# Real-Time Face Detection Using Boosting in Hierarchical Feature Spaces

Dong Zhang[1]
[1]IDIAP Research Institute
Rue du Simplon 4
1920 Martigny, Switzerland
zhang@idiap.ch

S.Z. Li[2] and Daniel Gatica-Perez[1]
[2]Microsoft Research Asia
No. 49, Zhichun Road, Haidian
100080 Beijing, China
szli@microsoft.com

## Abstract

*Boosting-based methods have recently led to the state-of-the-art face detection systems. In these systems, weak classifiers to be boosted are based on simple, local, Haar-like features. However, it can be empirically observed that in later stages of the boosting process, the non-face examples collected by bootstrapping become very similar to the face examples, and the classification error of Haar-like feature-based weak classifiers is thus very close to 50%. As a result, the performance of a face detector cannot be further improved. This paper proposed a solution to this problem, introducing a face detection method based on boosting in hierarchical feature spaces (both local and global). We argue that global features, like those derived from Principal Component Analysis, can be advantageously used in the later stages of boosting, when local features do not provide any further benefit. We show that weak classifiers learned in hierarchical feature spaces are better boosted. Our methodology leads to a face detection system that achieves higher performance than a current state-of-the-art system, at a comparable speed.*

## 1. Introduction

In pattern recognition terms, face detection is a two-class (face/non-face) classification problem. As the face manifold is highly complex, due to the variations in facial appearance, lighting, expressions, and other factors [5], face classifiers that achieve good performance are very complex.

The learning-based approach constitutes the most effective one for constructing face/non-face classifiers [4, 6]. Recently, Viola and Jones proposed a successful application of AdaBoost to face detection [10, 9]. Li *et al.* extended Viola and Jones' work for multi-view faces using an improved boosting algorithm [2]. Both systems achieved a detection rate of about $91\%$, and a false alarm rate of $10^{-6}$ for frontal faces, with real-time performance on $320\times240$ images. The real-time speed and good performance can be explained by two factors. First, AdaBoost learning algorithms are used for learning of highly complex classifiers. AdaBoost methods [1] learn a sequence of easily learnable weak classifiers, and boost them into a single strong classifier via a linear combination of them. Second, the real-time speed is achieved by an ingenuous use of techniques for rapid computation of Haar-like features [3, 10]. Moreover, the use of cascade structures [10] further speeds up the computations.

In spite of their evident advantages, existing systems have limitations to achieve higher performance because weak classifiers become too weak in later stages of the cascade. Current approaches use bootstrapping to collect non-face examples (false alarms) to re-train the detection system (e.g. as the input of the next layer in a cascade system). However, after the power of a strong classifier has reached a certain point, the non-face examples obtained by bootstrapping are very similar to the face patterns, in any space of the simple Haar-like features. It can be empirically shown that the classification error of Haar-like feature-based weak classifiers approaches $50\%$, and therefore boosting stops being effective in practice.

To address this problem, we propose a method in which boosted weak classifiers are learned in a hierarchy of feature spaces. The power of weak classifiers can be increased by switching between these spaces, from local to global features, to an extent that boosting learning is still beneficial. In particular, we show that Principal Component Analysis (PCA) coefficients are quite effective at discriminating between face and non-face patterns, when embedded in a boosting algorithm at its later stages, unlike local features that do not provide any benefit. Although more expensive in computational terms, global features can be used only at very late stages of a cascade system, not affecting the real-time requirement. The result is a face detection system with higher detection rate and lower false alarm rate than a state-of-the-art, single feature, Adaboost-based detection system. Our approach is illustrated in Fig.1.

The rest of paper is organized as follows: Section 2 describes Adaboost learning in the Haar-like feature space, and motivates our work based on the limitations of cur-

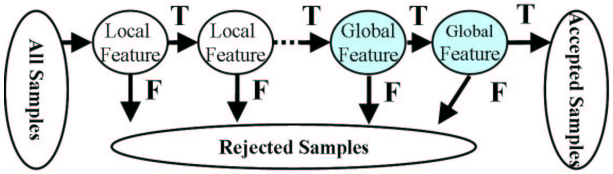


**Figure 1. Face detection framework**

rent methods. Section 3 introduces our approach. Results are presented in Section 4. Section 5 provides some concluding remarks.

