LETTER

Spatially Adaptive Logarithmic Total Variation Model for Varying Light Face Recognition

Biao WANG[†a)], ***Nonmember***, **Weifeng LI**[†b)], ***Member***, **Zhimin LI**[††c)], ***and*** **Qingmin LIAO**[†d)], ***Nonmembers***

SUMMARY In this letter, we propose an extension to the classical logarithmic total variation (LTV) model for face recognition under variant illumination conditions. LTV treats all facial areas with the same regularization parameters, which inevitably results in the loss of useful facial details and is harmful for recognition tasks. To address this problem, we propose to assign the regularization parameters which balance the large-scale (illumination) and small-scale (reflectance) components in a spatially adaptive scheme. Face recognition experiments on both Extended Yale B and the large-scale FERET databases demonstrate the effectiveness of the proposed method.

key words: *face recognition, illumination normalization, logarithmic total variation (LTV) model*

1. Introduction

Over the past few decades, face recognition has remained a very active topic in computer vision communities. Although lots of effective algorithms have been proposed, robust face recognition under variant illumination conditions, which are common in real-world applications, is still challenging [1]. To address this problem, numerous illumination normalization methods have been proposed. Most classical methods take the assumption that the reflectance component corresponds to relatively higher spatial frequencies, while the illumination part corresponds to low spatial frequencies. For example, in [2], the authors proposed to remove several DCT coefficients corresponding to low frequencies. Logarithmic total variation (LTV) [3] proposed by Chen *et al.* utilizes the edge-preserving capability of the total variation model to remove the illumination component. Tan *et al.* [4] proposed a simple and efficient method based on a pipeline of image preprocessing (PP) operations, in which the major component is the carefully designed bandpass filter (i.e. difference of Gaussian (DOG)). The recently proposed WeberFace by Wang *et al.* [5] argues that the relative gradient in the form of a modified Weber contrast is illumination-insensitive. In this letter, we point out the limitation of the LTV model and address it by assigning the regularization parameters which balance the large-scale (illumination) and small-scale (reflectance) components in a spatially adaptive scheme. In the following, We denote the proposed method as the Spatially Adaptive LTV (SA-LTV) model. Experimental results on both Extended Yale B and FERET databse demonstrate its effectiveness.

Manuscript received June 5, 2012.
Manuscript revised August 6, 2012.

[†]The authors are with the Visual Information Processing Laboratory, Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, Beijing 100084, China.
[††]The author is with Luohu Branch, Shenzhen Municipal Public Security Bureau, Shenzhen, China.
a) E-mail: wangbiao08@mails.tsinghua.edu.cn
b) E-mail: li.weifeng@sz.tsinghua.edu.cn
c) E-mail: 20942118@qq.com
d) E-mail: liaoqm@tsinghua.edu.cn (Corresponding author)
DOI: 10.1587/transinf.E96.D.155

2. Limitation of LTV

Lambertian reflectance model implies that a face image $f(x, y)$ could be expressed by

$$f(x, y) = r(x, y)i(x, y), \tag{1}$$

in which $f(x, y)$ is the image pixel value, $r(x, y)$ is the reflectance and $i(x, y)$ is the illuminance at each pixel (x, y). $i(x, y)$ depends on the lighting source, while $r(x, y)$ depends only on the albedo of the face, thus could be regarded as the illumination-insensitive part.

By taking logarithmic transform to both side of Eq. (1), we have:

$$\log(f(x, y)) = \log(r(x, y)) + \log(i(x, y)). \tag{2}$$

The TV-L^1 model can decompose an input image $\log(f(x, y))$ into large-scale component $u(x, y)$ and small-scale component $v(x, y)$. LTV model takes $u(x, y), v(x, y)$ as the approximation to $\log(r(x, y))$ and $\log(i(x, y))$, respectively:

$$\begin{aligned}\log(i(x, y)) \approx \min_u & \int |\nabla u| dxdy \\ & + \lambda \int |\log(f(x, y)) - u(x, y)| dxdy,\end{aligned} \tag{3}$$

