DOMAIN-PARTITIONING RANKBOOST FOR FACE RECOGNITION

Bangpeng YAO[1], Haizhou AI[1], Yoshihisa IJIRI[2], Shihong LAO[2]

[1]Computer Science and Technology Department, Tsinghua University, Beijing 100084, China
[2]Sensing and Control Technology Laboratory, Omron Corporation, Kyoto 619-0283, Japan
Email: ahz@mail.tsinghua.edu.cn

ABSTRACT

In this paper we propose a domain partitioning RankBoost approach for face recognition. This method uses Local Gabor Binary Pattern Histogram (LGBPH) features for face representation, and adopts RankBoost to select the most discriminative features. Unlike the original RankBoost algorithm in [1], weak hypotheses in our method make their predictions based on a partitioning of the similarity domain. Since the domain partitioning approach handles the loss function of a ranking problem directly, it can achieve a higher convergence speed than the original approach. Furthermore, in order to improve the algorithm's generalization ability, we introduce some constraints to the weak classifiers being searched. Experiment results on FERET database show the effectiveness of our approach.

Index Terms — Face recognition, RankBoost, Pattern classification

1. INTRODUCTION

Numerous algorithms for face recognition have been proposed for the past two decades [2]. Most of the existing effective algorithms solve face recognition in a two-class framework [3][4], in which differences or similarities of two samples from the same person and different persons compose intra-personal and extra-personal class respectively. Although some impressive results have been obtained in this framework, it does not strictly conform to the objective of the original face recognition task, which is broadly viewed as a multiclass problem where each class contains images of one individual. In this paper, we propose a ranking-based framework for face recognition that directly focuses on a multiclass problem by converting it to a ranking one.

RankBoost [1] is a boosting-based method for ranking problems which takes pairs of instances with relative rankings as its input. In [1], RankBoost learning minimizes its loss criterion with the assumption that all the weak rankers are designed in advance.

In this paper, we propose to implement RankBoost learning with domain-partitioning weak hypotheses (DP-RankBoost). The domain-partitioning technique was introduced in [5] to enhance the performance of AdaBoost algorithm, known as Real AdaBoost, which has been effectively applied to face detection problem [6]. Moreover, in order to improve the generalization ability of this algorithm, the optimization of DP-RankBoost is imposed with some constraints in this paper. The main contributions of the proposed approach are: 1) For each weak learner, with a partitioning of the feature domain, the sample space is mapped to another real-valued space to get a more precise prediction result. 2) The proposed method combines the design of weak rankers and the selection of learner's parameters to minimize the ranking loss, and therefore it is more consistent with the objective of a ranking problem. 3) Generalization ability of this algorithm is improved by introducing some constraints in the learning procedure.

The rest of this paper is organized as follows. The ranking-based framework is briefly presented in Section 2. In Section 3 we elaborate the domain-partitioning RankBoost algorithm. Experimental results are illustrated in Section 4. Finally in Section 5 is the conclusion.

2. A RANKING-BASED FRAMEWORK

The set of training images is given as $\{x_i, y_i\}_{i=1}^{N}$, where x_i is a training sample and $y_i \in \{1, \cdots, C\}$ is a class label. Given a set of features $\Phi = \{\varphi_k\}_{k=1}^{M}$, our goal is to select a subset of features and combine them to separate each class from the others.

The ranking-based framework for face recognition is shown in Figure 1. The training set is divided into C sub-problems, of which each focuses on correctly recognizing one specific individual through making the intra-personal similarities larger than extra-personal ones. Here each sub-problem can be regarded as a bipartite ranking problem [1], which targets to learn a ranking function that ranks the positive instances higher than negative ones. For each sub-problem, the intra-personal and extra-personal pairs are treated as positive and negative instances, respectively. In other words, the positive and negative samples for the cth ranking problem are

$$\begin{aligned} S_{c,+} &= \left\{ s_{i,j} \mid \left(y_i = c \right) \wedge \left(y_j = c \right) \wedge \left(i \neq j \right) \right\} \\ S_{c,-} &= \left\{ s_{i,j} \mid \left(y_i = c \right) \wedge \left(y_j \neq c \right) \right\} \end{aligned} \tag{1}$$

where $s_{i,j}$ is the similarity of x_i and x_j.

