Novel Face Detection Method Based on Gabor Features

Jie Chen[1], Shiguang Shan[2], Peng Yang[2], Shengye Yan[2], Xilin Chen[1] and, Wen Gao[1,2]

[1] School of Computer Science and Technology,
Harbin Institute of Technology, 150001, China
[2] ICT-ISVISION JDL for AFR, Institute of Computing Technology,
CAS, Beijing, 100080, China
{jchen, sgshan, pyang, syyan, xlchen, wgao }@jdl.ac.cn

Abstract. Gabor-based Face representation has achieved great success in face recognition, while whether and how it can be applied to face detection is rarely studied. This paper originally investigates the Gabor feature based face detection method, and proposes a coarse-to-fine hierarchical face detector combining the high efficiency of Harr features and the excellent discriminating power of the Gabor features. Gabor features are AdaBoosted to form the final verifier after the cascade of Harr-based AdaBoost face detector. Extensive experiments are conducted on several face databases and verified the effectiveness of the proposed approach.

1 Introduction

Over the past ten years, face detection has been thoroughly studied in computer vision research for its interesting applications, such as video surveillance, human computer interface, face recognition, and face image database management etc. 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]. Recently, the emphasis has been laid on data-driven learning-based techniques, such as [5], [8], [9], [10], [12], [13], [14], [15] and [19]. All of these schemes can be found in the recent survey by Yang [20]. After the survey, the methods based on boosting are much researched. Viola described a rapid object detection scheme based on a boosted cascade of simple features. It brought together new algorithms, representations and insights, which could broader applications in computer vision and image processing [16]. Boosting is simple and easy to implement. It has been proven that Boosting minimizes an exponential function of the margin over the training set. However, a strong classifier learnt by AdaBoost is suboptimal for the applications in terms of the error rate [1]. Therefore, some improved versions are developed. Li et al. proposed a FloatBoost-based algorithm to guarantee monotonicity of the sequential AdaBoost learning and developed a real-time multi-view face detection system [7]. C. Liu et.al developed a general classification framework called Kullback-Leibler Boosting [6]. Xiao et.al proposed a boosting chain algorithm, which can combine the boosting classifiers into a hierarchy “chain” structure [18].

In face recognition, the Elastic Bunch Graph Matching (EBGM) has attracted much attention because it firstly exploited the Gabor transform to model the local features

S.Z. Li et al. (Eds.): Sinobiometrics 2004, LNCS 3338, pp. 90–99, 2004.

of faces [17]. However, EBGM takes the complete set of Gabor features and then most of them are redundant for classification. For examples, Fasel pointed out in [3] that the Gabor features used in [17] are not the best ones for the representation of facial landmarks. In [4], Huang et.al proposed a classification-based face detection method using Gabor filter features. But they utilized only the Gabor features. It would decrease the detection speed of the resulting detector.

This paper is an attempt to solve this question by introducing the AdaBoost method into the Harr+Gabor feature-based face detection detector. That is to say, the resulting detector will consist of weak classifiers based on the Harr and Gabor features and is boosted by AdaBoost. AdaBoost selects a set of Harr features as the first part of the cascade detector to increase the speed and a small set of available Gabor features from the extremely large set as the second part of the cascade detector to decrease the false alarms of the detector. The final strong classifier, which combines a few hundreds of weak classifiers (Harr+Gabor features), is evaluated on the test set.

The remaining part of this paper is organized as following: In section 2, the Gabor representation of faces is introduced. Section 3 describes the boosting learning for features selection and classifiers construction. Experiments and analyses are demonstrated in section 4, followed by a small discussion. Conclusions and future work are presented in section 5.