$$\log(r(x, y)) \approx v(x, y) = \log(f(x, y)) - \log(i(x, y)), \tag{4}$$

where $\int |\nabla u| dxdy$ is the total variation (TV) of u, which ensures that u is smooth (smoothness); and $\int |\log(f(x, y)) - u(x, y)| dxdy = \|\log(f(x, y)) - u(x, y)\|_{L_1}$ ensures that u is close to $\log(f(x, y))$ (fidelity). The regularization parameter $\lambda > 0$ is a scalar balancing the smoothness and fidelity. The larger λ is, the more facial details retains in u, which is an approximation of the illumination component $\log(i(x, y))$ and will be discarded for robust face recognition. Cast shadows usually appears on the flat areas like forehead and cheeks, and to remove them, a larger λ is required. Although certain details on these areas will be lost, they are relatively less important for face recognition. However, for

Copyright © 2013 The Institute of Electronics, Information and Communication Engineers

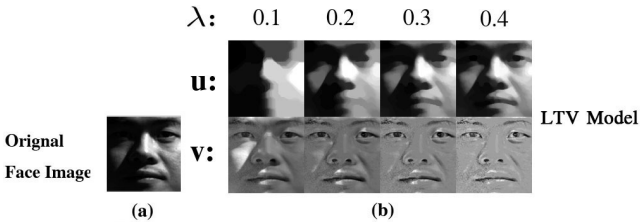

Fig. 1 Illustration of the limitation of LTV. (a). Original face image, (b). LTV outputs corresponding to different λ.

non-flat areas such as eyes,eyebrows, mouth, and nose tips, which are important for face recognition, a smaller λ is desired to avoid losing detail information. Figure 1 (a) illustrates an input face image under harsh illumination conditions, and Fig. 1 (b) illustrates the corresponding decomposition results of LTV model versus different regularization parameters, from which we could see that a smaller λ could preserve the details in non-flat areas but could not remove the cast shadows in flat area, while a larger λ could well remove the cast shadows in flat area but will lose much details in non-flat areas.

3. Spatially Adaptive LTV (SA-LTV) Model

From the aforementioned analysis, we could see that in order to remove the illumination without the loss of facial details, different face areas should be assigned with different regularization parameters according to the flatness of the area. Therefore, we propose the following spatially adaptive LTV (SA-LTV) model:

$$\log(i(x,y)) \approx \min_{u} \int |\nabla u| dxdy + \int \lambda(x,y)|\log(f(x,y)) - u(x,y)| dxdy. \quad (5)$$

in which $\lambda(x,y)$ is no longer a constant, and varies for different pixel positions. We define it as following:

$$\lambda(x,y) = \lambda_{non-flat} + (1 - mask(x,y))(\lambda_{flat} - \lambda_{non-flat}), \quad (6)$$

where $mask(x,y) \in [0,1]$ describes the flatness of pixel position (x,y): the closer $mask(x,y)$ to 0, the more flat it is; while the closer $mask(x,y)$ to 1, the more non-flat it is. As can be seen, LTV is a special case of SA-LTV when $\lambda_{non-flat} = \lambda_{flat}$.

To generate the $mask(x,y)$, we adopt the algorithm described in Table 1. Under harsh illumination conditions, the initial mask calculated by Step 1 and 2 will cover the facial areas corresponding to uneven illumination and cast shadows, as illustrated in Fig. 2 (a). To address this problem, the average mask calculated from all gallery samples is taken as a reference to determine the intrinsic flatness of these areas, just as illustrated in Fig. 2 (b) and Fig. 2 (c). The average flatness mask emphasizes the common non-flat areas shared by all the normally illuminated galleries, and will de-emphasize the "outliers" which are rarely present in the galleries. That is why it can be utilized as a reference to enhance the intrinsic discriminative details and reduce the side-effect resulted from the cast shadows. Moreover, in real-world applications, the calculation of $mask_{\bar{g}}(x,y)$ in Step 3 and 4 can be conducted off-line, and then can be directly applied to the probes. The results of SA-LTV are given in Fig. 2 (d) and Fig. 2 (e). By comparing them with Fig. 1 (b), we could clearly see that the proposed method achieves the best illumination removal effect while preserving most of the useful facial details.

Table 1 Generation of $mask(x,y)$ for a specific image.

