Face Samples Re-lighting for Detection Based on the Harmonic Images

Jie Chen[1], Yuemin Li[2], Laiyun Qing[3], Baocai Yin[2], and Wen Gao[1,3]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China
{chenjie, ymli, lyqing}@jdl.ac.cn

[2] Multimedia and Intelligent Software Technology Laboratory, Beijing University of Technology, Beijing 100022, China
ybc@bjut.edu.cn

[3] ICT-ISVISION JDL for AFR, Institute of Computing Technology, CAS, Beijing, 100080, China
wgao@ict.ac.cn

Abstract. Different environment illumination has a great impact on face detection. In this paper, we present a solution by the face relighting based on the harmonic images. The basic idea is that there exist nine harmonic images which can be derived from a 3D model of a face, and by which we can estimate the illumination coefficient of any face samples. To detect faces under the certain lighting conditions, we relight those original face samples to get more new face samples under the various possible lighting conditions by an illumination ratio image and then add them to the training set. By train a classifier based on Support Vector Machine (SVM), the experimental results turn out that the relighting subspace is effective during the detection under the diverse lighting conditions. We also use the relighting database to train an AdaBoost-based face detector and test it on the MIT+CMU frontal face test set. The experimental results show that the data collection can be efficiently speeded up by the proposed methods.

1 Introduction

Over the past ten years, face detection has been thoroughly studied in computer vision research for its interesting applications. Face detection is to determine whether there are any faces within a given image, and return the location and extent of each face in the image if one or more faces present [20].A hierarchical template matching method for face detection was proposed by Miao et al. [10]. Recently, the emphasis has been laid on data-driven learning-based techniques [8, 9, 14, 15, 19, 21]. However, different environment illumination has a great effect on face detection [20]. To build a robust and efficient face detection system, lighting variations is one of the main technical challenges.

In order to deal with the illumination variations on the faces, many methods have been exploited [1, 7, 11, 13, 16]. Recently, Basri et al. [1] and Ramamoorthi

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3332, pp. 585–592, 2004.

et al. [13] presented that the appearance of a convex Lambertian object can be well represented with a 9-D linear subspace. By using spherical harmonics and signal-processing techniques, they have shown that the set of images of a convex Lambertian object obtained under varying lighting conditions can be approximated by a 9-D subspace spanned by nine basis images of the object, called harmonic images, each of which corresponds to an image of the object illuminated under harmonic lights whose distributions are specified in terms of spherical harmonics [1, 13]. This discovery has greatly facilitated the modelling of generic illumination and provides the possibility to solve face recognition problem under varying lighting conditions, especially the outdoor environments. The 9-D subspace defined with harmonic images [1] and Harmonic Exemplars [22] provided the possibility to recognize facial images under the diverse lighting conditions. Based on the face relighting model proposed by Basri and Ramamoorthi, Qing proposed the method to relight faces based on illumination ratio image [12].Chen presents a genetic algorithm (GA)-based method to swell face database through re-sampling from existing faces and re-lighted the samples by the linear point light source [2].

In this paper, we propose a method for face relighting based on harmonic images [1, 13]. It is to relight face samples under any certain illumination to simulate the possible lighting variations of faces in the images. By this scheme, we can enrich the lighting variations of the training set. Using these newly produced samples together with the original, we train a classifier SVM and prove that the hit rates can be improved by this method.

The rest of this paper is organized as following: In section 2, we introduce the method to relight face samples to the certain illumination based on the harmonic images. The experiment results are described in Section 3. In Section 4, we give the conclusions.

2 Face Relighting with the Spherical Harmonics

As discussed in [1, 13], assuming a face is a convex Lambertian surface, we can denote the face image:

$$I(x, y) = \rho(x, y)\overrightarrow{n}(x, y)\overrightarrow{s} \tag{1}$$

where $\rho(x, y)$ is the albedo of the point (x, y) ; $\overrightarrow{n}(x, y)$ is the surface normal direction; $\overrightarrow{s}$ is the point light source direction, whose magnitude is the light source intensity.

