SODA-Boosting and Its Application to Gender Recognition

Xun Xu and Thomas S. Huang

Beckman Institute, University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA

Abstract. In this paper we propose a novel boosting based classification algorithm, SODA-Boosting (where SODA stands for Second Order Discriminant Analysis). Unlike the conventional AdaBoost based algorithms widely applied in computer vision, SODA-Boosting does not involve time consuming procedures to search a huge feature pool in every iteration during the training stage. Instead, in each iteration SODA-Boosting efficiently computes discriminative weak classifiers in closed-form, based on reasonable hypotheses on the distribution of the weighted training samples. As an application, SODA-Boosting is employed for image based gender recognition. Experimental results on publicly available FERET database are reported. The proposed algorithm achieved accuracy comparable to state-of-the-art approaches, and demonstrated superior performance to relevant boosting based algorithms.

1 Introduction

Automatic recognition of demographic properties, e.g. gender, age and ethnicity, from face images has many applications in intelligent surveillance, demographic statistics and human-computer interaction. Since early 1990s, gender recognition has attracted considerable attention of the computer vision and patter recognition community for a long time. Early works on this topic were mostly based on neural network [1,2,3,4], where promising performance (more than 90% in accuracy) were reported, although most experiments were conducted on rather small databases (consisting of dozens of images, except [4] where the FERET database was used). From the aspect of patter recognition, gender recognition is a typical two-class problem. In recent years, two most successful “off-the-shelf” classifiers, i.e. SVM [5,6] and AdaBoost [7,8,9], seem to have dominated in this area, because of their higher accuracy and robustness compared to earlier techniques. Both classifiers achieve comparably good recognition accuracy [9]. However, AdaBoost based gender recognizers are generally faster than SVM, which may be a desirable advantage for real-time applications.

In this paper, we present a novel classification algorithm for two-class problem, namely SODA-Boosting, and apply it to gender recognition. This algorithm is along the AdaBoost line. The main contribution lies in the methodology to discover the most discriminative features in an effective and efficient way. AdaBoost [10] is among the most influential recent advances in machine learning, and has

S.K. Zhou et al. (Eds.): AMFG 2007, LNCS 4778, pp. 193–204, 2007.

been widely adopted in computer vision problems. AdaBoost provides an elegant framework to aggregate weak classifiers into a strong one with theoretically provable generalization performance. In computer vision community, the most common practice of applying AdaBoost is defining a huge pool of candidate weak classifiers, and in each iteration seeking the best one through exhaustively traversing the whole feature pool. In order to evade the drawbacks of that approach such as computational load, the proposed algorithm takes a different approach, attempting to directly compute the discriminative features. The proposed algorithm is related to recent algorithms [11,12] which are based on similar motivation. However, SODA-Boosting makes more comprehensive hypotheses on the distribution of the two classes, resulting a stronger learning procedure. The effectiveness of the proposed algorithm is demonstrated by gender recognition experiments reported in this paper, where it achieved accuracy comparable to state-of-the-art gender recognizers, surpassing related boosting algorithms in performance.

Rest of this paper is organized as follows: In Section 2 the SODA-Boosting algorithm is introduced and discussed in detail. Section 3 presents experimental results of gender recognition on the FERET database [13], where SODA-Boosting is compared to many other algorithms. Finally we conclude the paper in Section 4 and briefly discuss future work.

2 SODA-Boosting

Boosting, especially AdaBoost (*Ada*ptive *Boost*ing) [10], is one of the most important and influential recent advances in machine learning, regarded by some researchers as the "best off-the-shelf classifier". During the recent years, it has been widely adopted in computer vision, resulting in many successful applications. As a meta-algorithm that constructs a strong classifier by "boosting" weak classifiers, the key of employing AdaBoost is to design appropriate "weak classifiers" (or "weak hypotheses"). In the computer vision community, the most common practice is defining a huge pool of candidate weak classifiers (depending on the domain knowledge for the problem of interest), and in each iteration seeking the best one (leading to the lowest classification error rate on the weighted training sample set) through traversing the whole feature pool. Such a practice actually treats AdaBoost as a feature selector. This approach has been popularized since successful deployment of Viola-Jones face detector [14]. However, it has several drawbacks:

