Boosted Metric Learning for Efficient Identity-Based Face Retrieval
Romain Negrel, Alexis Lechervy, Frédéric Jurie

To cite this version:
Romain Negrel, Alexis Lechervy, Frédéric Jurie. Boosted Metric Learning for Efficient Identity-Based Face Retrieval. 26th British Machine Vision Conference, Sep 2015, Swansea, United Kingdom. British Machine Vision Conference 2015 proceedings, 2015, <10.5244/C.29.139>. <hal-01272248>
Boosted Metric Learning for Efficient Identity-Based Face Retrieval

Romain Negrel
romain.negrel@unicaen.fr
Alexis Lechervy
alexis.lechervy@unicaen.fr
Frederic Jurie
frederic.jurie@unicaen.fr

GREYC, CNRS UMR 6072, ENSICAEN
Université de Caen Basse-Normandie
France

Abstract

This paper presents MLBoost, an efficient method for learning to compare face signatures, and shows its application to the hierarchical organization of large face databases. More precisely, the proposed metric learning (ML) algorithm is based on boosting so that the metric is learned iteratively by combining several weak metrics. Boosting allows our method to be free of any hyper-parameters (no cross-validation required) and to be robust with respect to overfitting. This MLBoost algorithm can be trained from constraints involving two pairs of vectors (quadruplets) with a quadratic complexity. The paper also shows how it can be included in a semi-supervised hierarchical clustering framework adapted to identity based face search. Our approach is validated on a benchmark relying on the Labelled Faces in the Wild (LFW) dataset supplemented with 1M face distractors.

1 Introduction and related Work

The performance of many visual recognition algorithms – e.g. face verification or person re-identification – heavily relies on the metric used to measure the similarity between input data. Recent works such as [5, 10, 12, 25, 31] have shown the interest of learning an optimal metric for a particular task using Metric Learning (ML). Most approaches learn a Mahalanobis metric based on an objective function whose constraints come either from a set of labeled examples or, more frequently, from sets of positive (in the same class) and negative (in different classes) pairs.

Presenting the exhaustive list of recent works that have used metric learning (ML) in such a context would be rather tedious, given their large number. However, we can mention a few of the most notable approaches such as DDML [2], RBML [18], Structural ML [2], PCCA [2], rPCCA [2], LMNN [3], LDML [2], ITML [2], KISSME [2], RS-KISSME [2], SML [2]. The reader interested in ML in general is referred to the recent survey by Bellet et al. [1]. The difference among these methods mainly lies in their objective functions which is task dependent or in the type of constraints of the objective function, (e.g. some require pairwise distances between points making large scale applications difficult).

The works of Shen et al. [2] and Bi et al. [2], which have introduced Boosting based approaches, deserve a particular attention due to the interesting properties they offer: i) they
are efficient and scalable, as no semidefinite programming is required and only the largest eigenvalue and corresponding eigenvectors are needed at each iteration; ii) like AdaBoost, they don’t have any parameters to tune and they are easy to implement as no sophisticated optimization techniques are involved. These approaches hence contrast with most of the commonly used ML methods e.g. [1, 2, 3, 4, 5] for which hyper-parameters, often introduced for regularization purposes, have to be adjusted by cross-validation.

However, one strong limitation of these two approaches [6, 7] is that they require having triplets for learning the metric, i.e., constraints expressed under the form $D(x_a, x_b) < D(x_a, x_c)$, where $x_a, x_b, x_c$ are three input vectors for which the label is known: positive and negative pairs within a constraint have to share a common vector. This is a limitation as for most of the verification task only training pairs are available (e.g., for person re-identification on the Viper dataset [11] only pairs of same/different persons are provided for training and, thus, using such triplets is not possible).

One of the key contributions of this paper is to propose a metric learning approach based on boosting allowing the use of pairs of points for training, i.e. which can use constraints such as $D(x_a, x_b) < D(x_c, x_d)$ while keeping the good properties of the approach in [2] (scalability, simplicity, no parameters, etc.). Our approach is based on the ranking algorithm RankBoost [6], known to offer three particularly interesting features: i) it requires no hyperparameters, ii) it has great robustness to overfitting (explained with a standard VC-dimensional analysis techniques [6, 9]), and iii) it uses a computational trick for reducing the complexity of the learning step. In the following, we show how to build on RankBoost [6] to efficiently address this metric learning problem.

