21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Illumination Normalization of Face Images with Cast Shadows

Tetsu Matsukawa Takahiro Okabe Yoichi Sato
Institute of Industrial Science, The University of Tokyo
E-mail; $\{$*te2, takahiro, ysato*$\}$*@iis.u-tokyo.ac.jp*

Abstract

We propose a method for extracting and combining small-scale and large-scale illumination insensitive features for face recognition that can work even in the presence of cast shadows. Although several methods have been proposed to extract such features, they are not designed to handle severe lighting variation on a face and thus fail to work if cast shadows are present. In this paper, we extend quotient image-based illumination normalization by explicitly taking cast shadows into account so that illumination insensitive large-scale features can be obtained. The experimental results show that the proposed method achieves favorable normalization results under difficult illuminations with cast shadows.

1. Introduction

Illumination variation is one of the most important remaining problems in face recognition. While the illumination subspace approach [1, 2] models the appearance of a face under varying illumination conditions from a set of images, the illumination normalization approach [3, 4, 5, 6, 7] extracts illumination-invariant/insensitive features from a single image and as a result is useful for a wide range of applications [8]. Hence, we focus on the latter approach in this paper.

Images of many objects including human faces are well approximated as a product of reflectance and shading components. The former is an intrinsic property of the object surface, and the latter depends on the object shape as well as on the illumination condition.

Shadings such as diffuse reflection components and attached shadows are low-frequency components of images [2]. Therefore, the existing techniques for illumination normalization use high-frequency components, *i.e.* *small-scale features* extracted from a single image for recognition, and discard its low-frequency components, *i.e.* *large-scale features*. For instance, SQI [3] normalizes an input image by dividing it by its smoothed version, and DCT [4] discards the low-frequency components of an input image. Recently, Xie *et al.* [9] reported that the method based on both small-scale and large-scale features works better than those based only on small-scale features. This result demonstrates that the large-scale features also contain useful information about identity of a person.

Shadows cast by facial parts such as a nose under harsh lighting or other objects such as a hat cause high-frequency components like shadow boundaries. It is reported that the variational image decomposition [5] can decompose an image into large-scale and small-scale features in which the shadow boundaries are well put into large-scale features even in the presence of cast shadows. In this way, we can obtain the small-scale feature that is less contaminated by cast shadows. However, the effects of cast shadows remain in the large-scale feature and it is difficult to normalize such features by existing methods.

To cope with this problem, our proposed method removes errors in the large-scale feature caused by cast shadows via quotient image-based normalization [3, 10]. More specifically, illumination-dependent components of the large-scale feature are approximately represented by a linear combination of basis functions. Our method uses error bases that can represent cast shadows in addition to illumination bases learned from bootstrap images. Then, the large-scale feature is normalized by removing the illumination-dependent components that take cast shadows into account. The use of such error bases is inspired by the dense error correction via L1 minimization [12], that is used in face recognition framework using sparse representation [13]. While their method requires multiple images under various illuminations for each gallery, our method requires only a single image.

2 Normalization using Large- and Small-Scale Features

A pixel intensity $I(x,y)$ is represented as a product of reflectance $\rho(x,y)$ and shading $S(x,y)$:

$$I(x,y) = \rho(x,y)S(x,y), \tag{1}$$

978-4-9906441-1-6

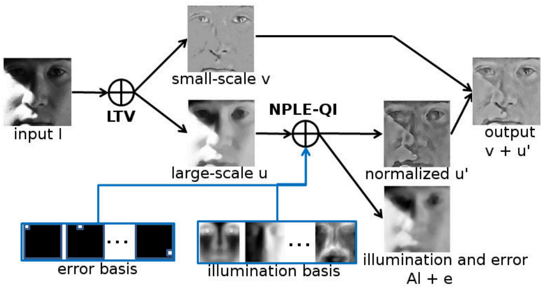

Figure 1. Block diagram of the proposed method.

where (x, y) is the coordinate of the image plane. Taking the logarithm in Eq.(1), we can convert the productive form into the additive form:

$$f(x, y) = v(x, y) + u(x, y), \tag{2}$$

where $f(x, y) = \log I(x, y)$, $v(x, y) = \log \rho(x, y)$, and $u(x, y) = \log S(x, y)$. We call v a small-scale feature and u a large-scale feature[1]. In this paper, this decomposition is obtained by LTV [5].