– The pre-defined feature pool restricts the optimality of the features that can ever be discovered, as the features that will be incorporated into the final classifier are strictly limited to this pool. For example, if linear weak classifiers are the candidates (which is the case for most applications of AdaBoost in computer vision), the feature pool is indeed only a very sparse sampling of the image space (due of the high dimensionality of the space). Even when the feature pool is "over-complete" (implying that its size is

larger than the dimensionality of the data space), there are still too few features to cover the whole space.
- It is computationally expensive. Usually the feature pool is huge, in order to avoid over-sparse sampling of the feature space. Exhaustive traversing such a huge pool in every iteration is clearly rather time consuming.
- The design of candidate weak features largely relies on certain domain knowledge of the problem at hand. Therefore the resulted classifier is not generic, i.e. could not be applied to other problems of different nature. For instance Viola-Jones classifier [14], where rectangular filters are used as weak classifiers, cannot be used in an audio related classification problem where the samples to be classified are MFCC coefficients.

Unlike the conventional practice discussed so far, SODA-Boosting takes another path to discover the weak classifiers. Instead of pre-defining a feature pool and *searching* for "good" features, it attempts to directly *compute* the weak classifiers that are ideal for classification purpose, in a computationally efficient way. In the SODA-Boosting algorithm, we limit the weak classifiers to be "linear-based", i.e. each weak hypothesis is reached by linearly projecting the sample onto a certain vector and thresholding the projection. The key of SODA-Boosting is how to learn such linear projection vectors and their corresponding thresholds, which is detailed in the following subsection.

2.1 SODA: Second Order Discriminant Analysis

In the training procedure of AdaBoost, in each boosting iteration one needs to learn a weak classifier to classify the weighted training samples. In our case, the goal is to seek a linear projection, and construct an effective weak classifier on that projection. Clearly, the weak classifier should have sufficient discriminative power. In SODA-Boosting, we seek such discriminative linear projections via two different techniques, FLD and MRC. With both techniques, the optimal linear projections can be computed in closed-form without exhaustive search or ad-hoc numerical optimization. As we shall see, since both FLD and MRC seek discriminative linear projections by utilizing statistical moments of (up to) second order, we categorize them under a common name SODA (Second Order Discriminant Analysis).

Fisher Linear Discriminant (FLD). FLD is the most well-known technique to find a discriminative linear projection [15]. Suppose we need to classify two classes, $\mathbf{X}^+, \mathbf{X}^- \subset \mathcal{R}^n$. In the training stage we have a labeled sample set $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}$ where $y_i \in \{+1, -1\}$ indicating $\mathbf{x}_i \in \mathbf{X}^+, \mathbf{X}^-$ respectively. In the boosting framework, each training sample is associated with a weight, say $\{w_1, w_2, \ldots, w_N\}$. In each iteration, the weights are normalized so that $\sum_{i=1}^{N} w_i = 1$. The weighted means of the two classes are given by:

$$\mathbf{m}_+ = \frac{1}{\sum_{y_i=+1} w_i} \sum_{y_i=+1} w_i \mathbf{x}_i, \tag{1}$$

and

$$\mathbf{m}_{-} = \frac{1}{\sum_{y_i=-1} w_i} \sum_{y_i=-1} w_i \mathbf{x}_i \tag{2}$$

respectively. The weighted scatter matrices are:

$$\mathbf{S}_{+} = \sum_{y_i=+1} w_i \left(\mathbf{x}_i - \mathbf{m}_{+}\right) \left(\mathbf{x}_i - \mathbf{m}_{+}\right)^T, \tag{3}$$

and

$$\mathbf{S}_{-} = \sum_{y_i=-1} w_i \left(\mathbf{x}_i - \mathbf{m}_{-}\right) \left(\mathbf{x}_i - \mathbf{m}_{-}\right)^T. \tag{4}$$

Defining within class scatter matrix

$$\mathbf{S}_W = \mathbf{S}_{+} + \mathbf{S}_{-}, \tag{5}$$

and between class scatter

$$\mathbf{S}_B = \left(\mathbf{m}_{+} - \mathbf{m}_{-}\right) \left(\mathbf{m}_{+} - \mathbf{m}_{-}\right)^T. \tag{6}$$

FLD is the vector $\mathbf{w}_{FLD}$ that minimizes criterion

$$J_{FLD}\left(\mathbf{w}\right) = \frac{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}. \tag{7}$$

