

Fast illumination normalization for robust eye localization under variable illumination

Fei Yang and Jianbo Su

Shanghai Jiao Tong University, No. 800, Dongchuan Road, Shanghai, China

E-mail: yangfei@sjtu.edu.cn

Abstract. Most eye localization methods suffer from illumination variation. To overcome this problem, we propose an illumination normalization technique as a preprocessing step before localizing eyes. This technique requires no training process, no assumption on the light conditions, and no alignment between different images for illumination normalization. Moreover, it is fast and thus effective for real-time applications. Experiment results verify the effectiveness and efficiency of the eye localization scheme with the proposed illumination normalization technique. © 2009 SPIE and IS&T. [DOI: 10.1117/1.3086868]

1 Introduction

Eye localization has been an intensive research topic for its wide applications in face recognition, driver fatigue monitoring systems, etc.¹⁻⁴ Unfortunately, it is still very hard to find an illumination-independent eye localization algorithm. Thus, the active infrared-based approaches are employed to diminish the influence of illumination variation.^{1,2} However, eye localization in visible-spectrum images is more relevant, since it is more practical in applications, without needing extra equipments. Consequently, it is often essential to normalize illumination before localizing eyes in visible-spectrum images of widely varied lighting. Jung *et al.*³ apply the self quotient image (SQI) to rectify illumination and get satisfactory eye localization results on their private face database. However, in SQI, the image is normalized by division over its smoothed version, which depends on the kernel size of the weighed Gaussian filter to a great extent. The kernel size is rather difficult to determine since the intrinsic information will be severely reduced if the kernel size is small or halo effects might appear if the kernel size is large. Although a multiscale technique is adopted to alleviate this problem, computational cost is added, and sometimes overcompensation regions still exist. In addition, the image noise would be amplified by the division operation, which also deteriorates the performance of eye localization. This letter proposes to take advantage of a more effective illumination normalization method, the logarithmic total variation (LTV) model,⁵ as a preprocessing step for eye localization, and it validates the performance of this eye localization scheme on some public face databases. In Sec. 2, the LTV model is presented to normalize illumination on the face. Since the computational

cost of the LTV model is expensive for real-time applications, a modified graph cut-based algorithm is proposed to solve the model in Sec. 3 so that the proposed preprocessing step is accelerated. Experimental results are given in Sec. 4, followed by conclusions in Sec. 5.

2 The LTV Model for Illumination Normalization

According to the Lambertian model, the captured face image $I(x, y)$ can be represented as

$$I(x, y) = \rho_{x,y} S_{x,y}, \quad (1)$$

where ρ is the albedo of the object surface, and $S(x, y)$ is the final light strength received at location (x, y) . The albedo ρ is the intrinsic representation of the captured face and is independent of the ambient lighting condition, which can be investigated for illumination-independent eye localization. Taking the logarithm of Eq. (1), we have:

$$\log(I) = \log(\rho) + \log(S). \quad (2)$$

If we denote $f = \log(I)$, $v = \log(\rho)$, and $u = \log(S)$, respectively, then

$$f = v + u. \quad (3)$$

Chen *et al.*⁵ argued that one of the differences between the intrinsic structure and the illumination pattern of a face image is the scale difference, and the intrinsic structure is usually smaller than the illumination artifacts and shadows. In a way, v promotes the variation patterns of the albedos of small-scale facial features. Thus, in order to eliminate the interference of ambient lighting in eye localization, one needs to extract v from f . We notice that the TV- L^1 model has shown its effectiveness for this task.⁵ Hence, we could use the TV- L^1 model to estimate u :

$$u = \operatorname{argmin}_u \int_{\Omega} (|\nabla u| + \lambda |f - u|) dx, \quad (4)$$

where λ is a penalty parameter. As λ increases, the term $|f - u|$ becomes more dominant, and thus u becomes more smoothing. One advantage of this LTV model is that the parameter λ , which depends only on the scale of image, is very easy to set. According to the LTV model, larger structures such as extrinsic illumination are left in u (or S), and ρ , which is taken as the output image of illumination normalization, can be obtained by $\rho = \exp(v) = \exp(f - u)$.

Therefore, the illumination normalization problem is transformed to how to solve the TV- L^1 model. Since partial differential equation-based algorithms often have numerical difficulties, Chen *et al.*⁵ cast Eq. (4) as a second-order cone program and solved it by the modern interior-point methods. But this iterative solution is expensive in both memory and computation time. To make this illumination normalization technique suitable for real-time applications, a graph cut-based algorithm is proposed to solve Eq. (4).