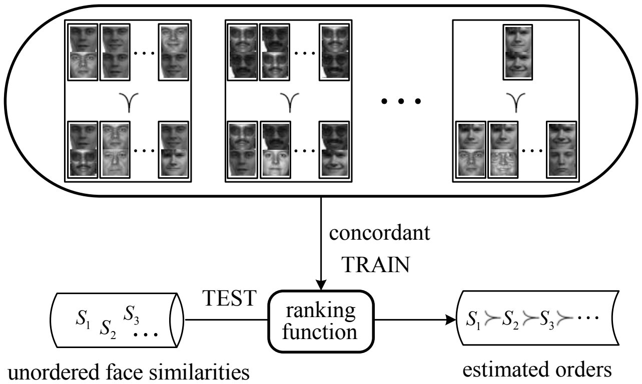


Figure 1. Ranking-based framework for face recognition

The inputs of the cth ranking problem are pairs of instances (s_0,s_1) from $S_{c,+}\times S_{c,-}$, provided with the information that s_0 should be ranked higher in each pair. In the ranking-based framework for face recognition, the input is a union of all the C bipartite ranking problems, *i.e.*,

$$S_+ \times S_- = \bigcup_{c=1}^{C}\left(S_{c,+} \times S_{c,-}\right). \tag{2}$$

3. DOMAIN-PARTITIONING RANKBOOST ALGORITHM

3.1. RankBoost for face recognition

The ranking problem can be solved by RankBoost [1], which is a boosting algorithm that operates in rounds. Given a pool of weak rankers, in round t RankBoost maintains a distribution D_t over $S_+\times S_-$ and selects the best ranker h_t with an importance α_t. After each round, D_t will be changed to assign higher importance to those instance pairs that are not correctly ordered by h_t. The final strong ranker H is a weighted combination of all the selected weak rankers.

It has been proved in [1] that in order to guarantee that the final combined ranking has a low ranking loss, in each iteration α_t and h_t should be chosen to minimize Z_t, which is the loss function in round t.

$$Z_t = \sum_{(s_0,s_1)} D_t\left(s_0,s_1\right)\exp\left(-\alpha_t\left(h_t\left(s_0\right)-h_t\left(s_1\right)\right)\right) \tag{3}$$

The RankBoost approach in [1] assumes that all the weak learners are designed in advance and Z_t is only used to guide the choice of α_t. Additionally, instead of greedily optimizing Z_t, this approach [1] expands the loss function in the training procedure as follows

$$Z_t \le \left(\frac{1-r_t}{2}\right)\mathrm{e}^{\alpha_t} + \left(\frac{1+r_t}{2}\right)\mathrm{e}^{-\alpha_t} \tag{4}$$

where $r_t = \sum_{(s_0,s_1)} D_t\left(s_0,s_1\right)\left(h_t\left(s_0\right)-h_t\left(s_1\right)\right)$.

3.2. Domain-partitioning RankBoost algorithm

Domain - Partitioning RankBoost Algorithm

Initialize: $\forall\left(s_0,s_1\right)\in S_+\times S_- : D_1\left(s_0,s_1\right)=1/\left|S_+\times S_-\right|$

For $t=1,\cdots,T$

1. For each feature ϕ_m do:
   a. Partition feature domain into: $X_1,\cdots,X_L$.
   b. Calculate: $W_{ij} = \sum_{(s_0,s_1)} D\left(s_0,s_1\right)$ for $s_0\in X_i, s_1\in X_j$.
   c. Find $c_1^m,\cdots,c_L^m$ to minimize: $Z^m = \sum_{i=1}^{L}\sum_{j=1}^{L} W_{ij}\exp\left(c_j^m - c_i^m\right)$.
2. Select $m^* = \arg\min_m Z^m$, let $Z_t = Z^{m^*}$, $h_t(s) = c_i^{m^*}$ when $s\in X_i$.
3. Update : $D_{t+1}\left(s_0,s_1\right) = \dfrac{D_t\left(s_0,s_1\right)\exp\left(h_t\left(s_1\right)-h_t\left(s_0\right)\right)}{Z_t}$.

Output the final ranking: $H(s) = \sum_{t=1}^{T} h_t(s)$.

Figure 2. Domain-Partitioning RankBoost algorithm

In this paper we propose a domain-partitioning RankBoost (DP-RankBoost) algorithm to greedily minimize Z_t. In each round, this approach combines the choice of α_t with the design and selection of h_t based on a partition of the feature domain. The algorithm is shown in Figure 2.