2 Gaborface

Gabor filter can capture salient visual properties such as spatial localization, orientation selectivity, and spatial frequency characteristics. Considering these excellent capacities and its great success in face recognition [11], we choose Gabor features to represent the face image besides the Harr-like features. Gabor filters are defined as following:

$$\psi_{u,v}(z) = \frac{\| k_{u,v} \|^2}{\sigma^2} e^{\left(-\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2} \right)} \left[e^{i\vec{k}_{u,v} z} - e^{-\frac{\sigma^2}{2}} \right], \tag{1}$$

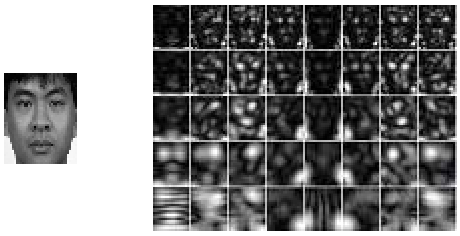
where $k_{u,v} = k_v e^{i\phi_u}$; $k_v = \frac{k_{\max}}{f^v}$ gives the frequency; $\phi_u = \frac{u\pi}{8}$, $\phi_u \in [0, \pi)$ gives the orientation; and $z = (x, y)$;

$$k_{u,v} = k_v e^{i\phi_u}, \tag{2}$$

where $e^{i\vec{k}_{u,v} z}$ is the oscillatory wave function whose real part and imaginary part are cosine function and sinusoid function, respectively. In equation 1, v controls the scale of Gabor filters, which mainly determines the center of the Gabor filter in the frequency domain; u controls the orientation of the Gabor filter.

In our experiment we use the Gabor filters with the following parameters [7]: five scales $v \in \{0,1,2,3,4\}$ and eight orientations $u \in \{0,1,2,3,4,5,6,7\}$ with $\sigma = 2\pi$, $k_{\max} = \pi/2$, and $f = \sqrt{2}$.

The Gaborface, representing one face image, is computed by convoluting it with corresponding Gabor filters. Figure 1 shows the Gaborface representation of a face image.



(a) (b)

Fig. 1. Gaborface representation for one face

3 AdaBoost-Based Detector Training

As we know, boosting learning is a strong tool to solve two-class classification problems. Noticing the great success of AdaBoost in face detection area, we exploited it in our method to combine weak classifiers into a final strong one.

We use AdaBoost to select a small set of Harr or Gabor features (or weak classifiers) from the original extremely high dimensional Harr or Gabor feature space to form a strong classifier. A strong classifier learned by AdaBoost, is formed by:

$$S(x) = \sum_{t=1}^{T} \alpha_t h_t(x), \tag{3}$$

where a_m is the combining coefficient and $h_t(x)$ is a threshold function. How to derive a_m and $h_t(x)$ will be discussed in the following section.

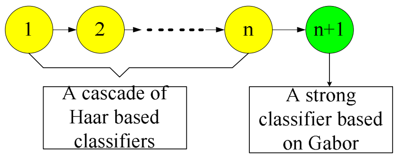


Fig. 2. A cascade classifier, combined by AdaBoost, has n+1 stages. The first n stages are n strong classifiers and each are composed of several weak classifiers based on Harr features. The $(n+1)^{st}$ stage is also a strong classifier and is composed of several weak classifiers but based on Gabor features

A large number of experimental studies have shown that classifiers combination can exploit the discriminating power of individual feature sets and classifiers. With the success of boosting in the application of face detection, boosting, as one of the most commonly used methods of combining classifiers based on statistical re-sampling techniques, boosting has shown its strong ability to discriminate the two-class problem. AdaBoost, a version of the boosting algorithm, is taken to solve this two-class problem. Herein, we use AdaBoost to train a strong classifier. A strong classifier is formed by AdaBoost, which combines a number of weak classifiers as shown in Fig.2.