3 Efficient Solution to the TV- L^1 Model

First, we show how to decompose Eq. (4) as several independent binary energy minimization problems. The images discussed in this letter are defined as $m \times n$ matrices in $\mathbb{Z}^{m \times n}$, where \mathbb{Z} denotes the set of nonnegative integers that

Paper 08175LR received Nov. 22, 2008; revised manuscript received Jan. 14, 2009; accepted for publication Jan. 16, 2009; published online Feb. 25, 2009.

1017-9909/2009/18(1)/010503/3/\$25.00 © 2009 SPIE and IS&T.

represent the grayscale levels of images, and $m \times n$ denotes the size of images. Let $f \in \mathbb{Z}^{m \times n}$ and $u \in \mathbb{Z}^{m \times n}$ denote the original and separated images, respectively. According to Eq. (3), each element of these matrices satisfies

$$f_{i,j} = v_{i,j} + u_{i,j}, \quad \text{for } i = 1, \dots, m, \quad j = 1, \dots, n. \quad (5)$$

Moreover, we assume that all images satisfy the Neumann condition on the boundary of the domain Ω , i.e., the differentials on the image edges are defined to be zero. This assumption can be guaranteed by padding the image using the boundary elements. In this letter, to simplify and accelerate our algorithm, we just use the 4-neighbors of $u_{i,j}$ to approximate the gradient of u at the location (i,j) . Consequently, the regularization term in Eq. (4) can be defined in the discrete case by

$$|\nabla u| = \sum_{i,j} (|u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|). \quad (6)$$

Suppose μ as given, and define $B_{i,j} = 1$ for $u_{i,j} \geq \mu$; otherwise, $B_{i,j} = 0$. Define $x^+ = \max\{x, 0\}$, where x is an arbitrary real number. There exists $|B_{i,j} - B_{k,l}| \equiv (B_{i,j} - B_{k,l})^+ + (B_{k,l} - B_{i,j})^+$. For each pair of neighboring pixels (i,j) and (k,l) , $|u_{i,j} - u_{k,l}|$ can be expressed in terms of the elements of $B_{i,j}$ over all grayscale levels $\mu = 0, 1, \dots, l_{\max}$ as follows:

$$|u_{i,j} - u_{k,l}| = \sum_{\mu=0}^{l_{\max}} |B_{i,j} - B_{k,l}| = \sum_{\mu=0}^{l_{\max}} [(B_{i,j} - B_{k,l})^+ + (B_{k,l} - B_{i,j})^+], \quad (7)$$

where $l_{\max} = \max_{i,j} \{u_{i,j}\} \leq 255$. In this way, the original problem is reformulated into several independent binary problems based on the decomposition of a function into its level sets. Hence, combining Eq. (6) with Eq. (7), the first term in the right part of Eq. (4) can be binarized as

$$\int_{\Omega} |\nabla u| dx = \sum_{\mu=0}^{l_{\max}} \sum_{i,j} \{[(B_{i,j} - B_{i+1,j})^+ + (B_{i+1,j} - B_{i,j})^+] + [(B_{i,j} - B_{i,j+1})^+ + (B_{i,j+1} - B_{i,j})^+]\}. \quad (8)$$

Similarly, we define $B'_{i,j} = 1$ for $f_{i,j} \geq \mu$; otherwise, $B'_{i,j} = 0$. For binary numbers $B'_{i,j}$ and $B_{i,j}$, there exists $|B'_{i,j} - B_{i,j}| \equiv (1 - B'_{i,j})B_{i,j} + B'_{i,j}(1 - B_{i,j})$. The second term of the right side in Eq. (4) can then be binarized as

$$\int_{\Omega} |f - u| dx = \sum_{i,j} |f_{i,j} - u_{i,j}| = \sum_{\mu=0}^{l'_{\max}} \sum_{i,j} [(1 - B'_{i,j})B_{i,j} + B'_{i,j}(1 - B_{i,j})], \quad (9)$$

where $l'_{\max} = \max_{i,j} \{f_{i,j}\} \leq 255$. As a result, Eq. (4) is reformulated by combining Eq. (8) with Eq. (9). For given input f and λ and a fixed level $\mu \in \{0, 1, \dots, l'_{\max}\}$, Eq. (4) can be rewritten as

$$u = \operatorname{argmin}_{\mu=0}^{l'_{\max}} E(B; f, \lambda, \mu), \quad (10)$$

$$E(B; f, \lambda, \mu) = \sum_{i,j} \{[(B_{i,j} - B_{i+1,j})^+ + (B_{i+1,j} - B_{i,j})^+] + [(B_{i,j} - B_{i,j+1})^+ + (B_{i,j+1} - B_{i,j})^+] + \lambda[(1 - B'_{i,j})B_{i,j} + B'_{i,j}(1 - B_{i,j})]\}. \quad (11)$$

