Asian Journal of Information Technology 4 (3): 49-55, 2005
Special Issue on Machine Learning in China - Selected Papers of CCML'04

Face Detection under Variable Lighting Based on Nine Point of Light

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

Abstract: Different environment illumination has a great impact on face detection and recognition. In this paper, we present a solution based on the nine points of light. The basic idea is that there exists a configuration of nine points light source directions which can be acquired by taking nine images of each individual under these single sources. Using this set of nine directions, we construct a linear subspace for each collected example by rendering it under these different lighting conditions. And then by sampling several examples randomly from the linear subspace, the collected example set can be multiplied. The multiplied sample set is used to train a Support Vector Machine (SVM), which is tested on a test set. It turns out that the resulting subspace is effective at detection under a wide range of lighting conditions. To verify the generalization capability of the proposed method, we also 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 also show that the data collection can be efficiently speeded up by the proposed methods.

Key words: Face detection, variable lighting, nine points of light, SVM

Introduction

Over the past ten years, face detection has been thoroughly studied in the research of computer vision 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 (Yang *et al.*, 2002). Recently, the emphasis has been laid on data-driven learning-based techniques (Li *et al.*, 2002; Liu, 2003; Rowley *et al.*, 1998; Rowley *et al.*, 1998; Schneiderman and Kanade, 2000; Sung and Poggio, 1998; Viola and Jones, 2001 and Yang *et al.*, 2000. However, different environment illumination has a great effect on face detection (Yang *et al.*, 2000). To build a robust and efficient face detection system, the problem of lighting variation is one of the main technical challenges facing system designers. In the past few years, many appearance-based methods have been proposed to handle this problem, and new theoretical insights have been reported in various publications, e.g. (Basri and Jacobs, 2001; Belhumeur and Kriegman, 1998 and Georghiades *et al.*, 2001). The main insight gained from these results is that there are both empirical and analytical justifications for using low dimensional linear subspaces to model image variations of human faces under different lighting conditions. Early work showed that the variations of the images of a Lambertian surface in a fixed pose, but under diverse lighting conditions so that no surface point is shadowed, is a three-dimensional linear subspace (Shashua, 1997). Under the Lambertian assumption, the set of images of an object under all possible lighting conditions forms the illumination cone, in the image space (Belhumeur and Kriegman, 1998). In a follow-up paper (Georphiades *et al.*, 1998) it was reported that the illumination cones of human faces can be approximated well by low-dimensional linear subspaces. More recently, by using the spherical harmonics and techniques from signal-processing, Basri and Jacobs have shown that for a convex Lambertian surface, its illumination cone can be accurately approximated by a 9-dimensional linear subspace (Basri and Jacobs, 2001). 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 the results obtained in (Basri and Jacobs, 2001) are excellent. In (Lee *et al.*, 2001), Lee etc. show that there exists a configuration of nine points light source directions such that by taking nine images of each individual under these single sources, the resulting subspace is effective at recognition under a wide range of lighting conditions. Since the subspace is generated directly from real images, potentially complex intermediate steps such as PCA and 3D reconstruction can be completely avoided and provide good recognition results.
The rest of this paper is organized as following: first the details of nine points of light are introduced, then the experiment results are reported and finally the conclusions are given.

Nine Points of Light

Harmonic Images: In this section, we briefly summarize the work presented in (Lee *et al.*, 2001). From the above

discussion, it follows that the set of all possible images of a convex Lambertian object under all lighting conditions can be well approximated by nine "harmonic images", which are formed under lighting conditions specified by the first nine spherical harmonics. Knowing the object's geometry and albedos, these harmonic images can be synthesized based on the ray-tracing, a standard technique.

Expressed in terms of (x, y, z), each spherical harmonic $Y_{lm}(x, y, z)$ is a polynomial in (x, y, z) of degree l. The first nine spherical harmonics in the Cartesian coordinates are:

$Y_{00} = 0.2821$ (1)

$(Y_{11}; Y_{10}; Y_{1-1}) = 0.4886\ (x;y;z)$; (2)

$(Y_{21}; Y_{2-1}; Y_{1-1}) = 1.093\ (xz, yz, xy)$; (3)

$Y_{20} = 0.3154\ (3z^2 - 1)$ (4)

$Y22 = 0.5462\ (x^2-y^2)$ (5)

Let H denote the linear subspace generated by the harmonic images and let C denote the illumination cone (Belhumeur and Kriegman, 1998). The volume of the intersection $C \cap H$ should be large. It is then natural to find whether there exits another 9-dimensional linear subspace R which is also good for face recognition. And a basis R of 9-dimensional linear subspace for a good face recognition method is provided in (Lee *et al.*, 2001).

