. Some examples of the filtered face irradiance images. (a) The varying illuminations. (b) The irradiance images under the varying illuminations. (c) The corresponding vertical derivative filtered irradiance images.

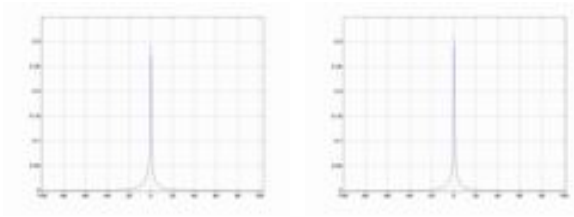

Figure 2. The distributions of the vertical derivative of the face irradiance images. (a) Natural illuminations. (b) Point light sources.

We select two sets of illumination environments. One set consists of 9 maps in Debevec's Light Probe Image Gallery (http://www.debevec.org/Probes/), which can be regards as the representatives of the natural illuminations. Debevec's maps represent diverse lighting conditions from 4 indoor settings and 5 outdoor settings. Because illumination statistics vary with elevation [7], we rotate each light probe horizontally per 30^o, resulting 12 different light conditions from one light probe. The other set is a point light source of different directions. The 64 directions of the flash are the same as the geodesic dome in the Yale B database [8].

We compute the nine low spherical harmonic coefficients L_{lm} ($0 \le l \le 2, -l \le m \le l$), for every illumination map as

$$L_{lm} = \int_{\theta_i=0}^{\pi}\int_{\phi_i=0}^{2\pi} L(\theta_i,\phi_i) Y_{lm}^{*} \sin\theta_i d\theta_i d\phi_i\,, \tag{5}$$

where $L(\theta_i,\phi_i)$ is the illumination intensity of the direction (θ_i,ϕ_i) . Then the irradiance image is computed with Eq. (4) using the average face shape. Finally, the derivate filter is applied to the log of the irradiance image to get the $\nabla \log E(x,y)$.

Figure 1 shows some examples of the varying illuminations and the corresponding filtered irradiance images. All the illuminations are scaled as $L_{00} = 60$ for better visualization. It can be seen that the filtered irradiances are almost zero (black pixels). Figure 2 shows the distri-

butions of the vertical derivatives for the two datasets. Though the illuminations vary very much, the two distributions have similar shape that are peaked at zero and fall off much faster than Gasussian, i.e., the outputs of the derivative filtered irradiances are very sparse. The statistics are similar with those of the natural images [7] and the natural illuminations [20]. Though human face is not convex strictly, the distribution does not change much.

2.2. The illumination insensitive measure

The above-mentioned observations that the filtered irradiance images are very sparse mean that the energy of the most of the pixels of the irradiance images is small. According to Eq. (3), at these pixels, the energy of the filtered image $\nabla \log I(x,y)$ is mostly caused by $\nabla \log R(x,y)$, i.e., $\nabla \log I(x,y) \approx \nabla \log R(x,y)$. The filtered reflectance image is invariant with illumination. Therefore, the filtered image $\nabla \log I(x,y)$ can be used as an illumination insensitive measure for face recognition. This measure had been used in [4], which were derived from the connection between the $\nabla \log I(x,y)$ and the gradient of the ratio of the two images. Then, the *Euclidean* distance map between the two images based on the feature $\nabla \log I(x,y)$ can be defined as

$$D1_{I_1,I_2}(x,y) = \left|\nabla \log I_1(x,y)\right| - \left|\nabla \log I_2(x,y)\right|, \qquad (6)$$

and used as distance of two images in face recognition in [4].

Actually, the distribution of the filtered face irradiance images and the filtered face reflectance images are very similar. Both of them have distribution similar with that of the natural images [7]. But for human faces, the locations of the pixels with larger intensity in filtered reflectance images are fixed, such as in the region of mouth, mustache and eyes. The irradiance images depend on the shape and the illumination. As the illuminations are random, the irradiance images are random, resulting in the random locations of the large intensity in the filtered irradiance images. Then the *min* operator on two or more filtered irradiance images of the same face can weaken the effect of varying illuminations. Therefore, we propose a new similarity map of two images based on the *min* operator as

$$S2_{I_1,I_2}(x,y) = \min\left(\left|\nabla \log I_1(x,y)\right|, \left|\nabla \log I_2(x,y)\right|\right) \qquad (7)$$

.

The *min* measure is something like Weiss' ML estimation of the reflectance image in [20]. In Weiss's ML estimation, the filtered reflectance image is the *median* of the filtered image sequences. Then the reflectance image can be recovered from the estimation of filtered reflectance image. However, the *median* operator needs at least three images per subject, which is not satisfied by many face recognition systems. In our *min* operator, if the two images are the same face, it is expected that $S2 \approx \nabla \log R$. Therefore the reflectance image can also be recovered from the similarity map $S2$. Only two images of the same face can give an estimation of the reflectance image. If more images are available, the result will be refined.