And it turns out that

$$\mathbf{w}_{FLD} = \mathbf{S}_W^{-1} \left(\mathbf{m}_{+} - \mathbf{m}_{-}\right). \tag{8}$$

The weak classifier associated with the FLD feature is given as

$$f_{FLD}\left(\mathbf{x}; \mathbf{w}_{FLD}, T\right) = \begin{cases} +1, & \mathbf{w}_{FLD}^T \mathbf{x} > T \\ -1, & else \end{cases}, \tag{9}$$

where optimal threshold T is chosen to minimize classification error

$$\varepsilon = \sum_{f_{FLD}(\mathbf{x}_i) \neq y_i} w_i \tag{10}$$

Maximal Rejection Classifier (MRC). FLD works well when the two classes are linearly separable (or at least approximately so) and when their covariance matrices are close to each other. In many practical problems, these conditions are not met. One especially interesting case is the "target detection" configuration [16], where one class (the *target*) is surrounded by the other (the *clutter*), as illustrated in Figure 1. This configuration is common in many computer vision problems, among which object detection is a natural example. As the two classes are by no means linearly separable, the best a linear projection can do to discriminate the two classes is to minimize the overlap between the projected samples from the two classes. An intuitive way to achieve this is seeking a projection vector on which the target class is "squeezed" whereas the clutter class is "pushed aside" as much as possible. Based on this idea, Elad et al [16] proposed a technique called Maximal Rejection Classifier (MRC) to find the discriminative projections, and applied it to face detection. In [12] Xu and Huang showed that face recognition can also be modeled as a two-class problem of this type, and proposed the MRC-Boosting algorithm where the MRC classifiers are aggregated via boosting.

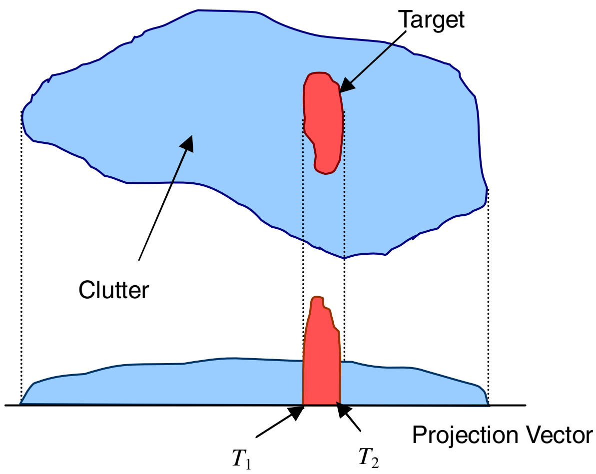


Fig. 1. Target detection configuration and MRC projection

If we treat class $\mathbf{X}^+$ as the target and $\mathbf{X}^-$ as the clutter, the MRC feature is the projection vector $\mathbf{w}_{MRC+}$ which minimizes criterion functional

$$J_{MRC+}(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_+ \mathbf{w}}{\mathbf{w}^T (\mathbf{S}_W + \mathbf{S}_B) \mathbf{w}}. \tag{11}$$

This criterion seems much similar to that of FLD (7), as it is also in the form of a generalized Rayleigh quotient. However, they chase completely different goals, leading to rather different projection vectors. The MRC feature $\mathbf{w}_{MRC+}$ is found through solving a generalized eigenvalue problem and picking the generalized eigenvector associated with the smallest eigenvalue. The weak classifier associated with MRC feature $\mathbf{w}_{MRC+}$ is

$$f_{MRC+}\left(\mathbf{x}; \mathbf{w}_{MRC+}, T_1^+, T_2^+\right) = \begin{cases} +1\,, T_1^+ \le \mathbf{w}_{MRC+}^T \mathbf{x} \le T_2^+ \\ -1\,, else \end{cases}, \tag{12}$$

where the thresholds T_1^+ and T_2^+ are chosen to minimize classification error, similar to (10). Note that unlike the FLD classifier, this weak classifier contains two thresholds, therefore it is not a linear classifier, but "linear-based" [16].