## 2. Boosting in Haar-like Feature Space

The basic form of AdaBoost [1] is for two class problems. A set of $N$ labeled training examples is defined as $(x_1, y_1), \ldots, (x_N, y_N)$, where $y_i \in \{+1, -1\}$ is the class label for the example $x_i \in \mathbb{R}^n$. For face detection, $x_i$ is an image sub-window of fixed size containing an instance of the face or non-face. AdaBoost assumes that a procedure is available for learning a sequence of *weak classifiers* $h_m(x)$ $(m = 1, 2, \ldots, M)$ from the training examples A *strong classifier* is a linear combination of the $M$ weak classifiers,

$$H_M(x) = \sum_{m=1}^{M} \alpha_m h_m(x), \tag{1}$$

where $\alpha_m \geq 0$ are combining coefficients. The AdaBoost learning procedure is aimed to compute the sets of coefficients $\{\alpha_m\}$ and classifiers $\{h_m(x)\}$.

In the early stage of face detection, the weak classifiers, which perform simple classification, are derived based on histograms of four basic types of Haar-like features. A total of 45891 features can be derived for a sub-window of size $20 \times 20$, for all admissible locations and sizes. Such features can be computed very efficiently from the integral image defined in [10]. The task of face detection is to classify every possible sub-window. A vast number of sub-windows result from the scan of the input image. For efficiency reasons, it is crucial to discard as many non-face sub-windows as possible at early stages, so that as few sub-windows as possible are further processed by later stages.

However, the classification power of the described system is limited when the weak classifiers derived from simple local features become too weak to be boosted, especially in the later stages of the cascade training. Empirically, we have observed that when the discriminating power of a strong classifier reaches a certain point, *e.g.* a detection rate of $90\%$, and a false alarm rate of $10^{-6}$, non-face examples collected by bootstrapping become very similar to those of face examples in terms of the Haar-like features. The histograms of the face and non-face examples for any feature can hardly be differentiated, and the empirical

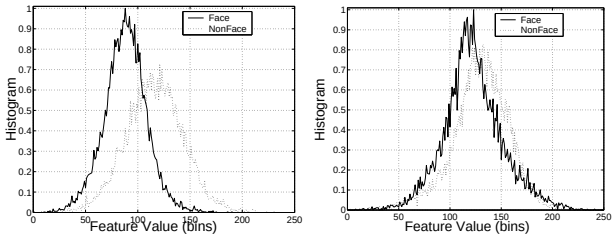


**Figure 2. Left: Empirical distribution of the face and non-face examples for the 5th Haar-like feature selected by boosting. The error rate is significantly lower than 50%. Right: Distribution for the 1648th Haar-like feature selected by boosting. The error rate is close to 50%.**

**Figure 3. The top five eigenfaces**

probability of misclassification of the weak classifiers approaches $50\%$. At this stage, boosting becomes ineffective because the weak learners are too weak to be boosted. This issue has been discussed in the past by Probably Approximately Correct (PAC) learning theory [8]. A specific example of this fact is illustrated in Fig.2. One way to address this problem is via the derivation of a stronger weak classifier in another feature space, which is more powerful and complementary to the local feature space. We propose to boost in PCA coefficient space. As we show in the next section, weak classifiers in this global feature space have sufficient classification power for boosting learning to be effective in the later stages of a cascade system.

## 3. Boosting in PCA Feature Space

When the local Haar-like features reach their limit, we would like to use another representation that is more discriminative between face and non-face examples. A fruitful alternative is to recourse to a global representation in the late stages of the cascade, such that these two feature spaces, one local and one global, complement each other.

Principal Component Analysis (PCA) is a classic technique for signal representation, used in the past for face recognition [7]. PCA can be summarized as follows. Given a set of face examples in $\mathbb{R}^N$ represented by column vectors, the mean face vector is substracted to obtain the vectors $\mathbf{x}_i \in \mathbb{R}^N$ $(i = 1, \ldots, m)$. The covariance matrix is then computed as $\mathbf{C} = \frac{1}{m} \sum_{j=1}^{m} \mathbf{x}_j \mathbf{x}_j^T$. Linear PCA diagonalizes the covariance matrix by solving the eigenvalue prob-

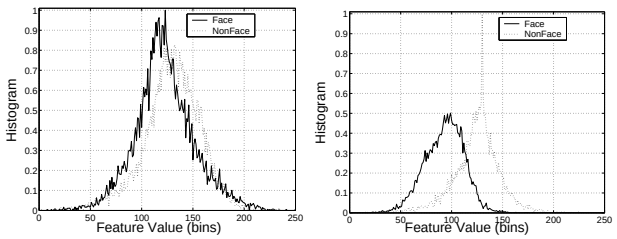


**Figure 4. Left: Empirical distribution of the face and non-face examples for the 1648th haar-like feature selected by boosting learning, whose error rate is almost 50% (same as Fig. 3(right)). Right: Distribution for the PCA features selected at the same stage of boosting; the error rate is significantly lower than 50%.**

lem $\lambda\mathbf{v} = \mathbf{C}\mathbf{v}$,

$$\lambda(\mathbf{x}_i \cdot \mathbf{v}) = (\mathbf{x}_i \cdot \mathbf{C}\mathbf{v}) \quad \forall i = 1, \ldots, m. \tag{2}$$