Input: Image $p(x,y) = \log(f(x,y))$,
the set **G** of all N_g gallery images.
Output: $mask_p(x,y)$ for $p(x,y)$.
1. Calculate the initial mask by thresholding the gradient magnitudes within a local window:

$$mask_p(x,y) = \sum_{(\hat{x},\hat{y}) \in W^d_{(x,y)}} \left(\delta\left(\sqrt{h_x^2(\hat{x},\hat{y}) + h_y^2(\hat{x},\hat{y})} - thresh \right) \right),$$

where $W^d_{(x,y)}$ denotes a local square window centered at (x,y) with size $d \times d$. $h_x(\hat{x},\hat{y})$ and $h_y(\hat{x},\hat{y})$ are the horizontal and vertical gradients at $(\hat{x},\hat{y})$, respectively. $\delta(k) = 1$ if $k \geq 0$ and $\delta(k) = 0$ otherwise. $thresh$ is a threshold to determine whether a pixel is flat or not.
2. Normalize $mask_p(x,y)$ to [0, 1]. (See Fig. 2 (a))
3. For each gallery image $g_i(x,y) \in \mathbf{G}$, calculate its flatness mask $mask_{g_i}(x,y)$ according to step 1 and 2, and we get the average flatness mask for galleries:

$$mask_{\bar{g}}(x,y) = \frac{1}{N_g} \sum_{i=1}^{N_g} mask_{g_i}(x,y).$$

4. Normalize $mask_{\bar{g}}(x,y)$ to [0, 1]. (See Fig. 2 (b))
5. Get the final mask for image $p(x,y)$:

$$mask_p(x,y) = mask_p(x,y) \cdot mask_{\bar{g}}(x,y),$$

where $\cdot$ denotes pixel-wise multiplication. (See Fig. 2 (c))

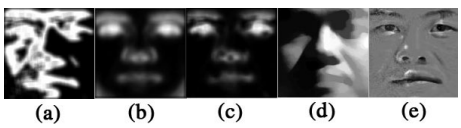

Fig. 2 Illustration of the proposed SA-LTV, and the input is the same as that of Fig. 1. (a). Initial mask, (b). Average flatness mask for galleries, (c). Final mask, (d) u, (e) v.

4. Experimental Results

Experiments are conducted on two publicly available face databases with variant illumination variations, namely, Extended Yale Face Database B [6] and the illumination subset of the FERET [7] database to illustrate the effectiveness of the proposed SA-LTV algorithm. All face images from the two databases are properly aligned, cropped and resized to 128×128. We will also compare our method with several state-of-the-art: DCT [2], LTV [3], Tan's preprocessing (PP) [4], and WeberFace [5]. The result of original images without any preprocessing (ORI) is provided as the baseline. The parameters of SA-LTV are empirically determined as following: $\lambda_{non-flat} = 0.1, \lambda_{flat} = 0.4, d = 7, thresh = 0.01$. For LTV, we report the best recognition results for variant λs.

After both gallery and probe samples are pre-processed by the aforementioned illumination normalization algorithms, we adopt the recently proposed local binary pattern (LBP) based face recognition scheme [8] for performance evaluation. The LBP operator was originally defined by encoding each pixel with 8 bit code, and each bit is obtained by thresholding the 3×3 neighborhood with the center pixel. Formally, we can define it as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} 2^n s(I_n - I_c), \tag{7}$$

in which (x_c, y_c) is the location of the center pixel, I_c and I_n are the intensity of the central pixel and its n-th neighbor, and $s(u)$ is 1 for $u \geq 0$ and 0 otherwise. There are two important extensions which makes LBP more powerful and widely used. The first one extends LBP to multi-scale by defining neighborhood of variant radii. For example, the encoding process of LBP with radius 2 and 8 neighbors is illustrated in Fig. 3. The second extension defines the so-called *uniform patterns*: a LBP code is 'uniform' if it contains no more than two 0-1/1-0 transitions. For example, the LBP code in Fig. 3 is non-uniform. It's the pioneering work of Ahonen *et al.* [8] that first successfully applied LBP to face recognition. To encode both texture and structure information for human face, the LBP coding map of a face image is divided into several nonoverlapping blocks and the histogram computed in each block is concatenated together. As suggested in [8], in our experiments, we utilize LBP pattern by thresholding 8 neighboring pixels in a circle of radius 2 and extracted the histograms in 8×8 blocks with 59 bins, each bin corresponding to a uniform pattern. Finally, the similarity of two LBP histogram is measured by the histogram intersection:

