Face Detection under Variable Lighting Based on Resample by Face Relighting

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

[1](Multimedia and Intelligent Software Technology Laboratory,
Beijing University of Technology, Beijing 100022, China)
[2](School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China)
[3] (Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China)
E-MAIL: ymli@jdl.ac.cn; chenjie@jdl.ac.cn;
lyqing@jdl.ac.cn; ybc@bjut.edu.cn; wgao@jdl.ac.cn

Abstract:

Different environment illumination has a great impact on face detection. In this paper, we present a solution based on face relighting technology. The basic idea is that there exist nine harmonic images that can be derived from a 3D model of a face, and by which we can estimate the illumination coefficient of any face samples. Using an illumination radio image, we can produce images under new lighting conditions. To detect faces under certain lighting condition, we relight original face samples to get more new faces under kinds of possible lighting condition and add them to the training set. Our experimental results on Support Vector Machine (SVM) turned out that the relighting subspace is effective on detection under variations of the lighting conditions. Moreover, if we relight original face samples to new samples under different illuminations, the collected example set will be multiplied. We use the expanded 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. The later experiment also verifies the generalization capability of the proposed method.

Keywords:

Face detection; illumination; relighting; SVM; harmonic images; Lambertian surface.

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 [19]. Recently, the emphasis has been laid on data-driven learning-based techniques. [11] [12] [14]. [13] [15] [18]. However, different environment illumination has a great effect on face detection [19]. To build a robust and efficient face detection system, the problem of lighting variation is one of the main technical challenges facing system designers.

There has been much work dealing with illumination variation in face recognition. The appearance based methods such as Eigenface [7] and Fisherface have also been used in face detection [19]. The image space of a 3D Lambertian surface has better result in face recognition [5][10]. Recently, using spherical harmonics and techniques from signal-processing, Basri [1] and Ramamoorthi [3] have got the analytic nine dimensional lighting space of a convex Lambertian surface expressed in terms of spherical harmonics, considering attach-shadows. By observing that the Lambertian kernel contains only low frequency components, they deduce that the first nine (low frequency) spherical harmonics capture more than 99% of the reflection energy. Using this nine-dimensional linear subspace, a straightforward recognition scheme can be developed and results obtained in [1] are excellent. Later, they proposed the way to render images under any lighting condition [2].

In this paper, we propose a method for face relighting under any certain illumination and use it for face detection. Based on the face relighting model proposed by Basri and Ramamoorthi, L.Y. Qing has proposed the method to relighting faces under the pre-defined canonical illumination to get its canonical form[4]. Using the illumination coefficient of original face and sample faces, we can get more face samples under any lighting condition. We use this in face detection under certain lighting condition and in enlarging the training set of face. The method is easy to realize and doesn't influence the speed of any face detection system.

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

2. Face relighting with spherical harmonics

From the above discussion, assuming a face is a convex Lambertian surface, we get the image of the face:

$$I(x,y) = \rho(x,y)\vec{n}(x,y)\vec{s}\,, \qquad (1)$$

where ρ (x,y) is the albedo associated with point *(x, y)* in the image, $\vec{n}(x,y)$ is the surface normal direction and $\vec{s}$ is the point light source direction and whose magnitude is the light source intensity.

It has been shown in [1] and [4] that the BRDF of Lambertian surface is a low-pass filter for a curved convex diffuse object in a distant isotropic lighting field. Therefore the lower nine spherical harmonics can approximate its irradiance environment map well:

$$\begin{aligned} E(\theta,\phi) &\approx \sum_{l=0}^{2}\sum_{m=-l}^{l} E_{lm}Y_{lm}(\theta,\phi) \\ &= \sum_{l=0}^{2}\sum_{m=-l}^{l} A_l L_{lm} Y_{lm}(\theta,\phi) \end{aligned}\,, \qquad (2)$$

where A_l is spherical harmonics coefficients of the diffuse reflection function. The analytic form of A_l and more details of deriving the space-frequency reflection equation are given in [1], [4].

2.1 Lighting estimation

In order to analyze with spherical harmonics easily, we use a 3D generic mesh to get the information of the geometric shape of the face. Given a 2D image, to build the correspondence between the vertices of the 3D mesh and the 2D image, we use ASM method to get feature points and align these points between the mesh and the 2D image. Then the rest of the vertices are aligned with image warping technique. An example of the feature points automatically labeled is given in Fig. 1. Because human faces share similar 3D shape, the warping results are good enough to be used in the experiment.

Fig. 1. An example of Feature points labeling. The face image is from CAS-PEAL face database [20].

Since the irradiance E is dominated by low frequency components of lighting, we need only estimate the lower nine spherical harmonic coefficients L_{lm}.