The eigenvalues are then sorted in descending order, and the first $M \leq N$ principal components $\mathbf{v}_k$ ($1 \leq k \leq M$) are used as the basis vector of a lower dimensional subspace, forming the transformation matrix $\mathbf{T}$ (Fig.3). The projection of a point $\mathbf{x} \in \mathbb{R}^N$ into the $M$-dimensional subspace can be calculated as $\theta = (\theta_1, \ldots, \theta_M) = \mathbf{x}^\top \mathbf{T} \in \mathbb{R}^M$. Its reconstruction from $\eta$ is $\hat{\mathbf{x}} = \sum_{k=1}^{M} \theta_k \mathbf{v}_k$, and constitutes the best approximation of the $\mathbf{x}_1, \ldots, \mathbf{x}_m$ in any $M$-dimensional subspace in the minimum squared error sense.

In Adaboost learning, each weak classifier is constructed based on the histogram of a single feature derived from PCA coefficients $(\theta_1, \ldots, \theta_M)$. At each round of boosting, one PCA coefficient -the most effective to discriminate the face and non-face classes- is selected by Adaboost. Note that the boosting algorithm selects features derived from PCA based on their ability to discriminate face and non-face samples, rather than on the rank of their eigenvalues. Therefore, some PCA features corresponding to small eigenvalues may be selected in the earlier stages instead of those with larger eigenvalues.

As stated earlier, the distributions of the two classes in the Haar-like feature space almost completely overlap in the later stages of the cascade training. In that case, we propose to switch features spaces and construct weak classifiers in the PCA space. Empirically, we have found that in such space, the distributions of the face and non-face classes have smaller overlap, given the same set of non-faces obtained by bootstrapping and used for training of later cascade stages. This situation is illustrated in Fig. 4. We can observe that the two classes are better separated, and therefore we can expect that weak classifiers based on PCA coefficients are "boostable".

One question regarding cascade boosting in hierarchical feature spaces is at which stage of the cascade we should decide to switch from Haar-like to PCA features (we refer to such stage as the *switching stage*). It is well-known that PCA features are much more expensive in computational terms than Haar-like features. On one extreme. if we used PCA features in the very early stages of boosting, we would have to extract PCA features from a very large number of sub-windows, and the speed of the face detection system would be unacceptably slow. On the other extreme, if we used PCA features in the very late stages of boosting, the performance improvement gained from their usage would be limited. Therefore, we determine the switching stage based on the tradeoff between speed and performance improvement.

## 4. Results

For training purposes, a total of 11,341 face examples were collected from various sources, covering out-of-plane rotation in the range $[-20^\circ, +20^\circ]$. All faces were manually aligned by the eyes position. For each aligned face example, five synthesized face examples were generated by a random in-plane-rotation in the range $[-20^\circ, +20^\circ]$, random scaling in the range $[-10\%, +10\%]$, random mirroring and random shifting to +1/-1 pixel. This created a training set of 56,705 face examples. The face examples were then cropped and re-scaled to $20 \times 20$ pixels. For non-face examples, enough instances were collected from over 100,000 large images containing no faces.

In our experiments, two face detection systems were used. The first one was trained using only the Haar-like features. We refer to this system as *S-Boost* as it was only applied in a *Single* feature space. The second system was trained using both Haar-like features and PCA features. We refer to it as *H-Boost* due to the *Hierarchical* feature spaces we use. We compared the two classifiers on the complete CMU frontal face test set. The test set is composed of 130 images containing 510 faces, and has been also used to report results by the state-of-the-art systems [4, 10].

Fig.5 shows the ROC curves for both classifiers. Since changing the switching stage of H-Boost will affect both the system performance and speed, the mean and standard deviation were used to measure the performance of H-Boost, and obtained by running the system 10 times with different switching stages. We can see that H-Boost performs consistently better than S-Boost. On one hand, the detection rate of H-Boost is always higher than that of S-Boost, given the same number of false alarms. On the other hand, for a given detection rate, the false alarms of H-Boost are always fewer than those of S-Boost. The higher performance of H-Boost reflects the benefit of the usage of the PCA features in the late stages, which are more effective to discriminate face and non-face examples. Fig. 6 shows the curves