Computing the Linear Subspace R and the Nine Points of Light: R should satisfy the following two conditions (Lee *et al.*, 2001):

1. Minimize the angular distance between R and H;
2. Maximized the (unit) volume $C \cap R$. (The unit volume is defined as the volume of the intersection of $C \cap R$ with the unit ball).

In order to calculate conveniently, an R is computed as a sequence of nested linear subspaces $R_0 \subseteq R_1 \subseteq \ldots \subseteq R_i \ldots \subseteq R_9 = R$, where R_i $(i=0, 1,\ldots 9)$ is a linear subspace of dimension i and $R_0 = \varnothing$. The main steps are:

First, define EC_i as the set obtained by deleting i extreme rays from EC, and $EC_0 = EC$, where the set EC denotes the set of sample points on the hemisphere. R_i and EC_i are defined inductively.

Second, suppose that we have defined (or computed) R_{i-1} and EC_{i-1}. The sets EC_i and R_i are defined iteratively as follows:

$$x_i = \arg\max_{x \in EC_{i-1}} \sum_{k=1}^{l} \frac{dist(x^k, R_{i-1}^k)}{dist(x^k, H^k)} \quad (6)$$

For each x EC, x_k denotes the image of model k taken under a single light source with direction x, and k indexes the available face models.

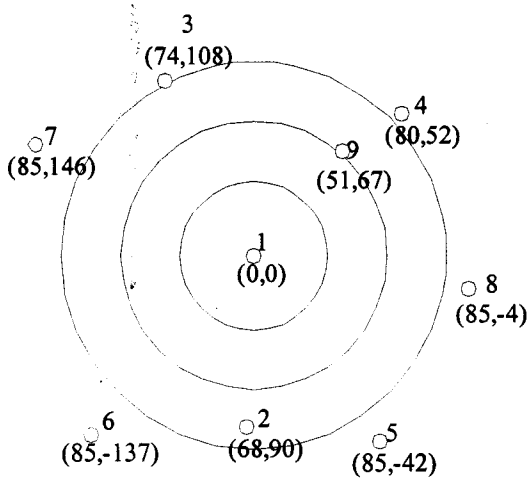


Fig. 1: The schematic of the universal configuration of nine light source directions with all 200 sample points (the projection from the hemisphere (ϕ, θ) onto the xy-plane in polar coordinates (r, θ) is: $\phi \rightarrow r$, $\theta \rightarrow \theta$). The cricles reprsent ϕ = 25°, 50° and 75°, respectively on the hemisphere. Where ϕ is the elevation angle (angle between the polar axis and the z-axis) and θ is the azimuth angle

Fig. 2: Some generated face samples

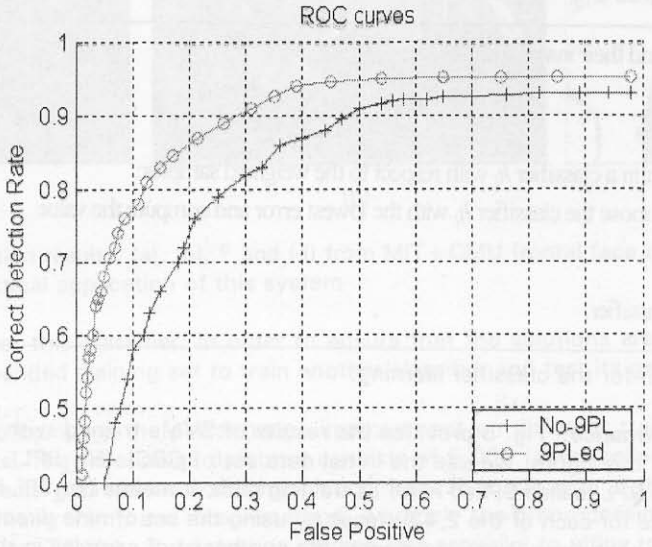


Fig. 3: The ROC curves on the testset.

Third, R_i is spanned by xi and R_{i-1}, and the set EC_i is defined as $EC_{i-1}|x_i$.
The algorithm terminates after $R_9 = R$ is computed.
Let H_k denote the harmonic plane of model k and R^k_{i-1} represents the linear subspace spanned by the images ($x^k_{i'}$...xki) of model k under light source directions (x_1,....,x_i). Note that dist(x, H) is always non-zero for all x ECi

since H is the harmonic plane. When computing $R1$, we define dist(x,R_0) = dist(x, .) to be 1. Therefore, the first element x_1 is the extreme ray in C that is closest to the harmonic plane H.
The resulting configuration of the nine directions is called the universal configuration. These directions, along with the 200 samples on the hemisphere, are {(0, 0), (68, 90), (74, 108), (80, 52), (85, 42), (85, 137), (85, 146), (85, 4), (51, 67)}, which are shown in Fig. 1.