By folding α_t into h_t and omitting the subscript t for simplicity, the goal in each round is to minimize

$$Z = \sum_{(s_0,s_1)} D\left(s_0,s_1\right)\exp\left(-\left(h\left(s_0\right)-h\left(s_1\right)\right)\right). \tag{5}$$

In this approach, each weak ranker is associated with a partition of the feature domain S into disjoint blocks $X_1, \cdots, X_L$. The prediction of a weak ranker only depends on which block a given instance falls into. Let $c_i=h(s)$ for $s\in X_i$, then Equation 5 can be written as

$$Z = \sum_{i=1}^{L}\sum_{j=1}^{L} W_{ij}\exp\left(c_j - c_i\right) \tag{6}$$

where W_{ij} is defined as in Figure 2. Since Equation (6) cannot be minimized by an analytical solution of $\{c_1,\cdots,c_L\}$, we optimize this equation with Steepest Descent Algorithm.

From Equation (6) it can be seen that, even though the feature domain is only divided into L disjoint blocks, the weighted distributions of instances are estimated in $L\times L$ regions. This guarantees that, a specific region of a weak ranker gives the same output for any instance, no matter this instance is positive or negative.

3.3. LGBPH feature based LUT weak learner

Local Gabor Binary Pattern Histogram (LGBPH) is a facial representation method [7], in which a face image is first convolved by multi-resolution and multi-orientation Gabor filters and then encoded with LBP operator. The LGBPH features are obtained by estimating histograms of local sub-windows in LGBP images. Effectiveness of LGBPH relies on several aspects, including the Gabor decomposition, the LBP operator, and the local spatial histogram modeling.

In this work, we calculate LGBPH similarity with histogram intersection, of which the value is ranged from 0 to the window size. The Look-up-Table (LUT) classifier for each feature is defined with a partition of the similarity domain. Assuming that all the LGBPH features correspond to the same window size, then all the LGBPH similarity domains share the same division standard. The output of each block is determined by the solution of Equation (6).

The similarity domain is divided into 16 blocks in our experiment. Instead of equally dividing the similarity space, we divide it in the way that makes the number of samples in each block approximately equal.

3.4. Constrained Domain-Partitioning RankBoost

Since the weak classifiers are defined based on the similarity of two images, we expect that the blocks which correspond to large similarities have large outputs. In other words, assuming that a larger subscript of X_i represents the larger similarity value, we try to let $c_i \geq c_j$ when $i > j$.

In order to achieve this goal, we introduce some constraints in the minimization of Equation (6) as follows

$$\left.\begin{aligned} &\min\left(Z = \sum_{i=1}^{L}\sum_{j=1}^{L} W_{ij} \exp\left(c_j - c_i\right)\right) \\ &c_{i+1} - c_i \geq 0 \ \ \left(i = 1, 2, \cdots, L-1\right) \end{aligned}\right\} \quad (7)$$

The DP-RankBoost algorithm with constraints is called Constrained DP-RankBoost (CDP-RankBoost) algorithm in this paper. We use Multiplier Method [8] to get an optimized solution of Equation (7).