The proposed approach is validated on the identity face retrieval task, such as defined by Bhattarai et al. [6]. This task consists in retrieving the face(s) of the same person to a query, from a large database of faces. The task is evaluated both in terms of scalability and accuracy. The idea presented in [6] bears similarity with the work of Verma et al. [29], who proposed a method for learning hierarchical similarity metrics using a taxonomy. However, unlike Verma et al., the method proposed by Bhattarai et al. does not require any preexisting taxonomy. The experimented validation in [9] shows that using the proposed ML algorithm, allows to improve the performance by a large margin within the same framework as [6].

## 2 Proposed method : from RankBoost to boosted ML

### 2.1 Metric learning with boosting

Our approach, referred to as MLBoost, lies in the family of Mahalanobis ML. This means that the metric $D_W(x_i, x_j) = (x_i - x_j)^\top W (x_i - x_j)$ is parametrized by the positive semi-definite (PSD) matrix $W$ and $(x_i, x_j) \in \mathbb{R}^D \times \mathbb{R}^D$. This metric has two very interesting properties: (i) it is convex in $W$ allowing to build convex objective functions; (ii) it can be used for the dimensionality reduction of the input data. Regarding this last point, we use the following decomposition $W = LL^\top$, with $L \in \mathbb{R}^{D \times d}$ and $d << D$ the dimension of reduced space. The data can be projected as: $y_i = L^\top x_i$ and can be compared with the Euclidean distance: $D_{LL^\top}(x_i, x_j) = (x_i - x_j)^\top LL^\top (x_i - x_j) = \|y_i - y_j\|_2^2$.

In the same way boosting builds strong (good) solutions to a problem by the combination of weak solutions [29], ML with boosting (e.g. [29]) consists in iteratively building a linear combination of weak Mahalanobis distances of rank 1, denoted by $D_Z^{(i)}$ (with rank($Z^{(i)}$) = 1 and tr($Z^{(i)}$) = 1): $D_W(x, y) = \sum_{t=1}^T \alpha^{(i)} D_Z^{(i)}(x, y)$ in the following.
2.2 Metric Learning with Boosting

We assume that the constraints consist of positive and negative pairs such as $D_W(p_{1i}, p_{2i}) < D_W(n_{1j}, n_{2j})$, with $(p_{1i}, p_{2i}) \in \mathcal{P}$ the set of similar pairs and $(n_{1j}, n_{2j}) \in \mathcal{N}$ the set of dissimilar pairs. This ML problem is closely related to the following ranking problem: $H < D \exp$. The objective at each iteration is then to compute the weak metric $D_Z$, which minimizes the following function:

$$\text{loss}(W) = \sum_{C} d_{ij}^{(1)} \left[ D_W(p_{1i}, p_{2i}) \geq D_W(n_{1j}, n_{2j}) \right],$$

with the function $[\pi]$ that returns 1 if predicate $\pi$ holds and 0 otherwise. At each iteration, the weights are updated according to:

$$d_{ij}^{(t+1)} = \frac{d_{ij}^{(t)} e^{\alpha^{(t)}(D_Z(p_{1i}, p_{2i}) - D_Z(n_{1j}, n_{2j}))}}{w^{(t)}},$$

with $D_Z^{(t)}$ the weak metric computed at iteration $t$, $\alpha^{(t)}$ its weight and $w^{(t)}$ the normalization factor chosen so that the constraint (1) is met. Equation (3) and the inequality $[x \geq 0] \leq e^x$, allow us to show that:

$$\text{loss}(W) \leq \prod_{t=1}^{T} w^{(t)}.$$

The objective at each iteration is then to compute the weak metric $D_Z^{(t)}$ and the weight $\alpha^{(t)}$ which minimizes the following function:

$$w^{(t)} = \sum_{C} d_{ij}^{(t)} e^{\alpha^{(t)}(D_Z^{(t)}(p_{1i}, p_{2i}) - D_Z^{(t)}(n_{1j}, n_{2j}))},$$

with $\text{rank}(Z^{(t)}) = 1$, $\text{tr}(Z^{(t)}) = 1$ and $\alpha^{(t)} > 0$. We note that for increasing the performance of the final metric, we only need to have $w^{(t)} < 1$.