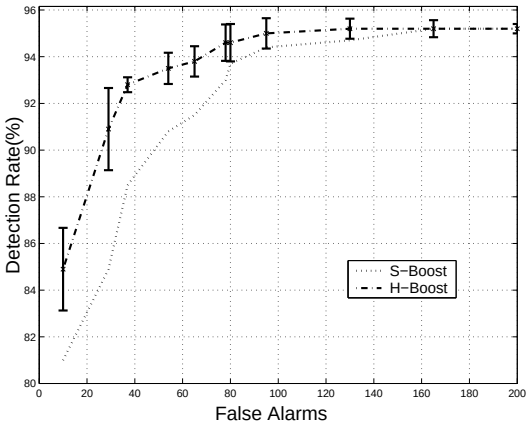


**Figure 5. ROC curves for S-Boost and H-Boost.**

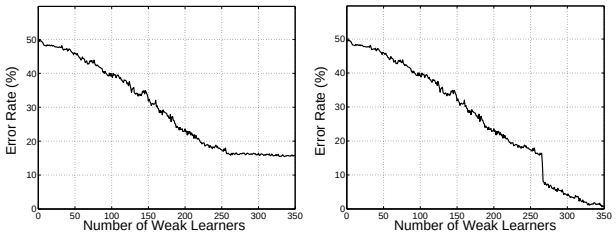


**Figure 6. Left: The error rate as a function the number of the selected weak classifiers using only Haar-like features. Right: The error rate using both Haar-like features and PCA features.**

of the error rate (average of false alarm rate and false rejection rate) as a function of the number of the selected weak classifiers in the switching stage from Haar-like features to PCA features. We can see that as the number of selected weak classifiers increased, the error rate always decreased. However, from the 265th weak classifier on, the error rate decreased only marginally for S-Boost, which indicates that any further selected weak classifiers could not discriminate face and non-face samples well. As a result, the selected weak classifiers contribute very little to the final strong classifier. On the contrary, switching from Haar-like space to PCA space decreased the error rate significantly. For H-Boost, boosting learning continued selecting weak learners in PCA space that discriminate face and non-face well, thus the error rate continues to decrease.

We test the speed of two face detection systems using a Pentium-P4 2.6GHz, 512MB RAM computer. Using a starting scale of 1.2 and a step size of 1.25, both systems can process 15 frames per second for $320\times240$ images. There are two facts that make the computational complexity of H-Boost comparable to that of S-Boost. First, a large majority of sub-windows are rejected by the first several layers in the cascade, so only a very small number of sub-window candidates will be verified by the later stages using PCA features. Second, the number of selected PCA features is far less than that of the Haar-like features selected by boosting at the same stage.

## 5. Conclusion

The paper introduced a novel boosting-based face detection algorithm in hierarchical feature spaces. Motivated by the fact that the weak learners based on the simple Haar-like features are too weak in the later stages of the cascade, we propose to boost PCA features in the later stages. The global PCA feature space complements the local Haar-like feature space. The algorithm selects the most effective features from PCA features using boosting, instead of ranking them according to their eigenvalues. The experiments on the CMU face test set showed that the proposed metholdogy can achieve better performance than a current state-of-the-art, single feature, Adaboost-based detection system, at a comparable speed.

## References

[1] Y. Freund and R. Schapire. “A decision-theoretic generalization of on-line learning and an application to boosting”. *Journal of Computer and System Sciences*, August 1997.

[2] S. Z. Li, L. Zhu, Z. Q. Zhang, A. Blake, H. Zhang, and H. Shum. “Statistical learning of multi-view face detection”. In *Proc. of the European Conference on Computer Vision*, Copenhagen, Denmark, May 28 - June 2 2002.

[3] C. P. Papageorgiou, M. Oren, and T. Poggio. “A general framework for object detection”. In *IEEE International Conference on Computer Vision*, Bombay, India, 1998.

[4] H. A. Rowley, S. Baluja, and T. Kanade. “Neural network-based face detection”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998.

[5] P. Y. Simard, Y. A. L. Cun, J. S. Denker, and B. Victorri. “Transformation invariance in pattern recognition - tangent distance and tangent propagation”. In *Neural Networks: Tricks of the Trade*. Springer, 1998.

[6] K.-K. Sung and T. Poggio. “Example-based learning for view-based human face detection”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998.

[7] M. A. Turk and A. P. Pentland. “Eigenfaces for recognition”. *Journal of Cognitive Neuroscience*, March 1991.

[8] L. Valiant. “A theory of the learnable”. *Communications of ACM*, 1984.

[9] P. Viola and M. Jones. “Asymmetric AdaBoost and a detector cascade”. In *Proc. of Neural Information Processing Systems*, Vancouver, Canada, December 2001.

[10] P. Viola and M. Jones. “Rapid object detection using a boosted cascade of simple features”. In *IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, December, 2001.