The illumination normalization framework based on small-scale and large-scale features (S&L) [9] normalizes the illumination variation mainly on the large-scale features u and recognition is done using both normalized large-scale features u' and small-scale features v. As shown in Fig.1, we propose an illumination normalization method that can work on this framework.

From next section, for the sake of simplicity, we raster scan pixel intensities of $v(x, y), u(x, y)$ and represent them as column vectors $\boldsymbol{v}$ and $\boldsymbol{u}$.

3 Shadow Insensitive Normalization

3.1 Quotient Image with Error Term

In this section, we describe the normalization method of large-scale feature $\boldsymbol{u}$ in the presence of cast shadows.

Let the bootstrap images consist of a number of subjects under various illumination conditions but do not contain severe cast shadows such as cast shadows by other objects. First, we decompose M bootstrap images by using LTV, and obtain the large-scale features $\{\boldsymbol{u}_m\}_{m=1}^{M}$. Then, we obtain the eigenvector matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_K]$ with the largest K eigenvalues by applying PCA to the matrix $[\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_M]$.

After that, we use the bases $\boldsymbol{A}$ to approximate the large-scale feature $\boldsymbol{u}$ of any image out of bootstrap images. If $\boldsymbol{u}$ is not contaminated by cast shadows, it is approximately represented by $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{l}$. Here, $\boldsymbol{l} = (l_1, l_2, \cdots, l_K)^T$ is the coefficients of linear combination. When severe contaminations such as cast shadows by other objects are present, $\boldsymbol{u}$ is not represented well by the learned bases from bootstrap images. To remove such effects, we incorporate an error term $\boldsymbol{e}$ and represent the large-scale feature as [2]

$$\boldsymbol{u} \sim \boldsymbol{A}\boldsymbol{l} + \boldsymbol{e}. \tag{3}$$

In a similar manner to a quotient image [10, 3], we can assume that the ratio between the large-scale feature and its approximation contains discriminative facial features. Since both large-scale and small-scale features are in the logarithmic domain, the ratio becomes subtraction. Therefore, the quotient image $\boldsymbol{u'}$ of the large-scale feature, *i.e.* the normalized large-scale feature can be defined by

$$\boldsymbol{u'} = \boldsymbol{u} - (\boldsymbol{A}\boldsymbol{l} + \boldsymbol{e}). \tag{4}$$

As described in the next section, we compute the approximation $(\boldsymbol{A}\boldsymbol{l} + \boldsymbol{e})$ on the basis of L1 minimization algorithm. Once we obtain $\boldsymbol{u'}$, we use the sum of the small-scale feature and the normalized large-scale feature $\boldsymbol{v} + \boldsymbol{u'}$ for face recognition.

We call this method *Non Point Light and Error Quotient Image (NPLE-QI)*. This is because the PCA illumination bases of quotient images are shown to be linearly related to spherical harmonics bases [2, 3] representing diffuse lighting, *i.e., Non Point Light* source [3].

3.2 Fitting via L1-Minimization

Next, we address how to estimate the coefficients $\boldsymbol{l}$ and error term $\boldsymbol{e}$ in NPLE-QI when large-scale feature $\boldsymbol{u}$ cannot be well represented as $\boldsymbol{A}\boldsymbol{l}$ due to cast shadows. We represent the error term by using a linear combination of basis functions:

$$\boldsymbol{e} = \boldsymbol{B}\boldsymbol{b}. \tag{5}$$

Here, $\boldsymbol{B}$ is a set of basis functions and $\boldsymbol{b}$ is their coefficients.

We assume that pixels contaminated by cast shadows can arbitrarily distribute according to objects that cast shadows on a face, and then we set $\boldsymbol{B} = \xi \boldsymbol{I}_e$. Here, the matrix $\boldsymbol{I}_e$ is the identity matrix of $P \times P$, where P is the number of image pixels. The scalar parameter ξ controls the contribution of the error term in optimization.