Table 1. The AdaBoost algorithm for classifier learning

Given labeled examples set S and two threshold, δ_1, δ_2 :

- $i = 1$; The first stage error rate $F_1 = 1$;
- Train the first n-stage strong classifier based on the Harr feature.
  - ❈ While the first i stages error rate $F_i > \delta_1$
    1) Train a strong classifier by AdaBoost based on the Harr features;
    2) Calculate the (i+1)st stage error rate e_{i+1};
    3) The first (i+1)st stages error rate $F_{i+1} = F_i \times e_{i+1}$ and $i = i + 1$;
  - ❈ Combine these n stages into a cascade detector
- Train the (n+1)st stage strong classifier based on the Gabor features.
  - ❈ While the (n+1)st stage error rate e_{n+1}, and $e_{n+1} \times \delta_1 > \delta_2$
    1) For each Gabor feature, *k,* train a classifier h_k with respect to the weighted samples as based on the Harr-like features;
    2) Calculate error rate
  - ❈ Combine all of these weak classifiers based on the Gabor features into a strong classifier.
- Combine the first n stages and the (n+1)st stage strong classifier into a resulting cascade detector.

The AdaBoost process is described in table 1. $S(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$ of table 1 is demonstrated as equation (3), where $\alpha_m \geq 0$ is the combining coefficient which is learned as in the table 1.

4 Experiments and Analyses

4.1 Performance Comparison

To compare the representative ability of Harr and Gabor features for discriminating face and non-face, we use two group feature sets. The first group consists of only Harr features while the second is composed of only Gabor features. 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 are 19×19 grayscale and re-normalized to 20×20. The data are available on the CBCL webpage [22]. Two different classifiers based on different feature set are trained on this training set as demonstrate in [16]. The detection performances on this test set are compared in Fig. 3. From the ROC curves one can find that the classifier based on the Gabor features outperforms the one based on the Harr distinctly. That is to say the Gabor features can describe the difference between the faces and non-faces more efficient than the Harr features.

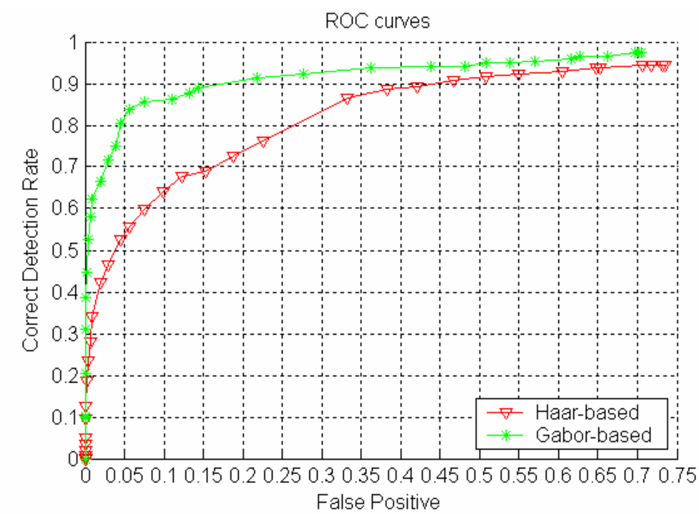


Fig. 3. The ROC curves for our detectors on the MIT face test set which are available on the CBCL webpage

4.2 Training the Detector

To compare the detector's performance improvement based on different feature set further, we use two group feature sets. The first group consists of only Harr features while the second is composed of both Harr and Gabor features. The first classifier is trained as described in [16] and the second is trained as shown in table 1. The face-image database consists of 6,000 faces (collected form Web), which cover wide variations in poses, facial expressions and lighting conditions. After the preprocessing by GA operations [2], we get 100,000 face images. For negative examples we start with 100,000 non-face examples from 16,536 images of landscapes, trees, buildings, etc. Although it is extremely difficult to collect a typical set of non-face examples, the bootstrap [15] is used to include more non-face examples during training.

In our experiment, all of the face samples are normalized to 20×20 and we use the Gabor filters with five scales and eight orientations. The number of Gabor features of each sample is $20\times20\times5\times8=16000$, from which the training algorithm would select tens of the most discriminant ones to form the last stage of the cascade detector as demonstrated in Fig 2 and table 1. To the second classifier, we run AdaBoost in 22 stages, a total of 4231 rounds, and got 4231 features. In this cascade, the first 21 stages are composed of weak classifiers based on Harr features with 4135 features while the last stage consists of 96 weak classifiers based on Gabor features.