Given an input image I (a column vector of n elements, n is the number of pixels in the image) of an object with constant albedo ρ, we can get the following equations:

$$I_{org}(x,y) = \rho(x,y)\sum_{l=0}^{2}\sum_{m-=-l}^{l} \Lambda_l A_l L_{lm} Y_{lm}(\theta,\phi)\,, \qquad (3)$$

where L_{lm} is the illumination coefficients.

By solving the least squares problem, we get the nine illumination coefficients scaled with the albedo, ρL, which approximate the illumination [4].

2.2. Samples relighting under certain lighting condition

As discussed above, given a face image, we can estimate its lighting condition using nine illumination coefficients. By these coefficients, it is just forward to relight face samples to new face images under such illumination with ratio image. The illumination ratio image is independent of the texture. Illumination ratio image between the original image and the relighting image is

$$r = \frac{I_{org}}{I_{new}} = \frac{\rho\vec{n}\cdot\vec{s}_{org}}{\rho\vec{n}\cdot\vec{s}_{new}} = \frac{\vec{n}\cdot\vec{s}_{org}}{\vec{n}\cdot\vec{s}_{new}}\,, \qquad (4)$$

where I_{new} means the new pixel value after relighting, $\vec{s}_{org}$, $\vec{s}_{new}$ mean the original and new light source directions respectively.

We have estimated $\vec{s}_{new}$ using nine illumination coefficients and with the same way we can also get the value of $\vec{s}_{org}$. Therefore, the relighting new image can get by following equation:

$$\begin{aligned} I_{new} &= \frac{I_{org}}{r} = \frac{\vec{n}\cdot\vec{s}_{new}}{\vec{n}\cdot\vec{s}_{org}} I_{org} \\ &\approx \frac{\sum_{l=0}^{2}\sum_{m=-l}^{l} E^{o}{}_{lm}Y_{lm}(\theta,\phi)}{\sum_{l=0}^{2}\sum_{m=-l}^{l} E^{n}{}_{lm}Y_{lm}(\theta,\phi)} I_{org} \\ &= \frac{\sum_{l=0}^{2}\sum_{m-=-l}^{l} A_l L^{o}{}_{lm}Y_{lm}(\theta,\phi)}{\sum_{l=0}^{2}\sum_{m-=-l}^{l} A_l L^{n}{}_{lm}Y_{lm}(\theta,\phi)} I_{org} \end{aligned}\,, \qquad (5)$$

where L^{o}_{lm}, L^{n}_{lm} means nine illumination coefficients of original image and the new image respectively.

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 try to first warp all faces to the generic shape and then compute the ratio-image for relighting the face images. It is then easy to reverse 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. Some examples of the relighting results are given in Fig.2

Fig. 2. Face relighting samples: the first line is the original faces and the second line is the relighting faces under the same illumination condition.

3. Face detection with sample relighting

Using the face relighting method, we can get samples under specific illumination conditions and deal with the illumination problems in face detection to some extent.

3.1 Face samples relighting for face detection

In our experiment, we use a test set of 400 faces and 23573 non-faces. The 400 faces are collected under four different lighting conditions. The detection rate on this set is usual low, as showed in the following ROC curve. The faces are available on the JDL webpage [22].

Our first training set is 4000 faces and 5000 non-faces from internet. We preserve the original images for face relighting.

Take 4 images of different lighting condition from the test set and 400 faces from training set. Estimate the illumination coefficients of the 4 test images. Using the relighting method, relight the 400 faces to get 1600 new face samples. Add these new samples to 3000 old training faces to construct new training set.

(a) original sample after normalization

(b) Relighting Samples with only DC coefficient

(c) Relighting Samples under lighting condition-1

(d) Relighting Samples under lighting condition-2

Fig. 3. Face Relighting Samples.

Because of the detected outliers of ASM during morphing, some new face samples are invalid. We delete those samples and get 1249 new available faces. There are 4249 faces and 5000 non-faces in the second training set. We choose only 400 faces for relighting to avoid over-fitting.

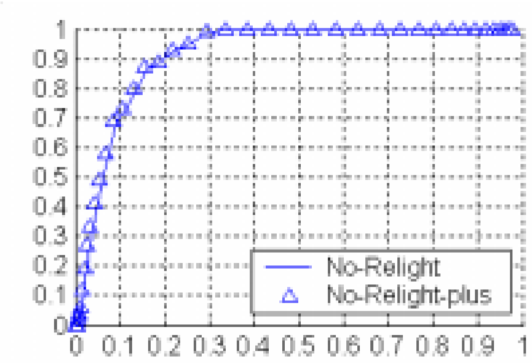
3.2. Comparing the solutions performance

We use SVM as a classifier and the gray value of the samples as feature. The samples are histogram normalized before feature abstraction to avoid the noise influence. We use LibSVM version 2.5[21] and trained the classifier with a RBF kernel.