Similarly, if we instead treat class $\mathbf{X}^-$ as the target and $\mathbf{X}^+$ as the clutter, we obtain the other MRC feature $\mathbf{w}_{MRC-}$ which minimizes

$$J_{MRC-}(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_- \mathbf{w}}{\mathbf{w}^T (\mathbf{S}_W + \mathbf{S}_B) \mathbf{w}}, \tag{13}$$

and the associated weak classifier

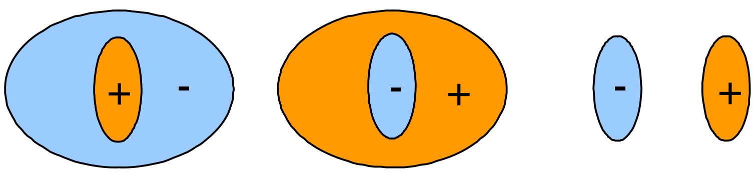
$$f_{MRC-}\left(\mathbf{x}; \mathbf{w}_{MRC-}, T_1^-, T_2^-\right) = \begin{cases} -1\,, T_1^- \le \mathbf{w}_{MRC-}^T \mathbf{x} \le T_2^- \\ +1\,, else \end{cases} \tag{14}$$

2.2 The SODA-Boosting Algorithm

The SODA-Boosting algorithm employs AdaBoost framework to aggregate the SODA classifiers (i.e., FLD and two kinds of MRC classifiers), leading to a strong classifier. In each boosting iteration, because we don't know which SODA feature is appropriate for current distribution of the two classes (represented by the weighted samples), three hypotheses are made, as illustrated in Figure 2. The first two hypotheses, MRC+ and MRC−, reflect the cases where $\mathbf{X}^+$ is surrounded by $\mathbf{X}^-$ and vice versa, respectively. The last hypothesis reflects the configuration where the two classes are well separated, and can be reasonably discriminated by FLD. For these hypotheses, we employ techniques discussed in the previous subsection to obtain corresponding weak classifiers respectively. Naturally, the one resulting in lowest classification error rate best models current distribution of the two classes, hence will be selected and included into the final strong classifier. The algorithm is listed in Algorithm 1.

2.3 Discussion

SODA-Boosting is closely related to two other boosting based classification algorithms, MRC-Boosting [12] and FisherBoost [11]. MRC-Boosting employs AdaBoost framework to aggregate MRC classifiers. As shown in [12], it works well for face recognition. However, just as the original MRC approach [16], MRC-Boosting was designed specifically for "target detection" type problems, as illustrated in Figure 1. However, for a general two-class discrimination problem, we don't really know whether the distribution of the two classes obey such configuration. Although many problems in computer vision, such as object detection and face recognition, can be modeled as "target detection", this assumption is



(a) MRC+ (b) MRC- (c) FLD

Fig. 2. Hypotheses made by SODA-Boosting on the distribution of the two classes

Algorithm 1. SODA-Boosting algorithm

Input: $\{(\mathbf{x}_i, y_i), i = 1, 2, \cdots, N : \mathbf{x}_i \in \mathcal{R}^n, y_i \in \{+1, -1\}\}$. The maximal number of weak classifiers K.

Initialize: $w_i = \begin{cases} 1/2N^+, & y_i = +1 \\ 1/2N^-, & y_i = -1 \end{cases}$, where N^+ and N^- are the numbers of positive and negative samples respectively.

for $k = 1, 2, \cdots, K$ **do**

Compute weighted means $\mathbf{m}_+$, $\mathbf{m}_-$ and scatter matrices $\mathbf{S}_+$, $\mathbf{S}_-$, using (1)∼(4) respectively.

Compute FLD feature using (8), obtain the associated weak classifier $f_{FLD}(\mathbf{x})$ according to (9), and calculate its classification error ε_{FLD}.

Compute MRC+ feature by minimizing (11), obtain the associated weak classifier $f_{MRC+}(\mathbf{x})$ according to (12), and calculate its classification error ε_{MRC+}.

Compute MRC− feature by minimizing (13), obtain the associated weak classifier $f_{MRC-}(\mathbf{x})$ according to (14), and calculate its classification error ε_{MRC-}.

Select weak classifier $f_k(\mathbf{x}) \in \{f_{FLD}, f_{MRC+}, f_{MRC-}\}$ with minimal classification error ε_k.