4.3 Detection Results

The resulting detectors, trained on different features, are evaluated on the MIT+CMU frontal face test set, which consists of 130 images showing 507 upright faces [12]. The detection performances on this set are compared in Fig. 4. From the ROC curves one can find that we get the detection rate of 90.37% and 8 false alarms with the detector trained based on the Harr+Gabor feature set. P. Viola reported a similar detection capability of 89.7% with 31 false detects (by voting) [16]. 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 [20]. Some detected results on these test sets are shown in Fig. 5.

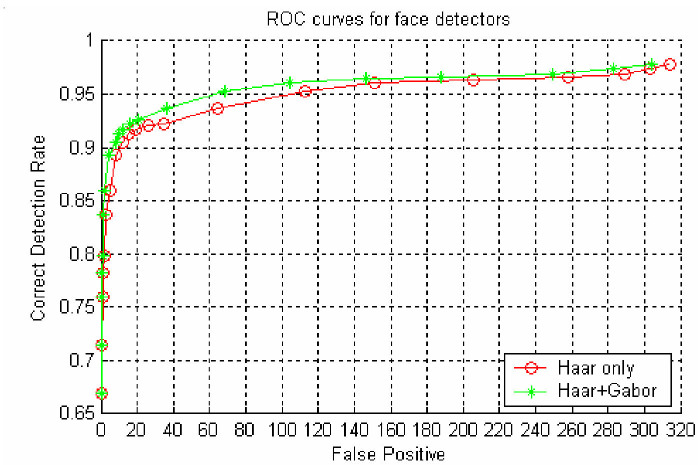


Fig. 4. The ROC curves for our detectors on the MIT+CMU frontal face test set

From the ROC curves we can also conclude that the detector trained on the Harr+Gabor feature set outperforms the detector trained only on the Harr feature set. The possible reason is that the Gabor feature can represent the difference between the face and non-face more efficient than the Harr feature only. Therefore, some false alarms by the classifier based on the Harr features in first part of the cascade are corrected by the classifier based on the Gabor features in the last stage.

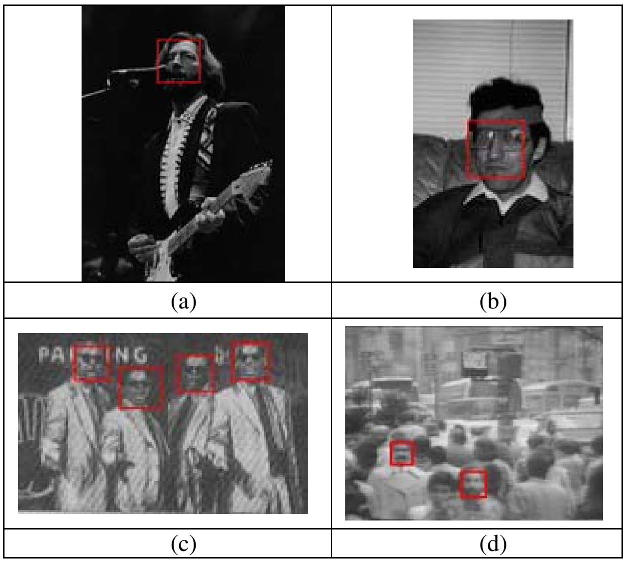
The detector trained on the Harr+Gabor feature set is also tested on the other test set except that on the MIT+CMU frontal face test set. This test set is selected from the CAS-PEAL Face Database, which includes large-scale face images with different sources of variations, especially poses, expressions, accessories, and lighting [21]. The CAS-PEAL face database contains 99,594 images of 1,040 individuals (595 males and 445 females) with varying Pose, Expression, Accessory, and Lighting (PEAL). For each subject, 9 cameras spaced equally in a horizontal semicircular shelf are mounted to simultaneously capture images across different poses in one shot.