**Weak learning algorithm:** The weak metric $D_Z^{(t)}$ is obtained by computing the solution of an approximation of Equation (5). Freund et al. [3] propose to approximate the exponential by the following lower bound:

$$e^{ax} \leq \left( \frac{1+x}{2} \right) e^a + \left( \frac{1-x}{2} \right) e^{-a},$$
for \( x \in [-1, 1] \) and \( \alpha \in \mathbb{R} \). Using this idea, obtain the following optimization problem:

\[
\arg\min_{\alpha, Z} \left( \frac{1-r(Z)}{2} e^\alpha + \frac{1+r(Z)}{2} e^{-\alpha} \right), \tag{7}
\]

where

\[
r(Z) = \sum_C d_{ij}^{(t)} (D_Z(n_{1j}, n_{2j}) - D_Z(p_{1i}, p_{2i})). \tag{8}
\]

The optimization problem (7) is minimized when \( \alpha = \frac{1}{2} \ln \left( \frac{1+r(Z)}{1-r(Z)} \right) \), showing that:

\[
\prod_{t'=1}^T w(t') \leq \sqrt{1-r(Z)^2}, \tag{9}
\]

Shen et al. [24] proposed to use eq. (8) to compute the weak metric by solving the following optimization problem:

\[
Z(t) = \arg\max_Z r(Z) \quad \text{s.t.} \quad \text{rank}(Z) = 1, \text{tr}(Z) = 1 \tag{10}
\]

By making the following variable change: \( Z = zz^T \), the previous problem is equivalent to:

\[
z(t) = \arg\max_z z^T A(t) z \quad \text{s.t.} \quad \|z\|_2 = 1 \tag{11}
\]

with \( A(t) = \sum_C d_{ij}^{(t)} ((n_{1j} - n_{2j})(n_{1j} - n_{2j})^T - (p_{1i} - p_{2i})(p_{1i} - p_{2i})^T) \). The optimization problem (11) is simply solved by the computation of the eigenvector corresponding to the largest eigenvalue of the \( A(t) \) matrix.

**Choosing \( \alpha(t) \):** once the weak metric \( D_Z(t) \) is computed with the aforementioned method, the best weight \( \alpha(t) \) of this weak metric is computed by solving the following problem:

\[
\alpha(t) = \arg\min_{\alpha} \sum_C d_{ij}^{(t)} e^{\alpha (D_Z(t)(p_{1i}, p_{2i}) - D_Z(t)(n_{1j}, n_{2j})).} \tag{12}
\]

This optimization problem is strictly convex and therefore has a unique minimum. We find the solution numerically via the golden section search method. We note that if \( \alpha(t) \) is negative or zero, we stop the algorithm. Indeed, we want the parameter matrix of the final metric to meet the positive-semidefinite constraint.

### 2.3 Reduction of the algorithmic complexity

A naive implementation of the presented MLBoost algorithm would induce a high computational complexity. Indeed, the complexity depends on the cardinality of \( C \), the set quadruplets. Under the assumption that the images are uniformly distributed across identities, we can approximate the cardinality of \( C \) by:

\[
|C| \approx |P||N| \approx |I|^{4/2c}, \tag{13}
\]
The optimization problem (12) allowing to compute \( \alpha^{(t)} \), which are the weights of this weak metric, is rewritten as:

\[
\alpha^{(t)} = \arg \min_{\alpha} \left( \sum_{p} u_{i}^{(t)} e^{\alpha (D_{Z}^{(t)}(p_{1i}, p_{2i}))} \right) \left( \sum_{N} v_{j}^{(t)} e^{-\alpha (D_{Z}^{(t)}(n_{1j}, n_{2j}))} \right).
\]

In this form, the equations have a much lower computational complexity which it depends on the sum of the number of positive and negative pairs : \(|P| + |N|\). The computational complexity is then of the order of \( O(|T|^2) \). The different steps of MLBoost are described in Algorithm 1.