Let $v_2 = \left|\nabla \log E_1(x,y)\right|$ and $v_2 = \left|\nabla \log E_2(x,y)\right|$, the probability is $p(v_1)p(v_2)$. We compare the two measures, $D1$ and $S2$ in the three possible cases:

1. Both v_1 and v_2 are small. According the distribution of the $p(v)$ in section 2.1, $p(v_1)$ and $p(v_2)$ are large. Therefore this is the most possible case. In this case, both the two measures, $D1$ and $S2$ are insensitive to the variations in illuminations.
2. v_1 is small and v_2 is large. Then $p(v_1)$ is large and $p(v_2)$ is small. This is the less possible case. The difference $\left|v_1 - v_2\right|$ is large and then the measure $D1$ is large. The result of the *min* operator, $\left|\min(v_1, v_2)\right|$, is small and the measure $S2$ is more insensitive to illumination. The same is for the case of v_1 is large and v_2 is small.
3. Both v_1 and v_2 are large. $p(v_1)$ and $p(v_2)$ are small and this is the least possible case. In this case, $\left|v_1 - v_2\right|$ is small and $\left|\min(v_1, v_2)\right|$ is large. Therefore the measure $d1$ is more insensitive to illumination than the measure $S2$. But this case is little possible and then the measure $S2$ is more invariable to illumination in total.

For face recognition, we define the similarity score of two images as

$$s2(I_1, I_2) = \sum_{x,y} S2_{I_1,I_2}(x,y). \qquad (8)$$

If the two images are of the same face, it is expected that $S2 \approx \nabla \log R$ and $s2 \approx \sum_{x,y} \left|\nabla \log R(x,y)\right|$, i.e., the intrapersonal similarity maps are just like the filtered reflectance image of faces. If the two images are from different faces, the similarity map $S2$ is more random and $s2$ is smaller.

3. THE EXPERIMENTAL RESULTS

The publicly available face database, the CMU-PIE database [19], is used in our experiments. We select the frontal images with lighting variations for the experi-

ments. There are total 68 persons included in the database. There are 21 flashes and the room lights to illuminate the face. The images were captured with the room lights on and off, resulting in 43 different conditions.

The images in CMU-PIE database under the 43 illuminations classified into two subsets: "illum" and "lights". There are some differences between the two subsets. In the "illum" subset, the room lights is turned off and the subject is not wearing glasses. While in the "lights" subset, the room lights is turned on. And the subject is wearing glasses if and only if they normally do. The flash numbers are 02-21. The flash number "01" corresponds to no flash (only for the "lights" subset). See the reference [19] for more details of the meaning of the flash number.

Some examples of the facial images under different illuminations are shown in Figure 3.

Figure 3. Some examples of the images in the CMU-PIE database. The first three images in the first row are from "illum" set and the images in the second row are from "lights" set. The last image in the first row is taken with just the room light (f01 in "lights" set).

Figure 4. Some examples of the recovered reflectance images. The images in the first row is the reflectance images recovered by median operator with three input images, the images in the second row are the results of the min operator with the same input of the first row. The images in the third row are the other results of min operator with two or four input images (the last one).

3.1. The recovered reflectance images

We compare the reflectance images recovered by Weiss's *median* operator and our *min* operator in this subsection. In this experiments, two filters: horizontal and vertical derivative filters, are used. Figure 4 show the reflectance image using different number of input images recovered by the *median* operator and the *min* operator. The input images are given in Figure 3. Note that the *min* operator needs at least two images while the *median* operator needs at least three images and the number of the images needs to be odd.

3.2. The illumination insensitive measure

To test the variations in illuminations only, both the gallery and probe are from the same subsets. In the experiments, we select the flash "11" in each subset as galleries for the both subsets. As the illumination of the gallery released by the CMU-PIE database is the same as that of the flash "01" in subset "lights", the results on the subset "lights" with the gallery flash "01" is also given.

Our experiments compare the three measures: the correlation of the image intensity I , $L2(\nabla \log I(x,y))$ and $Min(\nabla \log I(x,y))$.Face recognition is achieved by finding a nearest neighbor based on the image similarity or distance. As many facial features such as eyes, mouth, and eyebrows are horizontal, we use the vertical derivate in experiments only.

The experimental results of face recognition on the CMU-PIE database are listed in Table 1. Both the measure $L2(\nabla \log I(x,y))$ and $Min(\nabla \log I(x,y))$ improve the performances of face recognition much when the probes are taken under different illuminations. And the error rates using the measure $Min(\nabla \log I(x,y))$ are the lowest.

Table 1. Error rate comparisons between different measures on the CMU-PIE database.