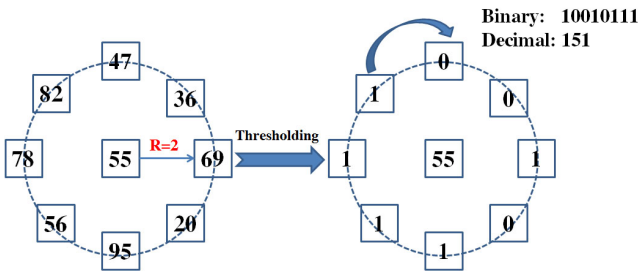

Fig. 3 An intuitive illustration of the LBP encoding process.

$$d(H_1, H_2) = \sum_i \min(h_1^i, h_2^i), \tag{8}$$

where H_1 and H_2 are the LBP spatial histograms for the gallery and probe sample respectively, and h_1^i, h_2^i are the corresponding i-th bin values. In summary, the overall face recognition framework is illustrated in Fig. 4.

Extended Yale Face Database B includes 38 subjects under 9 poses and 64 illumination conditions. Only the frontal images were chosen in our experiments. Totally there are 2,414 frontal images of 38 subjects under 64 illumination conditions. They are divided into five subsets according to the angle between the light source directions and the central camera axis: subset 1 (0° to 12°, 263 images), subset 2 (13° to 25°, 456 images), subset 3 (26° to 50°, 455 images), subset 4 (51° to 77°, 526 images), subset 5 (above 78°, 714 images). In our experiments, the images with the most neutral light condition ('A+00E+00') were used as the gallery, and images from subset 1-5 were used as the probes.

The corresponding results of several sample images from Extended Yale B processed by variant methods are given in Fig. 5. And the corresponding recognition rates of each methods for the five subsets are illustrated in Table 2.

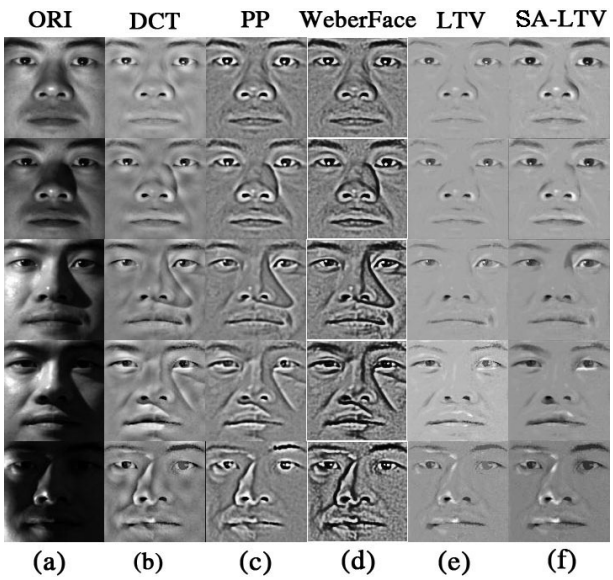

Fig. 5 Illumination normalization with different approaches on face images in the Extended Yale B database.

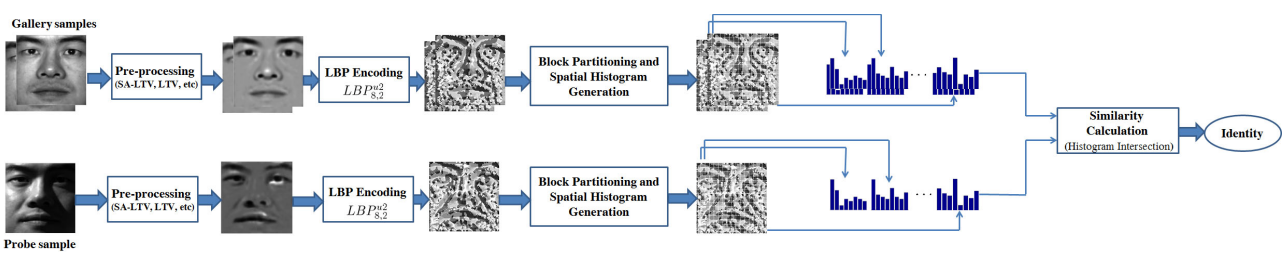

Fig. 4 The face recognition evaluation scheme based on variant illumination normalization methods.