**Algorithm 1** MLBoost with efficient learning implementation

1: \textbf{procedure} MLBOOST(\( X, P, N, \text{itersMax} \))
2: \quad \( t \leftarrow 1 \)
3: \quad \( L^{(t)} \leftarrow \emptyset \)
4: \quad \text{Initialize weights: } u_{i}^{(t)} = 1/|P|, \forall i \in 1 \ldots |P|, v_{j}^{(t)} = 1/|N|, \forall j \in 1 \ldots |N|.
5: \text{repeat}
6: \quad \text{Compute weak metric: } z^{(t)} \text{ by solving of eq. (11) + (16).}
7: \quad \text{Choose the best: } \alpha^{(t)} \text{ by solving of eq. (17).}
8: \quad \text{if } \alpha^{(t)} \leq 0 \text{ then}
9: \quad \quad \text{break}
10: \quad L^{(t+1)} \leftarrow \left[ L^{(t)}, \sqrt{\alpha^{(t)} z^{(t)}} \right]
11: \quad \text{Update weights: } u_{i}^{(t+1)} \text{ and } v_{j}^{(t+1)} \text{ with the eq. (15).}
12: \text{until } t < \text{itersMax}
13: \text{return } L

with \(|I|\) the total number of images and \(c\) the number of identities. The computational complexity, as a function of the number of images, is thus \( O(|I|^4) \).

Interestingly, Freund et al. [9] have introduced a way to rewrite the equations of Rank-Boost in order to reduce the computational complexity. It turns out these principles extend to our MLBoost algorithm. For doing this, we decompose \( d_{ij}^{(t)} \), the weights of quadruplets in \( C \), as follows:

\[
d_{ij}^{(t)} = u_{i}^{(t)} v_{j}^{(t)}, \quad (14)
\]

with \( u_{i}^{(1)} = 1/|P|, \forall i \) and \( v_{j}^{(1)} = 1/|N|, \forall j \). Equation (3) used to update the weights can be rewritten as:

\[
u_{i}^{(t+1)} = u_{i}^{(t)} e^{\alpha^{(t)} D_{Z}^{(t)}(p_{1i}, p_{2i})} / w_{P}^{(t)}, \forall i
\]

\[
v_{j}^{(t+1)} = v_{j}^{(t)} e^{-\alpha^{(t)} D_{Z}^{(t)}(n_{1j}, n_{2j})} / w_{N}^{(t)}, \forall j
\]

with \( w_{P}^{(t)} \) and \( w_{N}^{(t)} \) the normalization factors chosen such that \( \sum u_{i}^{(t+1)} = 1 \) and \( \sum v_{j}^{(t+1)} = 1 \). The expression of matrix \( A^{(t)} \) used to compute the weak metric (11) is rewritten as follows:

\[
A^{(t)} = \sum_{N} v_{j}^{(t)} \left( - (n_{1j} - n_{2j})(n_{1j} - n_{2j})^{\top} \right) - \sum_{P} u_{i}^{(t)} \left( - (p_{1i} - p_{2i})(p_{1i} - p_{2i})^{\top} \right).
\]

The optimization problem (12) allowing to compute \( \alpha^{(t)} \), which are the weights of this weak metric, is rewritten as:

\[
\alpha^{(t)} = \arg \min_{\alpha} \left( \sum_{P} u_{i}^{(t)} e^{\alpha (D_{Z}^{(t)}(p_{1i}, p_{2i}))} \right) \left( \sum_{N} v_{j}^{(t)} e^{-\alpha (D_{Z}^{(t)}(n_{1j}, n_{2j}))} \right).
\]
Algorithm 2

1: procedure HIERARCHICAL_TREE_CLUSTERING($X, y, nbPairs, nbCluster, treeDepth$)
2:   rootNode.idxC ← [1 ... N] and rootNode.idxML ← [1 ... N]
3:   WorkQueue.push(rootNode)
4: while WorkQueue.empty() == false do
5:   node ← WorkQueue.pop()
6:   ($P, N$) ← build_pairs($y[node.idxML]$, nbPairs)
7:   $L ← \text{RankBoost}(X[:], node.idxML, P, N, 1024)$
8:   node$L ← L$
9:   if node.depth < treeDepth then
10:      $X' ← L^\top X[:], node.idxC$
11:      $C ← \text{clustering}(X', nbCluster)$
12:      for $i = 1 \ldots nbCluster$ do
13:         $\text{childNode.idxC} ← \text{cluster_assign}(node.idxC, X', C, i)$
14:         $\text{childNode.idxML} ← \text{ml_selection}(y, \text{childNode.idxC})$
15:         node.addChild(childNode)
16:      WorkQueue.push(childNode)
17: return rootNode