By integrating the illumination term $\boldsymbol{A}\boldsymbol{l}$ and error term $\boldsymbol{e}$, and allowing the fitting error up to $\epsilon|\boldsymbol{u}|_2$, the coefficients $\boldsymbol{w} = (\boldsymbol{l}^T, \boldsymbol{b}^T)^T$ are given by

$$\min |\boldsymbol{w}|_1 \quad \text{subject to} \quad |\boldsymbol{u} - \boldsymbol{C}\boldsymbol{w}|_2 \leq \epsilon|\boldsymbol{u}|_2, \tag{6}$$

[1] The decomposition in Eq.(1) is not unique. For example, Chen *et al.* [5] consider high frequency components of reflectance as small-scale features.

[2] The error term is not restricted to negative values. Hence, this term removes other nuisances for recognition, such as specular high lights, and can remove fitting errors of $\boldsymbol{A}\boldsymbol{l}$.

[3] We performed PCA in logarithmic domain because it produces more stable results for the quotient image. We confirmed that the variation of the image under non-points lights can be modeled as low dimensional subspaces in logarithmic domain.

Algorithm 1 NPLE-QI for large-scale feature

Input: Large-scale feature $\boldsymbol{u}$.

$\hat{\boldsymbol{l}}$ = argmin $|\boldsymbol{u} - \boldsymbol{A}\boldsymbol{l}|_2$.
if $|\boldsymbol{u} - \boldsymbol{A}\hat{\boldsymbol{l}}|_2 > \epsilon|\ \boldsymbol{u}\ |_2$ **then**
  $\hat{\boldsymbol{w}}$ = arg min $|\boldsymbol{w}|_1$ s. t. $|\boldsymbol{u} - \boldsymbol{C}\boldsymbol{w}|_2 \leq \epsilon|\boldsymbol{u}|_2$.
  $\boldsymbol{u}' = \boldsymbol{u}$ - $\boldsymbol{C}\hat{\boldsymbol{w}}$.
else
  $\boldsymbol{u}' = \boldsymbol{u}$ - $\boldsymbol{A}\hat{\boldsymbol{l}}$.
end if

Output: Normalized large-scale feature $\boldsymbol{u}'$.

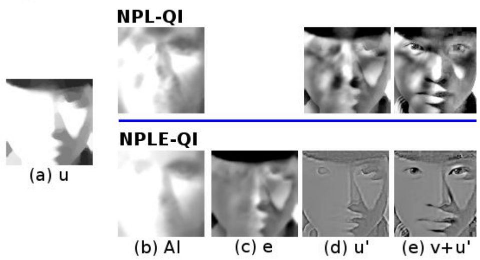

Figure 2. Normalization of Large-Scale Feature using NPLE-QI.

where $\boldsymbol{C} = [\boldsymbol{A}, \boldsymbol{B}]$. There are many algorithms that can solve this optimization problem. We use l1qc log-barrier function [11] of l1-magic. The actual algorithm of NPLE-QI is described in Algorithm.1.

An example of the results are shown in Fig2. The illumination term (b) and the error term (c) are computed from the original large-scale feature (a). As shown in (b), we can see that NPL-QI, which corresponds to the proposed method without the error term, cannot represent cast shadows by a hat. In addition, the effects of cast shadows propagate to non-shadowed pixels. On the other hand, we can see that the error term of NPLE-QI captures the cast shadows as shown in (c), and that the normalized large-scale feature obtained by using our proposed NPLE-QI is not sensitive to cast shadows as shown in (d). We use the sum of the quotient image (normalized large-scale feature) and small-scale feature (e) for face recognition.

4 Experiments

4.1 Setup

Datasets: We used four datasets: Extended Yale B [1], Multi-PIE [14], CAS-PEAL [15], and our own dataset termed CAST, which consists of face images with cast shadows. The Extended Yale B consists of images of 38 subjects taken under 64 different light source directions. Those images are classified into five subsets in accordance with the angle between the frontal direction and the light source direction. The Multi-PIE consists of images of 337 subjects taken under 21 different light source directions. The CAS-PEAL consists of images of 1040 subjects taken under various illumination conditions and with various accessories. We used a lighting probe set and chose images of subjects wearing hats from an accessory set. The CAST consists of images of 14 subjects taken under 50 different imaging conditions, with four different kinds of occluders such as hats and under various light source directions. In all datasets, only images of frontal faces are used.

Table 1. Recognition rates of various datasets.