Update weights: $w_i \leftarrow \frac{1}{Z_k} w_i \exp[-\alpha_k y_i f_k(\mathbf{x}_i)]$, where $\alpha_k = \frac{1}{2} \ln \frac{1-\varepsilon_k}{\varepsilon_k}$ and Z_k is a normalization factor to ensure $\sum_{i=1}^{N} w_i = 1$.

end

Output: Strong classifier $F(\mathbf{x}) = sgn[G(\mathbf{x})]$ where the classification function is $G(\mathbf{x}) = \sum_{k=1}^{K} \alpha_k f_k(\mathbf{x})$.

not likely to be true for a general two-class problem. When we don't have a compelling reason that the two classes form a "target detection" configuration, the features (and the associated weak classifiers) sought by MRC wouldn't be effective for classification. On the other hand, FisherBoost does not consider the "target detection" configuration at all, in each iteration it seeks a FLD classifier to discriminate the two classes. As we know, FLD is only effective when the two classes are linearly separable (or at least approximately so). For a complicated two-class problem, especially after the samples are re-weighted as the boosting procedure goes on, the distribution of the two classes may not be such case. At that point, FLD will fail to discover meaningful discriminants.

SODA-Boosting overcomes the limitation of MRC-Boosting and FisherBoost by including both MRC and FLD classifiers into consideration. FLD and MRC are complement to each other, working for rather distinct configurations of the two classes. Considering all these configurations has much stronger discriminative power than just considering one of them. As we shall see in Section 3, putting together FLD and MRC classifiers is actually *not* a trivial combination, it indeed leads to a much stronger learning procedure.

Besides MRC-Boosting and FisherBoost, another relevant approach which also attempts to directly *compute* weak classifiers is KL-Boosting [17]. KL-Boosting discovers discriminative features by seeking linear projection vectors on which the Kullback-Leibler divergence between the two classes are maximized. Unfortunately, such definition of discriminative power can not lead to a closed-form solution, hence an ad-hoc numerical optimization procedure was employed, which is computationally expensive and relatively difficult for implementation. Unlike KL-Boosting, SODA-Boosting (also MRC-Boosting and FisherBoost) is able to find the discriminative features in closed-form, using standard techniques in linear algebra, hence is more efficient.

3 Experiments

We applied the proposed SODA-Boosting algorithm to gender recognition, and report in this section experimental results on the FERET database [13].

In our experiments, 4109 frontal face images from the FERET database were aligned according to eye-corner coordinates supplied with the data, and finally normalized to 40×40. In order to normalize the illumination effect, all images were pre-processed to be of zero-mean and unit variance in pixel value. This data set consists of 703 male individuals and 498 female, the gender ground truth were manually labeled by viewing full face images (i.e. before the faces were aligned and cropped out).

In each run of experiments, we randomly partition the data into training and test set. Each time 80% of the individuals in the database were selected for training, all their images constitute the training set, and the remaining images were used for test. Note that we took a protocol similar to that in [9], where training and test data are separated based on individuals, instead of images themselves, hence any individual in the training set wouldn't have any images in the test set. This is in contrast to some earlier work e.g. [5] where images of one individual might appear in both training and test set (called "mixed data sets" in [9]), resulting in a more optimistic estimate of recognition accuracy. The protocol employed in our experiments and [9], on the other hand, is more close to the practical scenario where the trained gender recognizer will be tested on images of people it never saw before, leading to more accurate evaluation of the generalized performance.

We conducted gender recognition experiments with 10 independent random data partitions, and recorded the average accuracy of different approaches, as shown in Table 1. For comparison purpose, besides the proposed SODA-Boosting algorithm we also reported performance of other classifiers, including SVM, AdaBoost with rectangular filters (used in Viola-Jones face detector [14]) [7], and two algorithms related to the proposed one: FisherBoost [11] and MRC-Boosting [12].

For SVM (SVM-Light implementation [18]), RBF kernel was used (we also tried polynomial kernels, but the performance was inferior to RBF kernel). As reported by [9], the accuracy is sensitive to parameters C and γ. In our experiments, we took $C = 1/N$ (where $N = 1600$ is the dimensionality of the images)

