

Multi-View Face Detection under Complex Scene based on Combined SVMs

Peng Wang , Qiang Ji
Department of Electrical, Computer and System Engineering
Rensselaer Polytechnic Institute
Troy, NY, 12180

Abstract

A single face classifier has difficulty in detecting multi-view faces under real and complex scenes due to various poses, cluttering environment and small size of faces. In this paper, we propose a novel combination of SVMs to detect multi-view faces, using both cascading and bagging methods. In our method, the faces are divided into seven views. Each of them models a typical pose under complex scenes. By the modified bootstrap method applied in our method, a cascade of SVMs are constructed to quickly select face candidates from image with expected accuracy. Bagging of different SVMs can further eliminate the false detections that are difficult to handle by single SVM. Such combination of SVMs can effectively detect multi-view faces even with large rotation angles and heavy shadow. The experiment results show better accuracy and generalization performance over single classifier.

1 Introduction

Face detection is a two-class classification problem, which is to discriminate face patterns from background. Many factors, as concluded in [15], contribute to difficulties in face detection. Face patterns themselves include many elements that vary greatly with different persons and environment, such as face pose, skin color, face components and facial expression. The multiple poses make the assumption of convexity in feature space invalid. Furthermore, the unpredicted noise and illumination condition in the complex scene cause a lot of false detections.

According to [15], face detection methods are generally classified as knowledge based, feature invariant based, template matching and statistical appearance based methods [12, 7, 11]. In the real application where the faces are captured in some distance, those facial feature based methods have difficulties since that the face component, such as eye and mouth, and facial expression are blurred, shown as Fig. 3. Under such condition, the methods based on the whole

face appearance are more robust to noise and face variance.

Under complex environment, a single face classifier has limitations to detect multi-view faces. Combination of weak classifiers helps to improve both accuracy and generalization capability of a single classifier. Two methods, cascading and bagging, are well applied in face detection. In [14], critical features for face detection are selected through Adaboosting method. Then cascade of detectors are constructed to discard false face candidates before they go to next detector. A modified Adaboosting method, S-Adaboost, is also proposed to pick up the patterns hard to classify in complex environment [4]. Bagging method combines different classifiers together based on replicated training samples to improve the robustness and accuracy of a single classifier [3, 1].

Support vector machine (SVM) is a linear classifier in high dimensional feature space with minimal structural risk. SVM has already been successfully applied in face detection [8, 9]. However, problems occur when applying single SVM to detect multi-view faces. Since the profile faces and frontal faces have much different appearance, they are not convex in original feature space. Since SVM originally deals with two-class problem, the optimization in SVM become very difficult when all the face patterns with different poses are fused together.

One solution is to apply multiple SVMs for multiple poses, as in [6]. In the real scene, the face poses vary greatly so that many SVMs are needed to identify those poses. To reduce the complexity and improve the generalization performance while preserving accuracy, we propose the combination of SVMs, using cascading and bagging, to detect multi-view faces.

In our proposed method, a slightly revised bootstrap method is used to train a cascade of SVMs with expected detection rate. The face candidates are quickly pruned by removing most of non-face patterns. The remaining false detections, which are hard to classify by single SVM, are further eliminated by nonlinearly bagging SVMs. The experiments show that such combination can effectively improve detection rate and reduce false detection.

The paper is organized as follows. Section 2 introduces face detection based on SVM. Section 3 introduces our representation of multi-view faces and the combining methods used. Last two sections give experiment results and conclusion.