3 MLBoost for Semi-Supervised Hierarchical Clustering

In [2], Bhattarai et al. proposed a method for building hierarchical face representations for efficient identity-based face retrieval. The main idea is to use a ML algorithm for building a tree structured representation of a large set of faces, such that identical-identity faces belong to the same cluster at each tree level. The hierarchical clustering is built recursively: at each level, a new metric is learned using the faces belonging to a particular tree node which is then used to locally cluster the faces and to create the sub-branches.

The retrieval for a new query face is done by visiting the tree from top to bottom until a leaf is reached; if the current node is a non-leaf node, the face is projected into the corresponding subspace and compared with its centroids, which allows to move to the closest child node. When a leaf is reached, the face is compared with all the faces contained in the leaf using the corresponding local metric.

The main drawback of this method is that, when a local metric is learned within a particular node of the tree, no information is retained from the parent nodes. By doing this, the metrics become quickly too local (over specialized). Furthermore, misclassified faces (those going in to the wrong branches) can strongly impact the local metrics.

We therefore propose to improve [2] by combining the metrics learnt at different levels of the tree. A simple method for doing this would be to add an elastic term between the metric from the parent branch and the metric learned. However, the addition of these terms would add hyper-parameters to the metric learning problem and would require some additional tuning. We propose instead a simple method based on the selection of training samples: if one or more examples of an identity belongs to a given node, we use all the examples of this identity to learn the metric as described in Algo. 2. We see that, in each branch, we have two vectors of example indices: the one of the faces actually belonging to the branch (denoted as $\text{idxC}$ in Algo. 2 this vector is the same as that used in the algorithm proposed by [3]); the example indices used to learn the metric in the branch, denoted as $\text{idxML}$ in Algo. 2.
4 Experiments

Performance evaluation. In the considered scenario: a person has to be identified given an image of his face by querying a dataset containing faces of different identities. The performance should reflect the ability of the system to place the correct identity at the beginning of a list of candidate identities. The criterion used to evaluate the performance is the one used in [2], i.e., the mean of \( k \)-call@\( n \) (such as defined in [6]), with \( k = 1 \). We measure the performance for several values of \( n \), i.e., \( n \in \{1, 10, 20, 50, 100\} \).

Database: We use the well known Labeled Faces in the Wild (LFW) database by Huang et al. [16]. This database has more than 13000 images of over 4000 different identity. We use its aligned version with deep funneling [15]. In our experiments, we use the same set of images and queries than the ones used by Bhattarai et al. [2]. We keep only the images of identities with at least five examples each, all other images are not used during the experiments; it results in a set of 5985 images of 423 persons. Regarding the query set, we select one query image per identity (the same as Bhattarai et al.). The training set contains the selected images, excluding the query set. To evaluate the robustness with respect to large-scale databases, we use the set of million distractor faces kindly provided by Bhattarai et al. [2].

Image description: We evaluate the methods with two types of image signatures: (i) Local Binary Pattern (LBP) [20] and (ii) Fisher Vector (FV) [21, 22]. To extract the LBP signature, we use the same protocol as proposed by Bhattarai et al., each image being described by a vector of 9860 dimensions [2, 28]. To extract the FV signature, we use the implementation proposed by Simonyan et al. [25]. We use the publicly available source code\(^1\) with the default parameters and with our own training set. Each image is then described by a vector of 67584 dimensions. The two types of signatures (LBP and FV) are subsequently \( \ell_2 \)-normalized.

Training pairs: We select as many positive pairs as negative ones, uniformly distributed across the identities.

Dimension of the output space: The proposed algorithm does not allow to control explicitly the dimension of the output space, which is required for fair comparisons (i.e., for comparing two approaches for a given dimensionality of the projected features). Thus, we select the \( d \) best projectors by applying a singular value decomposition of the projectors resulting in a \( d \)-dimensional signature.