In space-frequency domain, Lambertian surface is a low-pass filter and the set of images of a Lambertian object under varying lighting can be approximated by a 9D linear subspace spanned by the harmonic images b_{lm} $(0 \leq l \leq 2, -l \leq m \leq l)$[13]. The harmonic images are defines as:

$$b_{lm}(x, y) = \rho(x, y)A_l Y_{lm}(\theta(x, y), \phi(x, y)), \tag{2}$$

where Y_{lm} is the spherical harmonic at the surface normal; (θ, ϕ) a pair of angles corresponding to the pixels, is the function of (x, y) and $0 \leq \theta \leq \pi$, $0 \leq \phi \leq 2\pi$; $A_l(A_0 = \pi, A_1 = 2\pi/3, A_2 = \pi/4)$ is the spherical harmonics coefficients.

The image under the arbitrary lighting can be written as:

$$I(x,y) = \sum_{l=0}^{2} \sum_{m=-l}^{l} L_{lm} b_{lm}, \tag{3}$$

where L_{lm} is the spherical harmonic coefficients of the specific lighting. The nine lower spherical harmonic coefficients L_{lm} ($0 \le l \le 2$) can be estimated as discussed in [13]. Given an input image $\mathbf{I}$ (a column vector of n elements, n is the number of pixels in an image), if the harmonic images of an object are known, then the coefficients of the illumination $\mathbf{L}$ can be solved by the least squares problem:

$$\widehat{\mathbf{L}} = \arg\min \|\mathbf{BL} - \mathbf{I}\|, \tag{4}$$

where $\mathbf{B}$ denotes the harmonic images, arranged as a $n \times 9$ matrix. Every column of $\mathbf{B}$ contains one harmonic image b_{lm} as in Eq. (2).

Assumes a convex Lambertian object in a distant isotropic illumination, its irradiance has been proved that its most energy is constrained in the three low order frequency components and its frequency form can be formulated as [1] [13]:

$$E(\theta,\phi) \approx \sum_{l=0}^{2} \sum_{m=-l}^{l} A_l L_{lm} Y_{lm}(\theta,\phi). \tag{5}$$

Once we have estimated the lighting of the original image as in Eq.(4), it is commonsense to relight it to the canonical illumination with the illumination ratio image. According to Eq. (2), (3), (5), illumination ratio image between the canonical image and the original image is defined as [12]:

$$IRI(x,y) = \frac{I_{can}(x,y)}{I_{org}(x,y)} = \frac{E_{can}(\theta(x,y),\phi(x,y))}{E_{org}(\theta(x,y),\phi(x,y))}, \tag{6}$$

where the subscripts are the illumination indexes, and E is the incident irradiance. Relighting image with the illumination ratio image can be rewritten as:

$$I_{can}(x,y) = IRI(x,y) \times I_{org}(x,y). \tag{7}$$

The ratio-image above defined is almost useless since it is only applicable to the original face. However, noticing that all faces have similar 2D and 3D shapes, we can firstly warp all faces to the generic shape and then compute the ratio-image for relighting the face images. It is then easy to reversely warp the relighting image back to its original shape. Currently, we just warp the 2D face image to a predefined mean shape, as defined in ASM. After the warp procedure, all face images are expected to have quite similar 3D shape. Using the face relighting method, we can get samples under the specific illumination conditions. Some relighting examples are illustrated in Fig. 1.

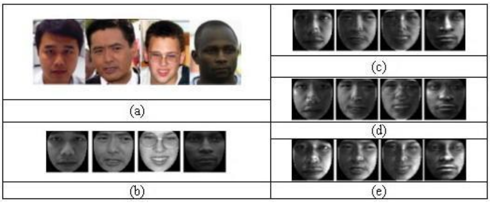


Fig. 1. Face samples relighting. (a) The original image; (b) cropped, normalized and masked face samples; (c) (d) (e) relighting under different lighting conditions.

3 Experiment Results

3.1 Face-Samples Preprocessing

The data set is consisted of a training set of 6977 images (2429 faces and 4548 non-faces) and two test sets. The first test set (Set1) is composed of 24045 images (472 faces and 23573 non-faces). All of these images were 19×19 grayscale and they are available on the CBCL webpage [24]. The second test set (Set2) is a subset of CAS-PEAL, which can be downloaded from [5]. And the subset includes: CAS-PEAL_Up_0, CAS-PEAL_Dn_0, CAS-PEAL_Mid_0, where CAS-PEAL_Up_0 means those images are illuminated by the 0 degree light source form overhead, CAS-PEAL_Dn_0 and CAS-PEAL_Mid_0 are also be captured under 0 degree light source. Some examples are shown in Fig.2, where the arrows denote the direction of the light source. Using the method presented above, we can relight those face samples in the training set.