Each subject is also asked to look up and down to capture 18 images in another two shots. It also considered 5 kinds of expressions, 6 kinds accessories (3 glasses, and 3 caps), and 15 lighting directions. A subset containing 24,018 images of 1,040 subjects is selected from this database because our detector is a frontal face detector. The detection results are listed in table 2. The frontal sub-directory consists of those images with only one frontal face in each image. The pose sub-directory is composed of those images with one face in multi-view but within 30^{o} in each image. The sub-directory PD is those images with one face looking down; the sub-directory PM is those images with one face looking horizontally; the sub-directory PU is those images with one face looking up.

Table 2. Detection rates from various sub-directory on Set1

Data Set			Faces	False alarms	Detection rates
Results in each sub-directory	Frontal		9029	66	96.42%
	POSE (within 30^{o})	PD	4998	18	94.74%
		PM	4993	35	99.78%
		PU	4998	135	98.06%
Total			24018	254	97.08%

From the table 2, one can find that the detector successfully detects 23,324 faces from the 24,018 images (each image contains only one face) with only 254 false



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

Fig. 5. Face detection on the MIT+CMU frontal face test set

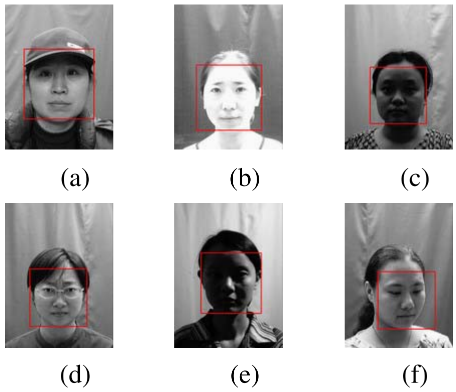


Fig. 6. Face detection results on some images from CAS-PEAL. (a) Accessories. (b), (c), (d), (e): lighting, multiple lamps and lanterns are used to cover varying lighting conditions, (f): pose

detects. The highest hit rate is in the PM sub-directory because all faces in this file are frontal and have fewer variations than that of the frontal sub-directory, which has more poses, expressions, accessories, and lighting variations. It is these variations that lead to more false alarms and less hit rates in the frontal sub-directory. From this table, one can also find the images with looking-down faces are a little harder for the detector to locate than the images containing looking-up faces. It may be that the looking-down faces have more effect on the eyes and therefore make it difficult to distinguish these faces from the background. Fig. 6 shows examples of the detected faces, where a square indicates a face region successfully detected. Note that the resolution of the images is 640×480, and the faces are detected at different scales.

5 Conclusion

This paper originally investigates the possibility to apply Gabor features to face detection motivated by the success of Gabor feature in face identification and verification area. Our experiments have shown that Gabor feature-based representation does have more powerful discriminating capability. However, Gabor feature is too high dimensional and its computation is too time-consuming. Therefore, we further proposed a coarse-to-fine hierarchical face detector by combining the popular Harr features and Gabor features using the cascade AdaBoosting method, in which the time-consuming Gabor-based face verifier accomplishes the final validation of few candidates. Extensive experiments are conducted on CMU+MIT face database to evaluate the efficiency and effectiveness of the framework, which shows that only tens of Gabor features are enough for decreasing the false alarm rate with an acceptable detection speed.

Our future work will focus on how to improve the face detection framework based on Harr+Gabor by trading off the efficiency and accuracy. Also, how to extract the few Gabor feature rapidly should also be considered.

Acknowledgement

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, National Hi-Tech Program of China (No. 2001AA114190, 2002AA118010 and 2003AA142140), and ISVISION Technologies Co., Ltd.