Thus, the problem of minimizing discretized Eq. (4) is decomposed into minimizing $E(B; f, \lambda, \mu)$ for all levels $\mu = 0, 1, \dots, l'_{\max}$. It is noted that the minimizer u^* of Eq. (4) can be constructed from the minimizers $\{B^*_{\mu} : \mu = 0, 1, \dots, l'_{\max}\}$ using the relationship⁶

$$u^*_{i,j} = \max[\mu | B^*_{\mu}(i,j) = 1]. \quad (12)$$

We then construct a directed capacitated graph corresponding to $E(B; f, \lambda, \mu)$ to find its minimizer B^*_{μ} at every level $\mu = 0, 1, \dots, l'_{\max}$. It is worth noting that the nodes/pixels in the graph are all binary and that the cost of each n -link connecting one pair of neighboring pixels equals 1 and a t -link connecting (i,j) with the source or the sink costs $\lambda B'_{i,j}$ and $\lambda(1 - B'_{i,j})$, respectively. In this way, a simplified two-terminal s - t graph representation of Eq.(11) is constructed, and then the minimizer B^*_{μ} is obtained via the min-cut algorithm on the graph.⁷

To sum up, by introducing divide-and-conquer methodology and a simplified graph representation, the minimizer u^* of Eq. (4) can be computed more efficiently. This method is essentially identical with Ref. 6 but is easier to understand and implement. Consequently, the LTV model-based illumination normalization technique is accelerated, called Fast LTV (FLTV) in this letter.

4 Experiments

Three well-known benchmark databases were chosen to evaluate the performance of the proposed eye localization scheme under both good and bad lighting conditions. In the Chinese Academy of Science: Pose, Expression, Accessory, Lighting (CAS-PEAL) face database,⁸ the Lighting and Normal subsets were used, which contain 2450 frontal face images under widely variable illumination and 1040 frontal face images under normal illumination, respectively. Yale face database B (Ref. 9), which contains 650 frontal face images was also adopted since it allows for testing under large variations of illumination, including strong shadow and side lighting. Another 3368 frontal face images under general illumination were chosen from the Face Recognition Technology (FERET) face database.¹⁰ All images are roughly cropped so that the facial regions are left and then resized to the appointed size. Then, illumination normalization is executed on the images. SQI and FLTV are both conducted here for comparison.

It is obvious that the darkest pixel in the eye region is most often a part of a pupil. Thus, this gray valley can be employed to localize eyes in face images. Generally, to suppress noise and alleviate the interference of other objects (e.g., hair, eye corner), a mean filter and a circular averaging filter are usually used to enhance the image. This simple eye localization approach requires no initialization and training process. Moreover, it is extremely fast and easy to implement and thus is widely used in practical applications. This approach is also used to test the illumina-

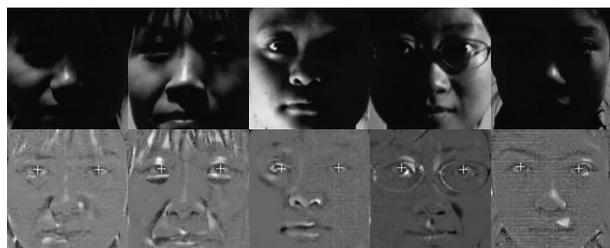


Fig. 1 Correct localization samples from CAS-Lighting subset.

tion normalization methods here. For higher accuracy and speed, we limit searching gray valleys to the top half of the face image.