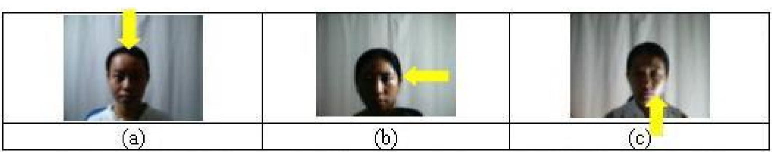


Fig. 2. Some test samples of CAS-PEAL. (a) is an example from CAS-PEAL_Up_0; (b) is a example from CAS-PEAL_Mid_0; (c) is a example from CAS-PEAL_Dn_0.

3.2 Comparing the Solutions Performance

To train and test the Support Vector Machines (SVMs) [18], we use the SVMFu version 2.001[23]. We train SVMs using the grayscale values and a polynomial kernel of degree 2. Fig. 3 provides the results for the classifier SVMs trained with different database and tested on Set1. In this figure, we use only the initial data set of CBCL (No-Light), and the initial database together with others 1000 face

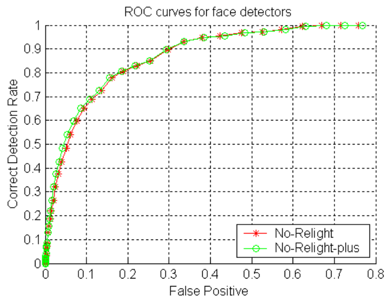


Fig. 3. The ROC curves by adding training face samples and being tested on Set1.

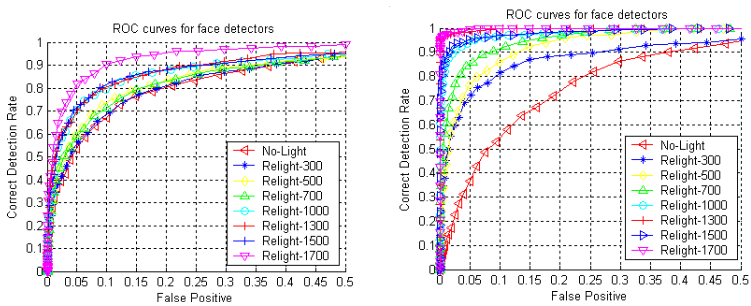


Fig. 4. The ROC curves of the trained SVM tested on the test sets. (a) The detect results on Set1; (b) the detect results on Set2.

samples collected from webpage (called No-Light-plus here). It means the No-Light has 2429 face samples, while No-Light-plus has 3429 face samples. The test set is the Set1 as discussed in Section 3.1. From these Receiver Operating Characteristic (ROC) curves in Fig. 3, we can find that the performance improvement of No-Light-plus is much limited compared with that of the No-Light. That is to say only expanding the number of the training set without considering the lighting variations can not improve the performance of the detector distinctly.

Fig. 4 provides the results for the classifier SVMs trained with others different database and tested on the testing sets. In these figures, we use only the initial data set of No-Light, and part of the initial database together with others different number face samples, which have been relighted by the methods demonstrated in section 2. Herein, No-Light is those face samples of CBCL database; Relight-300 means we substitute 300 relighting face samples for the same number of samples of CBCL database; and the same is of Relight-500, 700, . . . , 1700. Note all of these eight cases have the same face samples (that is to say 2429 positive examples.). The trained classifier SVMs, by these eight different positive sample sets and the same negative samples of the CBCL database, are tested on Set1 and Set2. The results are illustrated in Fig. 4. One can find that the presented method can improve the performance of the classifier. Comparing with the former results illustrated in Fig. 3, we can conclude that the relighting samples are more useful to train the face detectors under varying lighting conditions than simply adding the number of the training samples. However, the improvement of the performance on Set2 is more distinct than that of on Set1. It is because more lighting variations of Set2 contribute to it. It also demonstrates that this scheme will be more efficient for those images with diverse lighting conditions, which is just the problem of other face detectors. From these ROC curves, one can conclude that the performance improvement will decrease with the increasing of the substituted samples, for example, from No-light to Relight-300 compared with from Relight-1500 to Relight-1700. It may be that the certain number of lighting samples can represent the possible lighting variations and it makes this scheme more practical. We get the same results by using different original samples to do these experiments. That means the results are only related to the lighting conditions we relight the samples while it has little relations to the samples themselves.