Baseline (PCA, no metric learning): These experiments are intended to evaluate the performance obtained by the raw signatures (without metric learning), possibly projected in a low dimensional space by PCA. They are reported in Table 1, which compares the performance of the two signatures (LBP and FV), without distractors or with 100.000 and 1M distractors, and for different values of recall \( n \). The performances are given in the mean \( 1\text{-call}@n \) (%). Without distractors, the FV signatures perform better than LBP, while with distractors LBP outperforms FV. Projecting the signatures on a subspace of 4096 dimensions or more does not alter the performance significantly. In the following experiments, we pre-reduce the signatures at 4096 components with the PCA projectors.

Experimental validation of the proposed algorithm (exhaustive search)

Figure 1(a) illustrates some results on convergence of the algorithm (without distractors, using \( 2 \times 32768 \) pairs for training). Interestingly, we do not see any overfitting (this is confirmed by adding even more iterations). Figure 1(b) shows the performance of the algorithm\

---

\(^1\)http://www.robots.ox.ac.uk/~vgg/software/face_desc/
Table 1: Mean 1-call@n performance given by raw LBP and FV signatures (no ML), with/without PCA, with/without distractors.

<table>
<thead>
<tr>
<th>Sign.</th>
<th>Dim.</th>
<th>n=100</th>
<th>n=10</th>
<th>n=20</th>
<th>n=50</th>
<th>n=100</th>
<th>n=100</th>
<th>n=10</th>
<th>n=20</th>
<th>n=50</th>
<th>n=100</th>
<th>n=100</th>
<th>n=10</th>
<th>n=20</th>
<th>n=50</th>
<th>n=100</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>9860</td>
<td>31.9</td>
<td>53.7</td>
<td>60.5</td>
<td>68.4</td>
<td>74.7</td>
<td>31.4</td>
<td>52.7</td>
<td>59.3</td>
<td>67.8</td>
<td>73.3</td>
<td>31.4</td>
<td>52.7</td>
<td>59.3</td>
<td>67.8</td>
<td>73.3</td>
</tr>
<tr>
<td>LBP-PCA</td>
<td>8192</td>
<td>31.9</td>
<td>53.7</td>
<td>60.5</td>
<td>68.4</td>
<td>74.7</td>
<td>31.4</td>
<td>52.2</td>
<td>58.4</td>
<td>66.7</td>
<td>72.1</td>
<td>30.3</td>
<td>49.4</td>
<td>55.8</td>
<td>61.9</td>
<td>69.0</td>
</tr>
<tr>
<td>LBP-PCA</td>
<td>4096</td>
<td>31.7</td>
<td>52.7</td>
<td>60.3</td>
<td>68.1</td>
<td>74.2</td>
<td>15.1</td>
<td>31.4</td>
<td>39.5</td>
<td>51.3</td>
<td>60.3</td>
<td>12.2</td>
<td>29.1</td>
<td>33.3</td>
<td>42.3</td>
<td>50.1</td>
</tr>
<tr>
<td>LBP-PCA</td>
<td>32</td>
<td>16.1</td>
<td>33.3</td>
<td>43.3</td>
<td>55.8</td>
<td>66.7</td>
<td>15.1</td>
<td>31.4</td>
<td>39.5</td>
<td>51.3</td>
<td>60.3</td>
<td>12.2</td>
<td>29.1</td>
<td>33.3</td>
<td>42.3</td>
<td>50.1</td>
</tr>
<tr>
<td>FV</td>
<td>67584</td>
<td>57.9</td>
<td>77.5</td>
<td>83.5</td>
<td>86.6</td>
<td>92.8</td>
<td>31.9</td>
<td>52.2</td>
<td>58.4</td>
<td>66.7</td>
<td>72.1</td>
<td>30.3</td>
<td>49.4</td>
<td>55.8</td>
<td>61.9</td>
<td>69.0</td>
</tr>
<tr>
<td>FV-PCA</td>
<td>8192</td>
<td>55.8</td>
<td>75.2</td>
<td>81.8</td>
<td>89.1</td>
<td>92.9</td>
<td>12.5</td>
<td>21.5</td>
<td>22.9</td>
<td>26.5</td>
<td>30.0</td>
<td>11.1</td>
<td>13.7</td>
<td>14.4</td>
<td>16.3</td>
<td>17.0</td>
</tr>
<tr>
<td>FV-PCA</td>
<td>4096</td>
<td>56.5</td>
<td>75.2</td>
<td>81.8</td>
<td>89.1</td>
<td>92.9</td>
<td>17.4</td>
<td>21.5</td>
<td>22.9</td>
<td>26.5</td>
<td>30.0</td>
<td>11.1</td>
<td>13.7</td>
<td>14.4</td>
<td>16.3</td>
<td>17.0</td>
</tr>
<tr>
<td>FV-PCA</td>
<td>32</td>
<td>21.7</td>
<td>42.1</td>
<td>49.4</td>
<td>62.9</td>
<td>72.8</td>
<td>16.5</td>
<td>25.5</td>
<td>29.8</td>
<td>35.7</td>
<td>40.7</td>
<td>10.4</td>
<td>18.2</td>
<td>20.6</td>
<td>24.8</td>
<td>28.1</td>
</tr>
</tbody>
</table>