Method	Extended YaleB	Multi-PIE	CAS PEAL		CAST
			Light	Hat	
HE	55.87	57.92	4.37	25.00	18.43
TT [6]	86.67	99.46	18.28	24.73	76.00
DCT [4]	90.15	99.15	19.66	27.68	81.00
NDF [7]	90.57	**99.85**	21.89	32.50	79.71
LTV [5]	90.34	98.80	21.58	30.54	82.29
NPL-QI [3]	83.69	97.69	23.23	19.46	37.61
NPLE-QI	93.74	99.69	22.60	27.68	67.43
S&L(DCT) [9]	92.26	98.85	21.98	32.05	82.86
S&L(NPL-QI) [9]	90.83	98.62	**27.06**	21.07	49.71
S&L(NPLE-QI)	**94.71**	99.31	26.62	**33.04**	**84.57**

Recognition Protocol: All face images are cropped in accordance with the coordinates of eyes, and resized to 100×100 pixels with 256 gray levels. Nearest neighbor classifier based on normalized cross correlation is used for face recognition as it has been used in much previous related researches [5, 7, 9]. For each subject, only one image taken under an frontal lighting condition is registered as a reference image.

Bootstrap Images: We used images of 10 subjects taken under various light source directions as bootstrap images. The Multi-PIE is used as the bootstrap images for the Extended Yale B. The Extended Yale B is used as the bootstrap image for the other datasets.

Parameters: As default parameters, the number of illumination basis is set to 20, the ϵ is set to 0.03, and the ξ is set to 0.1. These parameters are commonly used for all datasets.

4.2 Results

We compared our proposed NPLE-QI with NPL-QI [3, 9]. For reference, middle of Table 1 lists the recognition rates of both methods not for large-scale features but for original images (refer as NPLE-QI and NPL-QI). We can see that NPLE-QI performs better than NPL-QI in many cases, especially when cast shadows are dominant in images. Below, we investigate the performance of NPLE-QI that is applied to large-scale features and combined with small-scale features (refer as S&L) in detail.

We also show the result images of S&L(NPL-QI) and S&L(NPLE-QI) in Fig. 3. We can confirm that S&L(NPLE-QI) has better visual quality.

Parameter ϵ: We investigated the recognition rates when the parameter ϵ of NPLE-QI changed. The results are shown in Fig. 4. The up-left figure shows the

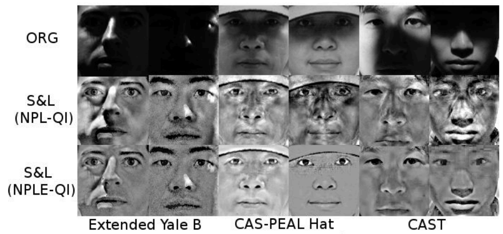

Figure 3. Example of illumination normalization results.

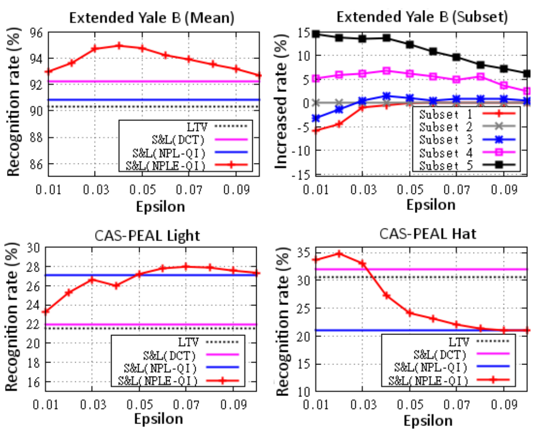

Figure 4. Results in different settings of ϵ.

mean recognition rates and the up-right figure shows the difference (gain) of the performance of S&L(NPLE-QI) from S&L(NPL-QI) for each subset of the Extended Yale B. The performance of S&L(NPLE-QI) is better than the other combinations (DCT and NPL-QI) and the original LTV. We can see that our proposed S&L(NPLE-QI) achieves much better results than S&L(NPL-QI) for subsets 4 and 5, *i.e.* for images taken under harsh lighting. Because cast shadows become dominant in such images and it is difficult to normalize them by using only the illumination bases learned from bootstrap images. Such degradation of NPL-QI is consistent with the result of [9]. In NPLE-QI, smaller ϵ removes more errors and such parameters are better when subset 4, 5. Similarly, the optimal parameter of ϵ in CAS-PEAL Hat is smaller than CAS-PEAL Light. This is because the bootstrap images (Extended Yale) contain various lighting conditions, but do not contain variations caused by hat.