Gallery	Probe	Measure	Error rate (%)
"illum" (f11)	"illum"	Correlation (I)	47.1
		L2($\nabla \log I(x,y)$)	12.6
		Min($\nabla \log I(x,y)$)	9.6
"lights" (f11)	"lights"	Correlation (I)	14.3
		L2($\nabla \log I(x,y)$)	0.4
		Min($\nabla \log I(x,y)$)	0.4
"lights" (f01)	"lights"	Correlation (I)	18.5
		L2($\nabla \log I(x,y)$)	3.2
		Min($\nabla \log I(x,y)$)	0.3

4. CONCLUSION

In this paper, we investigate the statistics of the derivative of the log irradiance images of human face and find that the distribution is very sparse. Based on this observation, we propose an illumination insensitive measure for face recognition based on the *min* operator of the derivatives of log of two images. We compare the proposed measure for reflectance images recovering with the *median* operator proposed by Weiss [19]. When the probes are collected under varying illuminations, our experiments of face recognition on the CMU-PIE database show that the proposed measure are much better than the correlation of image intensity and a little better than the Euclidean distance of the derivate of the log image used in [4].

5. ACKNOWLEDGEMENT

This research is partly sponsored by Natural Science Foundation of China (under contract No. 60332010), National Hi-Tech Program of China (No.2001AA114190 and No. 2002AA118010), and ISVISION Technologies Co., Ltd. Portions of the research in this paper use the CMU-PIE database. The authors wish to thanks everyone involved in collecting these data.

6. REFERENCES

[1] Y. Adini, Y. Moses and S. Ullman, "Face Recognition: The Problem of Compensting for changes in illumination Direction", *IEEE Trans. On PAMI*, Vol.19, No.7, pp.721-732, 1997.

[2] R. Basri and D. Jacobs, "Lambertian Reflectance and Linear Subspaces", *Proc. ICCV'01*, pp. 383-390, 2001.

[3] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs Fisherfaces: recognition using class specific linear projection", *IEEE Trans. On PAMI*, Vol.20, No 7, June, 1997.

[4] P.N. Belhumeur and D.W. Jacobs. "Comparing images under variable illumination", *Proc. CVPR'98*, 1998.

[5] V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model", *IEEE Trans. on PAMI*, Vol.25, No.9, pp.1-12, 2003.

[6] H.F. Chen, P.N. Belhumeur and D.W. Jacobs, "In Search of Illumination Invariants", *Proc. CVPR'00*, Vol.1, pp.1254-1261, June, 2000.

[7] R.O. Dror, T.K. Leung, E.H. Adelson, and A.S. Willsky, "Statistics of real-world illumination", *Proc. CVPR'01,* Kauai, Hawaii, December 2001.

[8] A.S. Georghiades, P.N. Belhumeur and D.J.Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Differing Pose and Lighting", *IEEE Trans. On PAMI*, Vol.23, No.6, pp.643-660, 2001.

[9] P. Hallinan, "A Low-Dimensional Representation of Human Faces for Arbitrary Lightening Conditions," *Proc. CVPR'94*, pp.995-999, Seattle, WA, 1994.

[10] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian Face Recognition", *Pattern Recognition*, Vol.33, pp.1771-1782, 2000.

[11] Y. Moses, Y. Adini and S. Ullman, "Face Recognition: the Problem of Compensating for Changes in Illumination Direction", *Proc. ECCV'94*, Vol.I, pp.286-296, 1994.

[12] P.J. Phillips, H. Moon, et al., "The FERET Evaluation Methodology for Face-Recognition Algorithms", *IEEE Trans. On PAMI*, Vol.22, No.10, pp.1090-1104, 2000.

[13] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, "FRVT 2002: Evaluation Report", *http://www.frvt.org/DLs/FRVT_2002_Evaluation_Report.pdf*, March 2003.

[14] S. Shan, W. Gao, B. Cao, D. Zhao, "Illumination Normalization for Robust Face Recognition against Varying Lighting Conditions", *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Nice, France, Oct.2003, pp157-164.

[15] M. Turk and A. Pentland, "Eigenfaces for Recognition", *Journal of cognitive neuroscience*, Vol.3, No. , pp.71-86, January 1991.

[16] R. Ramamoorthi and P. Hanrahan, "An efficient representation for irradiance environment maps", *Proc. SIGGRAPH'01*, pp.497-500, August 2001.

[17] A. Shashua, "On photometric issues in 3D visual recognition from a single 2D image", *IJCV*, Vol.21, No.1-2, pp.99–122, 1997.

[18] A. Shashua and T. Riklin-Raviv, "The Quotient Image: Class-Based Re-Rendering and Recognition With Varying Illuminations", *IEEE Trans. on PAMI*, pp.129-139, 2001.

[19] T. Sim, S. Baker and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database", *Proc. FG'00*, Washington, DC, May 2002.

[20] Y. Weiss, "Deriving intrinsic images from image sequences", *Proc. ICCV'01*, Vol.II, pp.68–75, Vancouver, Canada, July 2001.

[21] W.Y. Zhao, R. Chellappa, A. Rosenfeld, et al., "Face recognition: A literature survey", *UMD CfAR Technical Report CAR-TR-948*, 2000.