References

1. Buhlmann P. and Yu B.: Invited discussion on Additive logistic regressions: a statistical view of boosting (Friedman, Hastie and Tibshirani). *The Annual of Statistics*, 28(2): (2000) 377-386
2. Chen J., Gao W.. Expand Training Set for Face Detection by GA Re-sampling. *The 6th IEEE International Conference on Automatic Face and Gesture Recognition (FG2004)*. (2004) 73-79
3. Fasel I. R., Bartlett M. S, and Movellan J. R: A comparison of Gabor filter methods for automatic detection of facial landmarks. In Proceedings of the 5th International Conference on Face and Gesture Recognition, 2002
4. Huang L.-L., Shimizu A., and Kobatake H.: Classification-Based Face Detection Using Gabor Filter Features. The 6th IEEE International Conference on Automatic Face and Gesture Recognition (FG2004). (2004) 397- 402
5. Hsu R. L., Abdel-Mottaleb M., and Jain A. K.: Face detection in color images. IEEE Trans. Pattern Anal. Machine Intell., vol. 24, (2002).696–706
6. Liu C., Shum H. Y.: Kullback-Leibler Boosting. Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03). 2003.
7. Li S. Z., Zhu L., Zhang Z.Q., Blake A., Zhang H. J., and Shum H.: Statistical Learning of Multi-View Face Detection. In Proceedings of the 7th European Conference on Computer Vision. 2002.
8. Liu C. J.: A Bayesian Discriminating Features Method for Face Detection. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, June (2003) 725-740.
9. Osuna E., Freund R., and Girosi F.: Training support vector machines: An application to face detection. Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition, June 1997, pp. 130–136.
10. Papageorgiou C. P., Oren, M. and Poggio T.: A general framework for object detection. in Proc. 6th Int. Conf. Computer Vision, Jan. 1998, pp.555–562.
11. Phillips P. J., Moon H., Rauss P., and Rizvi. SA.: The FERET Evaluation Methodology for Face-Recognition Algorithms. Proceedings of Computer Vision and Pattern Recognition, Puerto Rico, (1997) 137-143
12. Rowley H. A., Baluja S., and Kanade T.: Neural Network-Based Face Detection. IEEE Tr. Pattern Analysis and Machine Intel. vol. 20, (1998) 23-38.
13. Rowley H. A., Baluja S., and Kanade T.: Rotation Invariant Neural Network-Based Face Detection. Conf. Computer Vision and Pattern Rec., (1998) 38-44.
14. Schneiderman H. and Kanade T.: A Statistical Method for 3D Object Detection Applied to Faces. Computer Vision and Pattern Recognition, (2000) 746-751.
15. Sung K. K., and Poggio T.: Example-Based Learning for View-Based Human Face Detection. IEEE Trans. on PAMI Vol.20. , No. 1, (1998) 39-51.
16. Viola P. and Jones M.: Rapid Object Detection Using a Boosted Cascade of Simple Features. Conf. Computer Vision and Pattern Recognition, (2001) 511-518.

17. Wiskott L., Fellous J., Kruger N., von der Malsburg C.: Face recognition by elastic bunch graph matching. IEEE Trans. PAMI, vol. 19, no. 7, (1997) 775-669,.
18. Xiao R., Li M. J., Zhang H. J.: Robust Multipose Face Detection in Images. IEEE Trans on Circuits and Systems for Video Technology, Vol.14, No.1 (2004) 31-41.
19. Yang M. H., Roth D., and Ahuja N.: A SNoW-Based Face Detector. Advances in Neural Information Processing Systems 12, MIT Press, (2000) 855-861.
20. Yang M. H., Kriegman D., and Ahuja N.: Detecting Faces in Images: A Survey. IEEE Tr. Pattern Analysis and Machine Intelligence, vol. 24, Jan. (2002) 34-58.
21. http://www.jdl.ac.cn/peal/index.html
22. http://www.ai.mit.edu/projects/cbcl/ software-dataset / index.html.