Experiment results

Face-Samples Preprocessing by 9PL: The data set consists of a training set of 6,977 images (2,429 faces and 4,548 non-faces) and a test set of 24,045 images (472 faces and 23,573 non-faces). The images were 19 x 19 grayscale. The data is available on the CBCL webpage (Zhou and Jiang, 2003).
Using the set of nine directions, we construct a linear subspace for each of the 2,429 faces by rendering it under these lighting conditions. In practice, the nine images should be real; however, due to the lack of samples, we have opted for rendering instead. We call this method the Nine Points of Light (9PL) method. Some results of the 9PL operations are shown in Fig. 2.

Learning with Support Vector Machines: The basic theory of SVMs (Vapnik, 1998). SVMs can perform pattern recognition for two-class problems. A non-linear transformation Φ(□) can be applied if the data is not linearly separable in the input space. This transformation maps the data points x $\in R^n$ of the input space into a high (possibly infinite) dimensional space R^p, which is called feature space. In the SVM classifier scheme, the mapping Φ(□) is implemented by a kernel function K(□□) which defines an inner product in R^N, i.e. K(x,t) = Φ(x)• Φ(t). The decision function of the SVM can be denoted in the form:

$$f(x)=\sum_{i=1}^{l}\alpha_i y_i K(x_i,x) \qquad (7)$$

where l is the number of data points in the training set, and $y_i \in$ (-1, 1) is the class label of the data point x_i. The coefficients a_i in Eq. (7) is solved by a quadratic programming problem (Vapnik, 1998). And then in this linear space, we can determine the separating hyper-plane with maximum distance to the closest points of the training set. These points are called support vectors. The optimal hyper-plane can separate these data in the feature space easily.
To train and test Support Vector Machines (SVMs), we use SVMFu version 2.001 (http://www.ai.mit.edu/projects/cbcl/ software-dataset/ index.html.). We trained SVMs using the grayscale values and a polynomial kernel of degree 2.

- Given example set S and their initial weights ;
- Do for t=1,……,T:
  1. Normalize the weights ;
  2. For each feature, j, train a classifier h_j with respect to the weighted samples;
  3. Calculate error , choose the classifier h_t with the lowest error and compute the value ;
  4. Update weights ;
- Get the final strong classifier: .

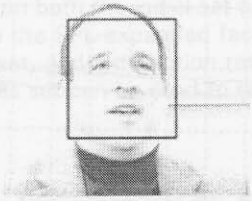
Fig. 4: The AdaBoost algorithm for the classifier learning.

Comparing the solutions performance: Fig. 3 provides the results of SVMs trained with different databases and tested with the testing set. In this figure, we use the initial data set of CBCL (No_9PL), and the initial database together with the results of the 9PL (called 9PLed here) as training data. It means No_9PL has 2,429 face samples, while 9PLed is a linear subspace for each of the 2,429 faces by using the set of nine directions and rendering each face image under these lighting conditions. Here, we double the number of of samples in the latter set. For the two cases, the trained classifiers are both tested on the testing set.
Form these Receiver Operating Characteristic (ROC) curves in Fig. 3, we can find that the performance of 9PLed is much better than that of the No_9PL.

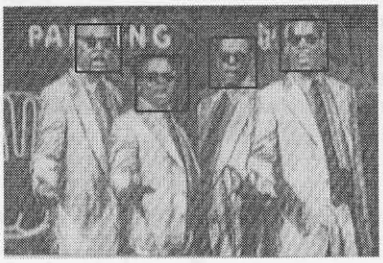
Evaluation of the generated samples:

The AdaBoost-based classifier: Considering that all the results of 9PLed are evaluated by the classifier SVM during

(a)



(b)

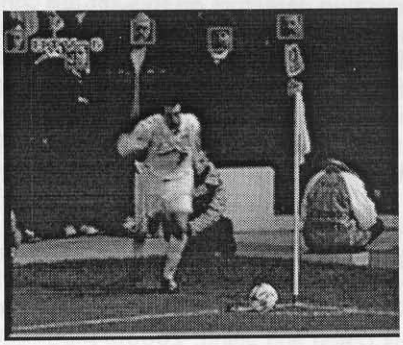


(c)

(d)

(e)

(f)



(g)

(h)

Fig. 5: Some face detection results; (a), (b), © and (d) from MiT + CMU frontal face test set while (e), (f), (g) and (h) from the practical application of this system

expanding, they may favor this classifier. In order to ensure that the solutions are independent to any special classifier, we use the expanded training set to train another classifier and test its generalization performance.