2 SVM based Face Detection

Since support vector machine(SVM) was proposed by Cortes and Vapnik [2], it has been successfully applied in many pattern recognition problems. SVM is a linear classifier in high dimension space with minimal structural risk. Suppose the pattern is given as (\mathbf{x}, y) , where \mathbf{x} is the feature vector and $y \in \{-1, 1\}$ is the class label, the linear decision function in SVM is defined as

$$f(\mathbf{x}) = \text{sgn}(\mathbf{w} \cdot \mathbf{x} + b) \quad (1)$$

Weight \mathbf{w} is the linear combination of Support Vectors(SVs)

$$\mathbf{w} = \sum_{i=1}^l \alpha_i \mathbf{x}_i \quad (2)$$

The decision boundary is determined by maximizing the structural margin between two classes. The optimization problem is equivalent to the quadratic programming problems [2]. ‘‘C-SVM’’ with soft margin hyperplane is introduced for the linearly inseparable case.

Kernel function in SVM plays the role of mapping feature vector to higher dimension space and dot production. Replace the dot products with kernel function $k(\mathbf{x}, \mathbf{x}_i)$, we obtain the linear discriminant function in high dimension feature space.

$$f(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^l \alpha_i y_i k(\mathbf{x}, \mathbf{x}_i) + b\right) \quad (3)$$

The most often used kernel functions in face detection are polynomial with degree d , Gaussian RBF and Multi Layer Perceptron [8]. A 2-d polynomial SVM is trained to detect frontal face in images. The face candidates are cut from images and preprocessed, then sent to SVM. The false detections are fed back to SVM and used as negative samples for later training. The single SVM shows good detection rate for high resolution images. However, its performance deteriorates for mixed-quality images (detection rate decreased to 74% with more false detections [8]).

To improve performance, combination of SVMs are more suitable for the complex pattern as multi-view faces under complex scenes. Five different types of SVMs are combined to improve accuracy [1]. The SVMs are trained with reduplicated samples. The detection result is the majority voting of those SVMs. The bagging results show reduction of false detections compared with single SVM. To

deal with multi-view faces, SVR is firstly applied to estimate face pose angle. Then specific pose SVM is selected to detect faces [6].

In next section, we propose a combination of SVMs incorporating both cascade and bagging to improve detection rate and generalization performance of single SVM.

3 Combination of SVMs to Detect Multi-View Faces

3.1 Multi-View Face Representation

The classification criterion that divides faces into two class, frontal and profile is very rough. In real environment, the face undergoes 3D motion. In most applications, the face rotation angles can not be recovered accurately from single 2-D image. Instead, the face angles are estimated from frontal and profile face detectors, such as in [13], or Support Vector Regression [6].

In our multi-view face detector, we divide face poses into seven views, shown as in Fig. 1. Each column in Fig. 1 represents a face view. Three views are frontal faces, i.e. vertical frontal, up frontal and down frontal faces. The profile faces are classified into four categories according to their approximate rotation angles.



Figure 1. Multi-view faces.

Our model of multiple views is suitable for faces under complex scene. Each of the seven poses approximates a typical rotation angle. In Fig. 1, The leftmost pose is with about left 90 degree. The faces in second left column have a smaller angle, approximately 45 degree. Each view also contains enough samples to handle face variance in the real scene, shown as faces in each column. From our experiments, classifier based on such pose model can achieve both good accuracy and generalization capability.

3.2 Cascading

For complex face detection, it is difficult to achieve both high detection rate and generalization capability in a single SVM. With more training samples, the number of support vectors increases dramatically. In our experiments, when training a 2-d polynomial SVM with data from real scene,

the number of SVs is about 600 for the first 5000 training samples. However, the number of SVs increases to 1500 when only 2000 more false detections are fed back. The reason is that the multi-view face patterns are not convex any more in feature space. When more non-face patterns are fed back, more support vectors are needed to construct the separate hyperplane. The increase of SVs is not affordable because it is very time and memory consuming.

Some methods, such as boosting, are proposed to obtain a “stronger” classifier from “weak” classifiers. Adaboosting is one of classical boosting methods in which the training samples are recursively chosen and gradually improve the accuracy of classifiers [10]. Actually since SVM is already a “strong” classifier when maximizing structural margin for all the training features, it is not a good candidate for boosting. In our method, we apply bootstrap method to train hierarchical SVMs so that they can quickly reject most non-face patterns while preserving face candidates under complex scenes.