3.3 Evaluation of the Generated Samples

Considering that the solutions performance of the relighting samples is evaluated by the classifier SVM above, they may favor this classifier. In order to verify that the solutions are independent to any special classifier, we use the relighting training set to train another classifier and test its generalization performance. AdaBoost has been used in face detection and is capable of processing images extremely rapidly while achieving high detection rates [19]. Therefore, we use the AdaBoost algorithm to train a classifier. A final strong classifier is formed by combining a number of weak classifiers. For the details of the AdaBoost based classifier, please refer to [3].

To compare the performance improvement on different training sets, we also use two different face training sets. The face-image database consists of 6,000 faces (collected form Web), which cover wide variations in poses, facial expressions and also in lighting conditions. To make the detection method less sensitive to affine transform, the images are often rotated, translated and scaled [4, 6, 8, 9, 14, 21]. Therefore, we randomly rotate these samples up to $\pm 15°$, translate up to half a pixel, and scale up to $\pm 10\%$. We get 12,000 face images. And these 12,000 face images constitute the first group *No-Relighting*. The second group *Relighting* is composed of 10,000 face images as above-mentioned and 2,000 relighting samples by the proposed method.

The non-face class is initially represented by 10,000 non-face images. Each single classifier is then trained using a bootstrap approach similar to that described in [17] to increase the number of negative examples in the non-face set. The bootstrap is carried out several times on a set of 12,438 images containing no faces.

The resulting detectors, trained on the two different sets, are evaluated on the MIT+CMU frontal face test set which consists of 130 images showing 507 upright faces [14]. The detection performances on this set are compared in Fig. 5.

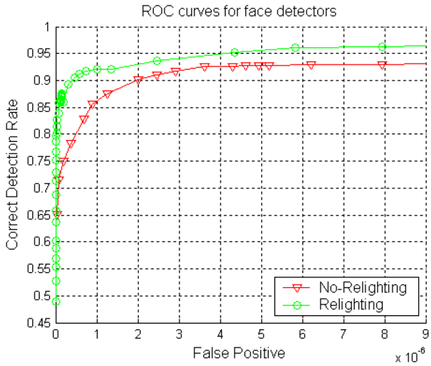
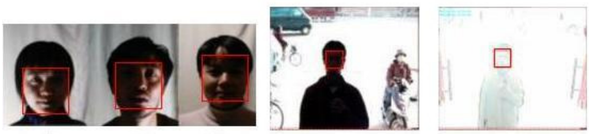


Fig. 5. The ROC curves of the AdaBoost-based classifiers are tested on test sets.



(a) (b) (c) (d) (e)

Fig. 6. Some face detection results; (a), (b), (c) are detected and cropped from CAS-PEAL database; (d), (e), (f), (g) from the practical applications.

From the ROC curves one can find that we get the detection rate of 90.8% and 15 false alarms with the detector trained on the set *Relighting*. Viola reported a similar detection capability of 89.7% with 31 false detects (by voting) [19]. However, different criteria can be used to favor one over another, which will make it difficult to evaluate the performance of different methods even though they use the same benchmark data sets [20]. Some results of this detector are shown in Fig. 6.

4 Conclusions

In this paper, we present a novel method to relight face training set. This scheme can improve the hit rates of the classifier under the diverse lighting conditions. The experiment results based on SVM and AdaBoost-based classifiers demonstrate the generation performance and its independence on the specific classifier.

Acknowledgements. This research is partially sponsored by Natural Science Foundation of China under contract No.60332010 and No.D070601-01, National Hi-Tech Program of China (No. 2001CCA03300, 2001AA114160, 2001AA114190 and 2002AA118010), and ISVISION Technologies Co. Ltd.