Comparison with Other Methods: We compared our proposed method with other illumination normalization methods; histogram equalization (HE), TT [6] [4], DCT [4], and NDF [7]. All the methods were implemented by us, and their parameters were carefully tuned so that the best performance could be obtained. The results are shown in Table 1. We can see that the recognition rates of our proposed S&L(NPLE-QI) are higher than those of the other methods in many cases, especially when the effects of cast shadows are severe such as Extended Yale B and CAST.

[4] In TT, only the illumination normalization stage of the whole recognition algorithm was used.

5 Conclusion

We proposed a method for illumination normalization of face images taken under harsh lighting and with shadows cast by other objects. On the basis of removal of errors including cast shadows, our method enables us to extract discriminative facial features from large-scale features of images. The experimental results demonstrated that the proposed method achieves favorable normalization results under difficult illuminations with cast shadows.

References

[1] A.S. Georghiades, P.N. Belhumeur and D.J. Kriegman, From few to many:Illumination cone models for face recognition under variable lighting and pose, *IEEE Trans. on PAMI*, vol.23, no.6, pp.643–660, 2001.

[2] R. Ramamoorthi, Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of Lambertian Object, *IEEE Trans. on PAMI*, vol.24, no.10, pp.1322–1333, 2002.

[3] H. Wang, S.Z. Li and Y. Wang, Generalized Quotient Images, In Proc.*CVPR*, pp.498–505, 2004.

[4] W. Chen, M.J. Er and S. Wu, Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain, *IEEE Trans. on SMC:B*, vol. 36,no.2, pp.458–466, 2006.

[5] T. Chen, W. Yin, X.S. Zhou, D. Comaniciu and T.S. Huang, Total variation models for variable lighting face recognition, *IEEE Trans. on PAMI*, vol.28, no.9, pp.1519–1524, 2006.

[6] X. Tan and B. Triggs, Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions, *IEEE Trans. on IP*, vol.19, no.6, pp.1635–1650, 2010.

[7] L.-H. Chen, Y.-H. Yang, C.-S. Chen and M.-Y. Cheng, Illumination Invariant Feature Extraction Based on Natural Images Statistics - Taking Face Images as An Example, In Proc. *CVPR*, pp.681–688, 2011.

[8] X. Tan, S. Chen, Z.-H. Zhou and F. Zhang, Face recognition from a single image per person: A Survey, *Pattern Recognition*, vol.39, pp.1725–1745, 2006.

[9] X. Xie, W.-S. Zheng and P.C. Yuen, Normalization of Face Illumination Based on Large- and Small- Scale Features, *IEEE Trans. on IP*, vol.20, no.7, pp.1807–1821, 2011.

[10] A. Shashua and T. Riklin-Raviv, The Quotient Image: Class-Based Re-Rendering and Recognition with Varying Illuminations, *IEEE Trans. on PAMI*, vol.23,no.12, pp.129–139,2001.

[11] E. Candes, J. Romberg, $l^1 magic$: Recovery of Sparse Signals via Convex Programming, http://www.acm.caltech.edu/l1magic/, 2005.

[12] J. Wright and Y,Ma, Dense Error Correction via L1 Minimization, *IEEE Trans. on Information Theory*, vol.56, no.7, pp.3540–3560, 2010.

[13] J. Wright, A.Y. Yang, A. Ganesh, S.S. Satr and Y. Ma, Robust Face Recognition via Sparse Representation, *IEEE Trans. on PAMI*, vol.31, no.2, pp.210–227, 2009.

[14] R.Gross, I.Matthews, J.Cohn, T.Kanade, and S.Baker, Multi-PIE, *Image and Vision Computing*, vol.28, no.5, pp.807–813, 2010.

[15] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang and D. Zhao, The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations, *IEEE Trans. on SMC*, Part A, vol.38, no.1, pp.149–161, 2008.