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 lighting conditions. First, we align all the collected images to reduce the extrinsic variations among them, and in our method, we apply the preprocessing proposed by Rowley *et al.* (1998) to align face samples. To make the detection method less sensitive to affine transform, we randomly rotate the samples up to $\pm 15°$, translate them up to half a pixel, and scale them up to $\pm 10\%$. Then, histogram equalization is performed, which maps the intensity values to expand the range of intensities. After these preprocessing, we get 12,000 face images, which constitute the first group No_9PL. The second group 9PLed is also a linear subspace for each of the 12,000 faces by using the set of the nine directions to render each face image. And we also double the number of the latter training set.

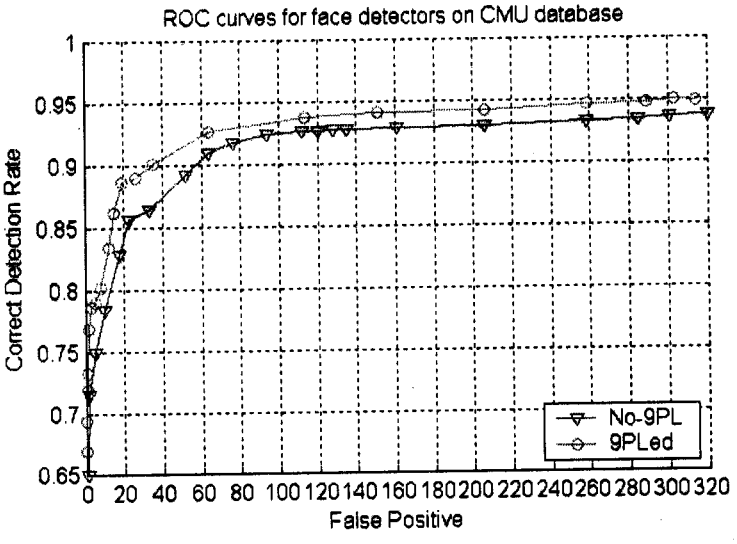


Fig. 6: the ROC curves for our detectors on the MIT + CMU frontal face test set.

The non-face class is initially represented by 12,000 non-face images. Each single classifier is then trained using a bootstrapping approach similar to that described in (Sung and Poggio, 1998) to increase the number of images in the non-face set. The bootstrapping is carried out several times on a set of 10,526 images containing no faces.

Results

The resulting detectors, trained on the two different sets separately, are evaluated on the MIT + CMU frontal face test set which consists of 130 images showing 507 upright faces (Rowley *et al.*, 1998). Some results are shown in Fig. 5.The detection performances of the two detectors on this set are compared in Fig. 6. 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 9PLed. P. Viola reported a similar detection capability of 89.7% with 31 false detects (by voting) (Viola and Jones, 2001). 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 method over another which will make it difficult to evaluate the performances of different methods even though they use the same benchmark data sets (Yang *et al.*, 2002).

Discussion

In fact, some recent works in machine learning have shown that under the certain conditions, using examples generated virtually to enhance the training set could benefit the trained classifiers (Zhou and Jiang, 2003 and Zhou and Jiang, 2004). The main reasons are discussed as following: The error rates (denoted by PE(C) here) result from a classifier can be broken into three terms (Zhou and Jiang, 2003). The first term, denoted by PE (C°) here, derives from the limited learning ability of the classifier. That is to say the trained classifier might still make some errors in prediction although the training set contains no noise and captures the whole target distribution. However, the value of PE (C°) may be extremely small in general. The second term results in the noised training set, denoted by Bias (C) here. The last term results in the finite training sample, denoted by Var (C) here. It is because the collected training set is always hard to capture fully the target distribution. Therefore, the errors of a classifier can be decomposed as:

$$PE^{©} = PE\,(C°) + Bias^{©} + Var\,(C) \qquad (8)$$

The proposed method tries to generate more samples to capture the target distributions because the original training set might be much noised and might not capture fully the target distribution. By the means the error rate Var (C) of the trained classifier can be decreased. In turns, the error rate PE (C) is lowered. And such an approach has already been used in face recognition with small samples (Chen *et al.*, 2004).