References

1. R. Basri and David W. Jacobs. Lambertian Reflection and Linear subspaces. IEEE transactions on Pattern Analysis and Machine Intelligence Vol. 25, No. 2. Feb. 2003.
2. J. Chen, X. L. Chen, W. Gao. Expand Training Set for Face Detection by GA Re-sampling. FG2004. pp. 73-79.
3. Y. Freund, R. E. Schapire. A decision-theoretic generalization of online learning and an application to boosting. Computational Learning Theory, 1995. pp. 23-37.
4. B. Froba and A. Ernst. Fast Frontal-View Face Detection Using a Multi-Path Decision Tree. Proc. AVBPA2003, pp. 921-928.
5. W. Gao, B. Cao, S. G. Shan, D. L. Zhou, X. H. Zhang, D. B. Zhao. The CAS-PEAL Large-Scale Chinese Face Database and Evaluation Protocols. Technical Report. 2004. (http://www.jdl.ac.cn/peal/JDL-TR-04-FR-001.pdf)
6. B. Heisele, T. Poggio, and M. Pontil. Face Detection in Still Gray Images. CBCL Paper #187. Massachusetts Institute of Technology, Cambridge, MA, 2000.
7. K. C. Lee, J. Ho, D. Kriegman. Nine Points of Lights: Acquiring Subspaces for Face Recognition under Variable Illumination. CVPR2001, pp. 519-526.
8. S. Z. Li, L. Zhu, Z.Q. Zhang, A. Blake, H. J. Zhang, and H. Shum. Statistical Learning of Multi-View Face Detection. In Proceedings of the 7th ECCV. 2002.
9. C. J. Liu. A Bayesian Discriminating Features Method for Face Detection. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, June 2003. pp. 725-740.
10. J. Miao, B.C. Yin, K.Q. Wang, et al., A Hierarchical Multiscale and Multiangle System for Human Face Detection in a Complex Background Using Gravity-Center Template, Pattern Recognition 32(7), 1237-1248, 1999
11. P. J. Phillips, P. Grother, R.J. Micheals, et. al.. FRVT 2002: Evaluation Report, http://www.frvt.org/DLs/FRVT_2002_ Evaluation_Report.pdf, March 2003.
12. L. Y. Qing, S. G. Shan, X. L. Chen. Face Relighting for face recognition under generic illumination. ICASSP2004 (Accepted).
13. R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object. Journal of the Optical Society of America. Vol.18, no.10, pp.2448-2459, 2001.
14. H. A. Rowley, S. Baluja, and T. Kanade. Neural Network-Based Face Detection. IEEE Tr. Pattern Analysis and Machine Intel, vol. 20, 1998, pp. 23-38.
15. H. Schneiderman and T. Kanade. A Statistical Method for 3D Object Detection Applied to Faces. Computer Vision and Pattern Recognition, 2000. pp. 746-751.
16. A. Shashua and T. Riklin-Raviv. The Quotient Image: Class-Based Re-Rendering and Recognition With Varying Illuminations. IEEE TPAMI, pp.129-139, 2001.
17. K. K. Sung, and T. Poggio. Example-Based Learning for View-Based Human Face Detection. IEEE Trans. on PAMI Vol.20. , No. 1, 1998. pp. 39-51.
18. V. N. Vapnik. Statistical Learning Theory. John Wiley & Sons, 1998.
19. P. Viola and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. Conf. Computer Vision and Pattern Recognition, 2001. pp. 511-518.
20. M. H. Yang, D. Kriegman, and N. Ahuja. Detecting Faces in Images: A Survey. IEEE Tr. Pattern Analysis and Machine Intelligence, vol. 24, Jan. 2002. pp. 34-58.
21. M. H. Yang, D. Roth, and N. Ahuja. A SNoW-Based Face Detector. Advances in Neural Information Processing Systems 12, MIT Press, 2000. pp. 855-861.
22. L. Zhang, D. Samaras. Face Recognition under Variable Lighting using Harmonic Image Exemplars. Proc. CVPR'2003, Vol. I, pp. 19-25, 2003.
23. http://www.csie.ntu.edu.tw/~cjlin/libsvm/
24. http://www.ai.mit.edu/projects/cbcl/software-dataset/index.html.