A few localization results on the Lighting subset of CAS-PEAL and Yale B face database are illustrated in Fig. 1 and Fig. 2, respectively. The upper images are the original images; the lower images are the eye localization results on corresponding illumination-normalized images using FLTV. Note that the same method used to select λ in Ref. 4 is also adopted here. It can be observed that there exist no over compensation regions in the normalized images. To evaluate the accuracy of eye localization, a general criterion⁴ to claim successful eye localization is adopted:

$$err = \frac{\max\{\|l_c - l'_c\|, \|r_c - r'_c\|\}}{\|l_c - r_c\|} < 0.25, \quad (13)$$

where l_c and r_c are the manually marked left and right eye positions, and l'_c and r'_c are the automatically located positions. Thus, the correct localization rates directly on the original images and the preprocessed images with SQI and FLTV on the four test sets are separately obtained and are shown in Table 1. It can be seen that SQI and FLTV can greatly improve eye localization accuracy on all test sets, and FLTV outperforms SQI under both good and bad illumination. In order to evaluate the computational cost, we get the average location time per image by calculating the mean of the total execution time. The average location times on a 128×128 face image using SQI and FLTV as a preprocessing step are 0.781 s and 0.057 s, respectively. It is easy to conclude that FLTV is much faster than SQI and is effective for real-time eye localization. In addition, it takes 6.53 s on average for the original LTV model to process a 128×128 image, which is much slower than the

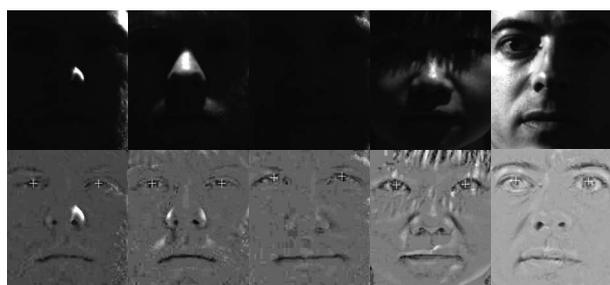


Fig. 2 Correct localization samples from Yale B.

Table 1 Correct localization rates on the four test sets.

Illumination preprocess	Bad illumination		Good illumination	
	CAS-Lighting	Yale B	CAS-Normal	FERET
None	50.8%	53.4%	69.4%	78.6%
SQI	95.5%	85.4%	99.3%	97.3%
FLTV	97.4%	86.5%	100.0%	98.1%

proposed FLTV and SQI methods. All the experiments are conducted with C++ on a Pentium D 2.8 GHz computer.

The experimental results demonstrate that our illumination normalization technique is reliable for robust eye localization under extreme lighting conditions. It can also greatly improve the eye localization accuracy on images under good lighting conditions. One reason may be that FLTV not only retains useful information for eye localization, but also eliminates the interference of hair of large size (structure) by leaving it in S . Therefore, using such a simple eye localization approach can achieve better or closer accuracy than other complicated eye localization algorithms.

5 Conclusion and Discussion

In this letter, we propose an illumination normalization technique as a preprocessing step before localizing eyes. This eye localization scheme is proven to be very fast and reliable under variable illumination. Motivated by the effectiveness and efficiency of the proposed illumination normalization technique, we might expect good performance when combining it with other existing eye localization algorithms, which is also our future work.

References

1. W. Hizem, Y. Yang, and B. Dorizzi, "Near-infrared sensing and associated landmark detection for face recognition," *J. Electron. Imaging* **17**(1), 011005 (2008).
2. Z. Zhu and Q. Ji, "Robust real-time eye detection and tracking under variable lighting conditions and various face orientations," *Comput. Vis. Image Underst.* **98**(1), 124–154 (2005).
3. S. U. Jung and J. H. Yoo, "A robust eye detection method in facial region," *Lect. Notes Comput. Sci.* **4418**, 596–606 (2007).
4. Z. H. Zhou and X. Geng, "Projection functions for eye detection," *Pattern Recogn.* **37**(5), 1049–1056 (2004).
5. 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.* **28**(9), 1519–1524 (2006).
6. J. Darbon and M. Sigelle, "Image restoration with discrete constrained total variation. Part I: fast and exact optimization," *J. Math. Imaging Vision* **26**(3), 261–276 (2006).
7. Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1124–1137 (2004).
8. W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The CAS-PEAL large-scale Chinese face database and evaluation protocols," *IEEE Trans. Syst. Man Cybern., Part A* **38**(1), 149–161 (2008).
9. A. Georghiades, D. Kriegman, and P. Belhumeur, "From few to many: generative models for recognition under variable pose and illumination," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 643–660 (2001).
10. P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(10), 1090–1104 (2000).