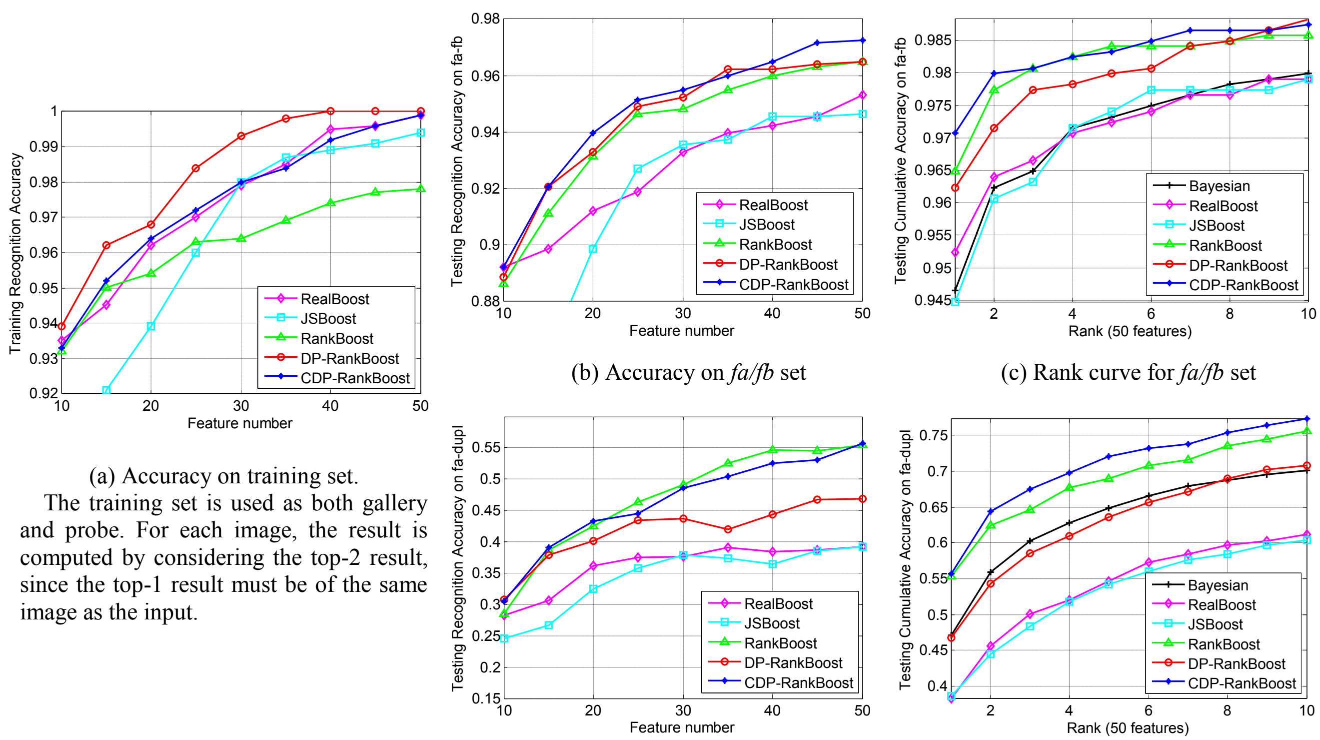
4. EXPERIMENTS

4.1. Face database and experimental setup

The proposed method is tested on FERET database [9]. The *fa* set is used as gallery, while *fb* and *dupI* are used as probe sets. The training set is the frontal images of FERET training set, which include 1002 images of 429 persons. All face images are cropped to the size of 64-by-64 pixels according to the positions of two eyes' center. Histogram normalization is applied to the cropped images.

We use the Gabor filters with three scales ($v = 0, 1, 2$) and four orientations ($\mu = 0, 2, 4, 6$), and both magnitude and phase features are used. The Gabor images are further encoded by $LBP_{8,2}$ to obtain LGBP images. Therefore we can obtain 24 LGBP images. LGBP histograms are estimated in the rectangular sub-windows with size 8-by-8 pixels. For each LGBP image we obtain 64 features by shifting the sub-window. From all the 24×64=1536 features we select 50 ones to construct the final ranker or classifier.

The recognition results of DP-RankBoost and CDP-RankBoost are compared with several other algorithms: Bayesian [3], Real AdaBoost [5], Jensen-Shannon Boosting [4], and the original RankBoost approach [1].



(a) Accuracy on training set.
The training set is used as both gallery and probe. For each image, the result is computed by considering the top-2 result, since the top-1 result must be of the same image as the input.

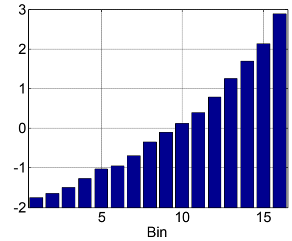
(b) Accuracy on *fa/fb* set

(c) Rank curve for *fa/fb* set

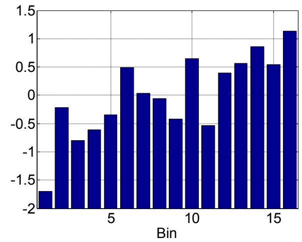
(d) Accuracy on *fa/dupI* set

(e) Rank curve for *fa/dupI* set

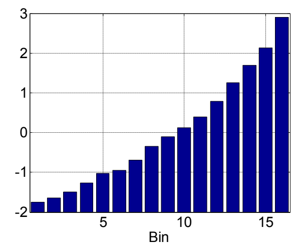
Figure 3. Experimental results. It is different from the previously published results because of the difference in experimental set.