- Given training set $S = (\mathbf{x}_i, y_i, p_i)$, where \mathbf{x}_i is the feature vector, $y_i \in \{-1, +1\}$ is the label of face and non-face samples. $p_i \in P = \{p^1, ..p^7\}$ indicates the face pose. It has no meaning for non-face samples.
- For $i = 1, \dots, m$,
 1. Initialization. Set $S_t = S$ and $S_r = \{\}$. Set expected detection rate θ_i and false rate η_i for i th SVM.
 2. Randomly select N training samples from S_t with appropriate ratio for different views. Move those samples to S_r .
 3. Train i th SVM using S_r .
 4. Test i th SVM using S_t , obtain detection rate θ_t and false rate η_t .
 5. Move those incorrectly classified patterns from S_t to S_r .
 6. If $\theta_{test} > \theta_i$ and false rate $\eta_{test} < \eta_i$, then train next SVM. Otherwise, repeat the above steps (3), (4) and (5).
- Cascade m SVMs to form hierarchical classifiers.

Figure 2. Bootstrap algorithm in our method

A slightly revised bootstrap method is applied here to obtain hierarchical SVMs and void over-fitting problems. Firstly, the face samples are selected from different views with appropriate ratio, which guarantees the cascading classifiers can detect face with multiple views. Instead of recursively changing the weight of training data and re-sample the whole data set, we only feed back those incorrectly classified face and non-face patterns so that SVM can quickly achieve high detection rate and moderate false rate, avoiding difficulty of optimization and increase of SVs. The al-

gorithm is illustrated as Fig. 2.

The hierarchical SVMs can can achieve high detection rate and low false rate. For example, in a three-layer cascade, suppose that each SVM achieves detection rate about 98% with false rate about 50%, 20% and 10%, the overall detection rate can be 94% with overall false rate of 1%.

3.3 Bagging

Due to the low resolution and complex background, there are still some false detections left after hierarchical SVMs. Bagging (“Bootstrap aggregating”) is another method to combine classifiers for better results [3].

The idea of bagging is to train multiple classifiers using randomly selected samples. Then majority voting of bagging results can obtain more robust classifier. In our methods, the training samples are also selected from bootstrap method stated before, different with reduplicatively selected training data as in [5].

SVM	Poly	Poly	RBF	Bagging
Parameters	$d = 3$	$d = 2$	σ^2	N/A
Number of SVs	870	862	784	N/A
Detection Rate	90.7%	89.5%	91.6%	92.5%
False Rate	1.59%	2.19%	1.40%	0.85%

Table 1. Test performance of bagging SVMs

We applies 14 SVMs with different kernel types and parameters in our bagging. The test performance of different types of SVMs and their bagging are given in Table 1. We can observe that compared with each single SVMs, bagging methods can improve the the detection rate and suppress more false detections.

4 Experiments

Thousands of images are captured for training and testing in the real scenes, such as laboratory, campus, shopping mall and other public areas. Training faces are cut by hand. They are more like “head” since more components, such as hair and ear, are included in our samples, as shown in Fig. 1. Non-face data is retrieved from background images. Totally we have about 3,000 face samples and above 8,000 non-face samples for training.

Typical faces in our images have small size. The size of head ranges from about 20 by 20 to 35 by 35. The facial features sometimes are blurred, or even not visible for the profile faces with large angle. All the face and non-face samples are resized to 23 by 23 gray scale images, then used to train combined SVMs.

In detection, the patches from image are firstly sent to hierarchical SVMs. Non-faces will be discarded as many

as possible while face patterns are preserved. The remaining patches are examined by bagging SVMs. The bagging SVMs eliminate those non-faces that are difficult to eliminate by single SVM.