Conclusions

In this paper, we present a novel method to enrich face sample set by applying a configuration of nine points light source directions. Using this set of nine directions, we construct a linear subspace for each collected example by rendering it under these different lighting conditions. These new generated samples can simulate the wide variations of faces in different lighting conditions. We use some face samples without 9PL and the same samples with 9PL to train a SVM detector respectively, and compare their performances on an independent test set. The

performances of the detector trained on both the initial set and the results of 9PL are much better than that trained only on the initial set. Finally, we use the 9PL-expanded face set to train an AdaBoost-based classifier and test it on the MIT + CMU frontal face test set, and a detection rate of 89. 8% is achieved with only 22 false alarms. It shows that the expanded face samples set can be used to train classifiers other than SVM and can further improve the performance of the classifier.

Acknowledgement

This research is partially sponsored by NSFC (contract No. 60332010) and the National Hi-Tech Program of China (No.2001AA114190, No.2002AA118010, and 2003AA142140). This work is also sponsored by ISVISION Technologies Co., Ltd.The authors also acknowledge Prof. Z.-H. Zhou for his comments and suggestions to improve this paper.

References

Basri, R. and D. Jacobs, 2001. " Lambertian reflectance and linear subspaces," in Int. Conf. on Computer Vision, 2: 383–390.

Belhumeur, P. and D. Kriegman, 1998. "What is the set of images of an object under all possible lighting conditions," in Int. J. Computer Vision, 28: 245–260.

Chen, S., D. Zhang and Z. H. Zhou, 2004. Enhanced $(PC)^2A$ for face recognition with one training image per person. Pattern Recognition Letters, 25: 1173-1181.

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

Freund, Y. and R. E. Schapire, 1995. A decision-theoretic *generalization of online learning and an application to boosting*. In Computational Learning Theory. pp: 23–37

Georghiades, A., D. Kriegman and P. Belhumeur, 1998. " Illumination cones for recognition under variable lighting: Faces," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition.

Georghiades, A., D. Kriegman and P. Belhumeur, 2001. "From few to many: Generative models for recognition under variable pose and illumination," in IEEE Transactions on Pattern Analysis and Machine Intelligence, pp: 643–660.

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

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

Li, S. Z., L. Zhu, Z.Q. Zhang, A. Blake, H. J. Zhang and H. Shum, 2002. Statistical Learning of Multi-View Face Detection. In Proceedings of *the 7th European Conference on Computer Vision*. 2002.

Liu. C. J., 2003. A Bayesian Discriminating Features Method for Face Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 25: 725-740.

Vladimir N. Vapnik, 1998. Statistical Learning Theory. John Wiley & Sons.

Rowley, H. A., S. Baluja and T. Kanade, 1998. Neural Network-Based Face Detection. IEEE Tr. Pattern Analysis and Machine Intel., 20: 23-38.

Rowley, H. A., S. Baluja and T. Kanade, 1998. Rotation Invariant Neural Network-Based Face Detection. Conf. Computer Vision and Pattern Rec., pp: 38-44.

Schneiderman, H. and T. Kanade, 2000. A Statistical Method for 3D Object Detection Applied to Faces. Computer Vision and Pattern Recognition, pp: 746-751.

Shashua, A., 1997. " On photometric issues in 3D visual recognition form a single image," *Int. J. Computer Vision*, 21: 99–122.

Sung, K. K. and T. Poggio, 1998. Example-Based Learning for View-Based Human Face Detection. IEEE Trans. on PAMI 20: 39-51.

Viola, P. and M. Jones. 2001. Rapid Object Detection Using a Boosted Cascade of Simple Features. Conf. Computer Vision and Pattern Recognition, pp: 511-518.

Yang, M. H., D. Roth and N. Ahuja, 2000. A SNoW-Based Face Detector. Advances in Neural Information Processing Systems 12, MIT Press, pp: 855-861.

Yang, M. H., D. Kriegman and N. Ahuja, 2002. Detecting Faces in Images: A Survey. *IEEE Tr. Pattern Analysis and Machine Intelligence*, 24: 34-58.

Zhou, Z. H. and Y. Jiang, 2003. Medical diagnosis with C4.5 rule preceded by artificial neural network ensemble. IEEE Transactions on Information Technology in Biomedicine, 7: 37-42.

Zhou, Z. H. and Y. Jiang, 2004. NeC4.5: neural ensemble based C4.5. IEEE Transactions on Knowledge and Data Engineering, 16: 770-773.

http://www.ai.mit.edu/projects/cbcl/ software-dataset/ index.html.