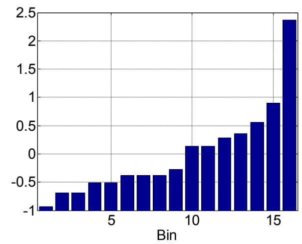
(a) DPRankBoost, 1st LUT



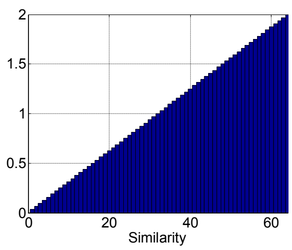
(b) DPRankBoost, 15th LUT



(c) CDPRankBoost, 1st LUT



(d) CDPRankBoost, 15th LUT



(e) RankBoost in [1]

Figure 4. Illustrations of weak classifiers' outputs. Y-axis is the output value. For the original RankBoost, all the weak learners have the same similarity-output relationship.

4.2. Results and analysis

The recognition results are shown in Figure 3. Figure 3(a) shows that DP-RankBoost outperforms all other algorithms on training set. It can obtain 100% accuracy with only 40 weak learners. The fact that DP-RankBoost converges faster than Real AdaBoost and JSBoost demonstrates the superiority of our ranking-based framework compared with the two-class one. Since the original RankBoost algorithm does not greedily optimize the training loss, it does not yield a faster convergence speed, although it is based on the same ranking-based framework. It can also be seen that even though some constraints are imposed on CDP-RankBoost, it converges fast on training set.

On testing sets, DP-RankBoost does not yield comparable results as CDP-RankBoost and original RankBoost. Figure 4 shows that with the learning iteration goes on, the weak learners selected by DP-RankBoost may give the learners with smaller similarities larger outputs. According to the theory of boosting, these regions do help in greedily optimizing the training loss. However, they do not perform well on unseen data. Since the training set of FERET does not contain all the types of variations exist in testing set, DP-RankBoost performs worse than CDP- RankBoost and original RankBoost, especially on *dupI* which contains large variations that are not included in training images.

It can be seen from Figure 3 that CDP-RankBoost outperforms all the other algorithms on testing sets, with one exception that the original RankBoost performs better in some situations of Figure 3(d). The constraints make those bins with large similarity values have large output values so that the generalization ability is improved. Figure 4(d) shows that some consecutive bins have the same output, which implies that without the constraints, the boosting approach might give large outputs to those bins which correspond to small similarity values, as in Figure 4(b).

5. CONCLUSIONS

In this paper, we propose to implement RankBoost learning with a partition of the feature domain. Two kinds of approaches, DP-RankBoost and CDP-RankBoost are proposed to perform face recognition. The experimental results show that, comparing with some other boosting-based algorithms, DP-RankBoost converges the fastest on training set, while CDP-RankBoost yields the best results on testing set.

The CDP-RankBoost is in fact an approach which searches the solution of RankBoost in a local zone which is defined with certain heuristic knowledge. Since it combines the merits of boosting learning and the constraint of heuristic knowledge, the CDP-RankBoost approach improves recognition results in our experiment.

6. ACKNOWLEDGEMENTS

This work is supported in part by National Science Foundation of China under grant No.60332010, No.60673107, and it is also supported by a grant from Omron Corporation.

7. REFERENCES

[1] T. Freund, R. Iyer, R.E. Schapire, and Y. Singer. "An efficient boosting algorithm for combining preferences," *J. Mach. Learn. Res.*, 4:933-969, 2003.

[2] B. Moghaddam. "Face recognition by humans and machines: a tutorial survey," *CVPR 2001 Short Course*.

[3] B. Moghaddam, T. Jebara, and A. Pentland. "Bayesian face recognition," *Pattern Recogn.*, 33(11):1771-1782, 2000.

[4] X. Huang, S.Z. Li, and Y. Wang. "Jensen-Shannon boosting learning for object recognition," *CVPR 2005*, pp.144-149.

[5] R.E. Schapire and Y. Singer. "Improved boosting algorithms using confidence-rated predictions," *Mach. Learn.*, 37(3):297-336, 1999.

[6] C. Huang, B. Wu, H. Ai, and S. Lao. "Omni-directional face detection based on real AdaBoost," *ICIP 2004*, pp.593-596.

[7] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. "Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition," *ICCV 2005*, pp.786-791.

[8] M.R. Hestenes. "Multiplier and gradient methods," *J. Optim. Theory Appl.*, 4:303-320, 1969.

[9] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss. "The FERET evaluation methodology for face recognition algorithms," *IEEE T. PAMI*, 22(10):1090-1104, 2000.