Fig. 3 provides the results for the SVMs classifier trained on two different training sets. In this figure, *No-Relight* shows the result of the initial data set and *Relight* shows the result of the initial database together with the relighting samples. For the two cases, the trained classifiers are tested on the same testing set.

From these Receiver Operating Characteristic (ROC) curves in Fig. 3, we can find that the performance of *Relight* is much better than that of *No-Relight*.

Another experiment was done to verify this result from anther view. We add 3000 face samples from cmu training set[new] to the original 2429 face samples from mit training set. Also we add another 4000 non-face from cmu training set. So we get a new training set including 5429 face samples and 8548 non-face samples. Training on the samples, the result was show in Fig. 4.



The result shows that the performance don't get much improvement although the number of training set increased. Comparing with the former result, we can conclude that the relighting samples was more useful for the face detection under varying lighting conditions than simply adding number of training samples.

We get the same result using different original samples to do the experiment. That means the result is only related to the lighting condition we adding to the samples but have no relation with the samples themselves.

3.3. Evaluation of the generated samples

3.3.1. The AdaBoost-based classifier

Considering that all the solutions of relighting samples are evaluated by the classifier SVM during expanding, they may favor this classifier. In order to verify that the solutions are independent to any special classifier, we use the expanded 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 [17]. Therefore, we use the AdaBoost algorithm to train a classifier. A final strong classifier is formed by combining a number of weak classifiers which is described in Fig. 4 following the notation in [6]. For the details of the AdaBoost based classifier, please refer to [20].

- Given example set ***S*** and their initial weights ω_1;
- Do for t=1,……,T:
  1. Normalize the weights ω_t;
  2. For each feature, j, train a classifier h_j with respect to the weighted samples;
  3. Calculate error ε_t , choose the classifier h_t with the lowest error and compute the value α_t;
  4. Update weights ω_{t+1};
- Get the final strong classifier: $h(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$.

Fig. 4. The AdaBoost algorithm for classifier learning[17].

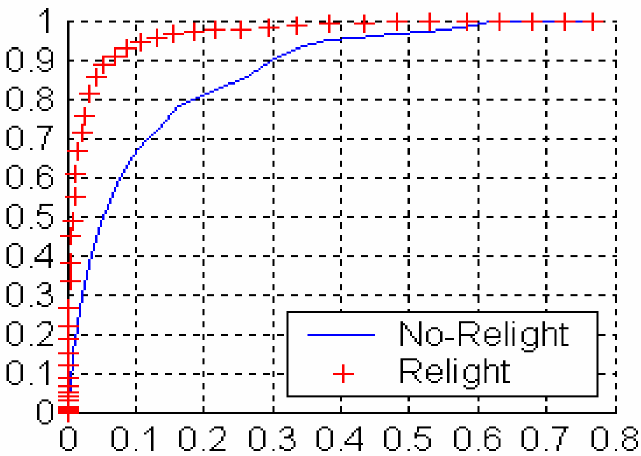


Fig. 3. The ROC curves on the test set

3.3.2. Training the detector.

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 [8, 9, 11, 12, 14, 19]. Therefore, we randomly rotate these samples up to ±15°, translate up to half a pixel, and scale up to ±10%., we get 12,000 face images. And these 12,000 face images constitute the first group *No-Relighting*. The second group *Relighting* is rendering from 1000 face samples and choose 40 different lighting conditions. We separate the 1000 face to 10 groups and each group relight different illumination. Together with original faces, we get 16,000 faces (4000 new faces).

The non-face class is initially represented by 5,000 non-face images. Each single classifier is then trained using a bootstrapping approach similar to that described in [19] to increase the number of images in the non-face set. The bootstrapping is carried out several times on a set of 8,736 images containing no faces.

3.3.3. Detection Results

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]. Some results are shown in Fig. 5.The detection performances on this set are compared in Fig. 5. From the ROC curves one can find that we get the detection rate of 89.8% and 22 false alarms with the detector trained on the set by *Relighting*. P. Viola reported a similar detection capability of 89.7% with 31 false detects (by voting) [17]. However, different criteria (e.g. training time, number of training examples involved, cropping training set with different subjective criteria, execution time, and the number of scanned windows in detection) 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 [19].