Several segments of video captured in school and mall are reserved for testing our face detector. The frames that contain multiple persons with various poses are selected for testing. The detection results are shown in Table 2. The detection rate is approximately 93.6% for “school” sequence and 93.1% for “mall” sequence. Some face detection results are shown as Fig. 3, where the multiple faces with different poses are accurately located from the complex background and heavy shadow.

sequence	person number	detected faces	false detections
“school”	47	44	28
“mall”	58	54	12

Table 2. Face detection results

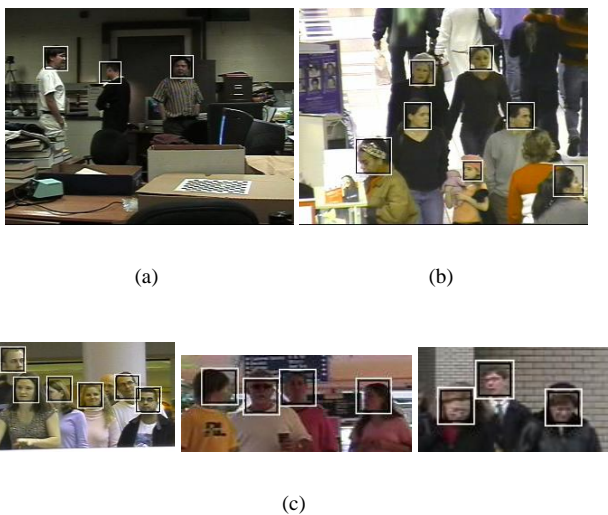


Figure 3. Some face detection results. (a),(b): detected faces in whole frames. (c): More results.

5 Conclusion

This paper proposes a combination of SVMs to detect multi-view faces in complex real scenes. The faces are represented by seven typical views. Different SVMs are combined together through cascade and bagging methods. Hierarchical SVMs can select multi-view face candidates from images quickly. The multi-view face patterns and hard-to-classify non-face patterns in complex environment are fur-

ther discriminated through nonlinear combination of bagging SVMs. The experiments show the accuracy and robustness of our method. Future work will focus on improving the computation efficiency of SVMs to achieve real-time performance.

References

- [1] L. Buciu, C. Kotropoulos, and I. Pitas. Combining support vector machines for accurate face detection. *ICIP 2001*, 1:1054–1057, 2001.
- [2] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3).
- [3] B. Draper and K. Baek. Bagging in computer vision. *CVPR*, pages 144–149, 1998.
- [4] J. L. Jiang and K.-F. Loe. S-adaboost and pattern detection in complex environment. *CVPR*, 1:413–418, 2003.
- [5] H.-C. Kim, S. Pang, H.-M. Je, D. Kim, and S. Y. Bang. Pattern classification using support vector machine ensemble. *ICPR*, 2:160–163, 2002.
- [6] Y. Li, S. Gong, and H. Liddell. Support vector regression and classification based multi-view face detection and recognition. In *Automatic Face and Gesture Recognition, IEEE International Conference on*, pages 300–305, 2000.
- [7] C. Liu. A bayesian discriminating features method for face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(6):725–740, 2003.
- [8] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. *CVPR*, pages 130–136, 1997.
- [9] H. Sahbi and N. Boujemaa. Coarse-to-fine support vector classifiers for face detection. *ICPR*, 3:359–362, 2002.
- [10] R. Schapire. A brief introduction to boosting. *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, 1999.
- [11] H. Schneiderman and T. Kanade. A statistical method for 3d object detection applied to faces and cars. *CVPR*, 1:746–751, 2000.
- [12] K.-K. Sung and T. Poggio. Example-based learning for view based human face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(1):39–51, 1998.
- [13] R. C. Verma, C. Schmid, and K. Mikolajczyk. Face detection and tracking in a video by propagating detection probabilities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(10):1216–1228, 2003.
- [14] P. Viola and M. Jones. Robust real-time object detection. *ICCV*, pages 747–747, 2001.
- [15] M.-H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.