Table 1. Gender recognition accuracy

Algorithm	**Accuracy**
SODA-Boosting	92.82%
SVM ($\gamma = 1$)	92.38%
SVM ($\gamma = 10^5$)	93.34%
AdaBoost (Viola-Jones)	92.67%
MRC-Boosting (+)	88.69%
MRC-Boosting (−)	89.76%
FisherBoost	89.54%

which is a good choice according to results reported in [9] and two different γ values, $\gamma = 1$ which is the default value of SVM-Light and $\gamma = 10^5$ which was shown by [9] to be optimal (through exhaustive parameter tuning). The number of support vectors varies across runs, on average 860 for $\gamma = 10^5$ and 1450 for $\gamma = 1$ respectively. For all the boosting based algorithms, $K = 500$ weak classifiers were used. The performance of SODA-Boosting (with 500 features) is better than SVM with default γ, and is slightly inferior, although still comparable, to SVM with optimal parameter setting (RBF kernel with $\gamma = 10^5$ and $C = 1/1600$). However, note that SVM employs more features (i.e. support vectors) than SODA-Boosting.

When compared to other boosting based algorithms, SODA-Boosting consistently worked better. For MRC-Boosting, we conducted experiments with two versions (marked with $+/-$) considering male and female as "target" respectively, and both of them achieved inferior performance to SODA-Boosting. In Figure 3 we compare the accuracy of different boosting algorithms as the number of weak classifiers increases. It can be seen that SODA-Boosting cleanly exceeded FisherBoost and two versions of MRC-Boosting everywhere. The asymptotic accuracy of SODA-Boosting was very similar to conventional AdaBoost based algorithm [7]. However, thanks to the effort of directly seeking most discriminative features, SODA-Boosting reached the same accuracy with fewer features than [7]. Although we did not compare SODA-Boosting to recent approach [9] directly, we conjecture that the comparison would reach similar conclusion, as [9] shares the same nature (i.e. boosting very weak features) as [7] and they achieved similar performance. It should be noted that although SODA-Boosting requires fewer features for the same accuracy, it is usually slower than boosting approaches like [7,9] *in classification stage*, because those approaches require much less computation per feature. However, those approaches are specifically designed to classify images, relying on weak features that are all pre-designed based on certain domain knowledge, hence bounded to certain classification problems. On the contrary, SODA-Boosting is a generic algorithm, which is, in theory, applicable to any two-class problem, just like SVM.

The point especially worth noticing is that SODA-Boosting cleanly exceeded both MRC-Boosting and FisherBoost in performance, although the latter two also aggregate FLD and MRC classifiers respectively. This convincingly demonstrates that putting together FLD and MRC weak classifiers is *not* a trivial combination, it indeed leads to a stronger learning procedure, as we mentioned in Section 2.

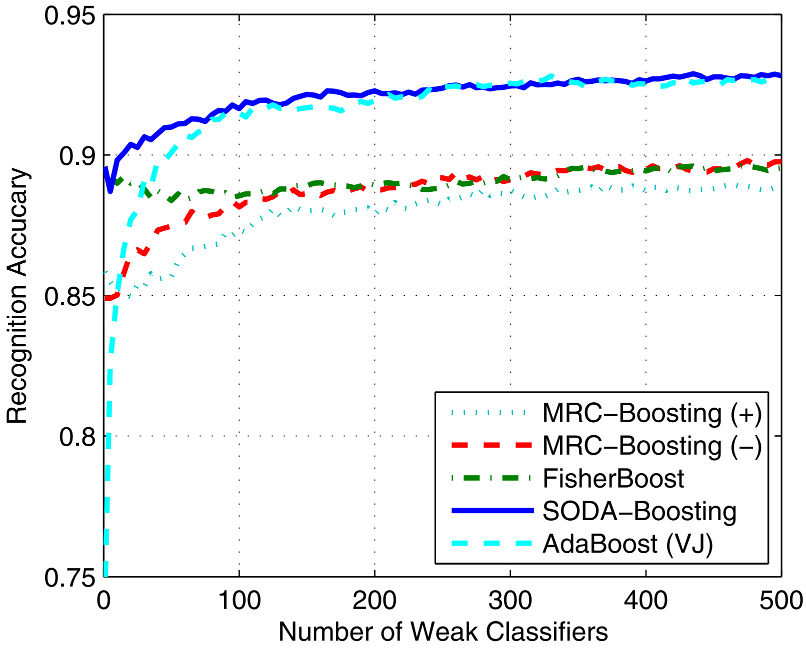


Fig. 3. Accuracy of different boosting algorithms as the number of classifiers increases