4. Conclusions

In this paper, we present a novel method to expand face training set by applying relighting technology on sample faces. The method can be used to deal with illumination problem in face detection. Unlike used in face recognition, face relighting under specific illumination based on harmonic images is used in face detection to produce more face samples under poor illumination. With one or several images under an illumination, we resample much more face samples under this illumination so that we can enhance the detection rate in the end. Also the method can enlarge our training set in general. The experiment results based on SVM and Adaboost-based classifier show that the relighting samples are available.

The current relighting images are depended on the shape information of face, therefore the result are influenced by the feature points labeling. The original illumination also has some effluence on the new relighting images. Better relighting technology will enhance the result again.

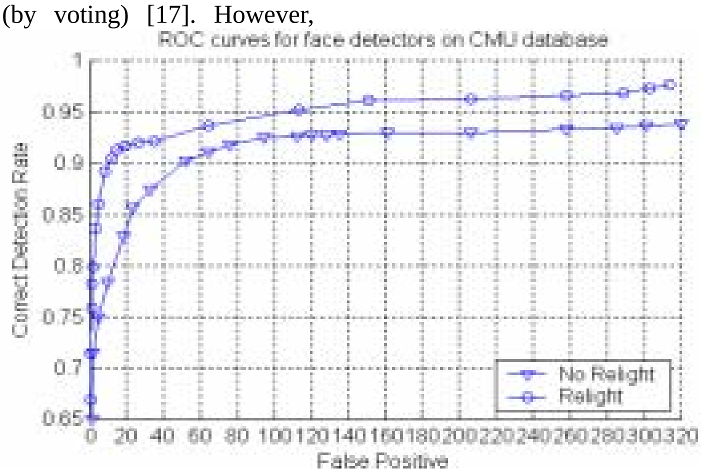
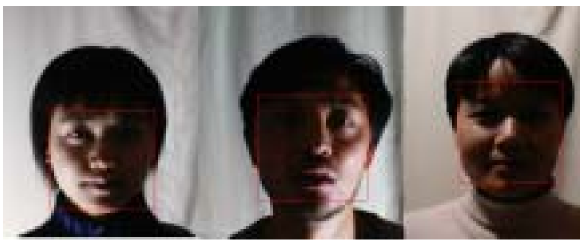


Fig.5. The ROC curves for our detectors on the MIT+CMU frontal face test set.

(a) (b) (c)



(d) (e)

(f) (g)

(h) (i)

Fig.6. Some face detection results; (a), (b), (c) are the results of the first experiment, the faces are from CSA-PEAL[22]; (the faces were cut form the original images) and (d), (e), (f), (g), (h), (i) from the practical application of the second experiment.

■ 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 D. Jacobs, "Lambertian reflectance and linear subspaces," Int. Conf. on Computer Vision, vol. 2, 2001, pp. 383 -390.

[2] 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.

[3] 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 A, vol.18, no.10, pp.2448-2459, 2001.

[4] Laiyun Qing, Shiguang Shan, Xilin Chen, "Face Relighting for face recognition under generic illumination.", icassp04.

[5] A. Shashua. "On photometric issues in 3D visual recognition from a single 2D image",Iernational Journal of Computer Vision, vol.21, no.1-2, pp.99-122, 1997.

[6] P. Belhumeur and D. Kriegman, "What is the set of images of an object under all possible lighting conditions," in Int. J. Computer Vision, vol. 28(3), 1998, pp. 245-260.

[7] M. Turk and A. Pentland, "Eigenfaces for Recognition",Journal of cognitive neuroscience, vol. 3, no. 1, pp71-86, 1991.

[8] B. Froba and A. Ernst, "Fast Frontal-View Face Detection Using a Multi-Path Decision Tree", Proc. Audio-and-Video-based Biometric Person Authentication (AVBPA '2003), 2003, pp. 921-928.

[9] B. Heisele, T. Poggio, and M. Pontil. Face Detection in Still Gray Images. CBCL Paper #187. Massachusetts Institute of Technology, Cambridge, MA, 2000.

[10] K. C. Lee, J. Ho, D. Kriegman. "Nine Points of Lights: Acquiring Subspaces for Face Recognition under Variable Illumination". IEEE Conf. On Computer Vision and Pattern Recognition, 2001, pp. 519-526.

[11] 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 European Conference on Computer Vision. 2002.

[12] 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.

[13] V. N. Vapnik, Statistical Learning Theory, John Wiley & Sons, 1998.

[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, On photometric issues in 3D visual recognition form a single image, Int. J. Computer Vision, vol. 21, pp. 99-122, 1997.

[17] 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.

[18] 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.

[19] 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.
[20] Y. Freund, R. E. Schapire, “A decision-theoretic generalization of online learning and an application to boosting”, Computational Learning Theory, 1995. pp. 23-37. Springer-Verlag.
[21] http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[22] http://www.jdl.ac.cn/peal/index.html