Table 2 Recognition rates (%) on Extended Yale B.

Methods	S1	S2	S3	S4	S5	avg
ORI	100	100	97.6	62.0	35.0	72.0
DCT [2]	100	100	96.5	89.9	85.2	92.8
PP [4]	100	100	99.8	96.4	83.5	94.3
WeberFace [5]	100	100	100	**96.8**	86.6	95.3
LTV [3]	100	100	98.7	91.4	86.2	93.8
SA-LTV(Ours)	**100**	**100**	**100**	95.1	**94.4**	**97.3**

Table 3 Recognition rates (%) on the illumination subset of FERET.

ORI	DCT [2]	PP [4]	WeberFace [5]	LTV [3]	**SA-LTV(Ours)**
81.4	91.8	95.9	96.3	94.3	96.3

Our SA-LTV improves the overall average recognition rate from 72.0% to 97.3%. For subset 5, which is really challenging due to harsh illumination and shadows, SA-LTV improves the recognition rate from 35.0% to 94.4%, significantly better than the other approaches.

To further testify the proposed method on practical applications, we conduct experiment on the FERET database, one of the most commonly used large-scale face database. The gallery set Fa consists of $1,196$ images of $1,196$ subjects. The probe sets for standard FERET database contain four parts, each corresponding to variations of expression, illumination, short-term aging and long-term aging. Since we only focus on the illumination variations, experiments are only conducted on the illumination subset Fc, which contains 194 probe images under varying illumination conditions. The corresponding recognition rates of each methods are illustrated in Table 3. As can be seen, the proposed SA-LTV gives better result than LTV and is comparable to the recently proposed PP and WeberFace.

5. Conclusion

We propose a novel illumination normalization approach, SA-LTV, for face recognition under variant illumination conditions. It's an extension to the classical LTV model and can retain more facial details by adapting the regularization parameters according to the flatness of different facial areas. Experimental results on Extended Yale B and the illumination subset of the FERET database demonstrate the effectiveness of the proposed method.

Acknowledgements

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions, which are helpful to improve both technical and the literary quality of this paper. We also thanks the FERET Technical Agent, the U.S. National Institute of Standards and Technology for providing the FERET database. This work was supported by the Shenzhen-Hongkong Innovation Circle Project under grant No. ZYB200907070030A.

References

[1] J. Kittler, X. Zou, and K. Messer, "Illumination invariant face recognition: A survey," Int. Conf. Biometrics: Theory, Applications, and Systems, pp.1–8, 2007.
[2] J. Jiang and G. Feng, "Robustness analysis on facial image description in DCT domain," Electron. Lett., vol.43, no.24, pp.1354–1356, 2007.
[3] T. Chen, X.S. Zhou, D. Comaniciu, and T.S. Huang, "Total variation models for variable lighting face recognition," IEEE Trans. Pattern. Anal. Mach. Intell., vol.28, no.9, pp.1519–1524, 2006.
[4] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," IEEE Trans. Image Process., vol.19, no.6, pp.1635–1650, 2010.
[5] B. Wang, W. Li, W. Yang, and Q. Liao, "Illumination normalization based on weber's law with application to face recognition," IEEE Signal Process. Lett., vol.18, no.9, pp.462–465, 2011.
[6] P. Belhumeur, A. Georghiades, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.6, pp.643–660, 2001.
[7] P.J. Phillips, H. Moon, P. Rizvi, and P. Rauss, "The feret evalu- ation method for face recognition algorithms," IEEE Trans. Pattern. Anal. Mach. Intell., vol.22, no.10, pp.1090–1104, 2000.
[8] T. Ahonen, A. Hadid, and M. Pietikäinen: "Face description with local binary patterns: Application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.12, pp.2037–2041, 2010.