4 Conclusion and Future Work

In this paper, we proposed a novel boosting based classification algorithm called SODA-Boosting. Unlike conventional AdaBoost based algorithms widely used in computer vision e.g. [14], SODA-Boosting does not involve any exhaustive search across a huge feature pool, instead it attempts to directly compute discriminative features in a computationally efficient way. Compared to recent approaches [11,12] with similar motivation, the proposed algorithm takes into consideration several distinct hypotheses on the distribution of the two classes, resulting a stronger learning procedure. Application to gender recognition shows that SODA-Boosting achieves considerably better performance than those approaches, reaching accuracy comparable to that of state-of-the-art gender recognizers. In this work, we conducted experiments on the FERET database which only consists of frontal face images obtained under controlled condition (e.g. regular illumination). However, in practical applications, image quality seldom reaches such an ideal level. The images captured from real-world videos usually, if not always, are of rather poor quality, which means low resolution, arbitrary illumination, imaging noises, and most importantly, the faces are often captured in non-frontal views. As future work, we are interested in applying the proposed algorithm to these challenging settings.

References

1. Golomb, B.A., Lawrence, D.T., Sejnowski, T.J.: Sexnet: A neural network identifies sex from human faces. In: NIPS-3: Proceedings of the 1990 conference on Advances in neural information processing systems 3, pp. 572–577. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1990)
2. Cottrell, G.W., Metcalfe, J.: Empath: face, emotion, and gender recognition using holons. In: NIPS-3: Proceedings of the 1990 conference on Advances in neural information processing systems 3, pp. 564–571. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1990)
3. Tamura, S., Kawai, H., Mitsumoto, H.: Male/female identification from 86 very low resolution face images by neural network. Pattern Recognition 29(2), 331–335 (1996)
4. Gutta, S., Huang, J.R.J., Phillips, P.J., Wechsler, H.: Mixture of experts for classification of gender, ethnic origin, and pose of human faces. IEEE Trans. Neural Networks 11, 948–960 (2000)
5. Moghaddam, B., Yang, M.: Learning gender with support faces. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5), 707–711 (2002)
6. Graf, A.B.A., Wichmann, F.A.: Gender classification of human faces. In: Bülthoff, H.H., Lee, S.-W., Poggio, T.A., Wallraven, C. (eds.) BMCV 2002. LNCS, vol. 2525, pp. 491–500. Springer, Heidelberg (2002)
7. Shakhnarovich, G., Viola, P.A., Moghaddam, B.: A unified learning framework for real time face detection and classification. In: FGR 2002. Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, p. 16. IEEE Computer Society, Washington, DC, USA (2002)
8. Wu, B., Ai, H., Huang, C.: Lut-based adaboost for gender classification. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 104–110. Springer, Heidelberg (2003)
9. Baluja, S., Rowley, H.: Boosting sex identification performance. In: AAAI-IAAI 2005 (2005)
10. Schapire, R.E., Freund, Y., Bartlett, P., Lee, W.S.: Boosting the margin: a new explanation for the effectiveness of voting methods. In: Proc. 14th International Conference on Machine Learning, pp. 322–330. Morgan Kaufmann, San Francisco (1997)
11. Tu, J., Zhang, Z., Zeng, Z., Huang, T.: Face localization via hierarchical condensation with fisher boosting feature selection. In: CVPR 2004. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 02, pp. 719–724 (2004)
12. Xu, X., Huang, T.S.: Face recognition with MRC-Boosting. In: ICCV 2005. 10th IEEE International Conference on Computer Vision, vol. 2, pp. 1770–1777 (2005)
13. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. 22(10), 1090–1104 (2000)
14. Viola, P., Jones, M.J.: Robust real-time object detection. In: IEEE Workshop on Statistical and Theories of Computer Vision, IEEE Computer Society Press, Los Alamitos (2001)

15. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley-Interscience, Chichester (2000)
16. Elad, M., Hel-Or, Y., Keshet, R.: Pattern detection using a maximal rejection classifier. Pattern Recognition Letters 23(12), 1459–1471 (2002)
17. Liu, C., Shum, H.Y.: Kullback-leibler boosting. In: CVPR 2003. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 01, p. 587. IEEE Computer Society Press, Los Alamitos (2003)
18. Joachims, T.: Making large-Scale SVM Learning Practical. In: Advances in Kernel Methods - Support Vector Learning, MIT Press, Cambridge (1999)