(a) Performance vs number of boosting rounds

(b) Performance vs number of training pairs

Figure 1: Performance of the proposed ML approach

as a function of the number of pairs used to learn the metric (no distractors, 1000 rounds of boosting). We observe that the evolution of performance is sub-logarithmic in the number of pairs.

Table 2 compares the performance of the proposed MLBoost algorithm to the performance of the state-of-the-art PCCA (Pairwise Constrained Component Analysis) [2, 19] with and without distractors for different dimensions of the final signature. For fair comparison, we use the signatures kindly given by the authors of [2]. We apply 1000 rounds of boosting. The performance of the proposed approach improves over PCCA by about 20% (LBP, n = 10). We observe, in our experiments, that PCCA tends to overfit and that the performance could certainly be improved by adding a regularization term, such as proposed by Xiong et al. [3]. However, adding a regularization term would add a second hyperparameter to PCCA, whereas our method has no hyper-parameter at all.

<table>
<thead>
<tr>
<th>Sign.</th>
<th>Dim.</th>
<th>n=1</th>
<th>n=10</th>
<th>n=20</th>
<th>n=50</th>
<th>n=100</th>
<th>n=1</th>
<th>n=10</th>
<th>n=20</th>
<th>n=50</th>
<th>n=100</th>
<th>n=1</th>
<th>n=10</th>
<th>n=20</th>
<th>n=50</th>
<th>n=100</th>
</tr>
</thead>
<tbody>
<tr>
<td>FV-MLBoost</td>
<td>32</td>
<td>44.2</td>
<td>70</td>
<td>78.7</td>
<td>85.8</td>
<td>91</td>
<td>28.4</td>
<td>47.3</td>
<td>52.7</td>
<td>60</td>
<td>63.6</td>
<td>20.3</td>
<td>32.6</td>
<td>38.1</td>
<td>44</td>
<td>50.1</td>
</tr>
<tr>
<td>LBP-MLBoost</td>
<td>32</td>
<td>31</td>
<td>54.1</td>
<td>63.4</td>
<td>74</td>
<td>83.2</td>
<td>26.2</td>
<td>43.7</td>
<td>49.2</td>
<td>57.7</td>
<td>64.1</td>
<td>22.9</td>
<td>34.3</td>
<td>39.5</td>
<td>46.1</td>
<td>51.8</td>
</tr>
<tr>
<td>LBP-PCCA [8]</td>
<td>32</td>
<td>-</td>
<td>32.1</td>
<td>41.3</td>
<td>53.4</td>
<td>63.5</td>
<td>-</td>
<td>23.6</td>
<td>28.3</td>
<td>41.1</td>
<td>48.9</td>
<td>-</td>
<td>13.4</td>
<td>18.2</td>
<td>26.7</td>
<td>33.8</td>
</tr>
</tbody>
</table>

Table 2: Comparison with PCCA [8].
Hierarchical tree structure for efficient large-scale identity-based face retrieval

This last part of this experimental section focuses on the use of our MLBoost algorithm for face retrieval, in the context of semi-supervised hierarchical clustering, with/without distractors. Here again, we use LBP and FV signatures and set the number of training pairs to $2 \times 16384$ and the dimension of the output space to 32. We restrict ourselves to a binary clustering tree of 3 and 4 levels (denoted with suffix “-d3” and “-d4”). Table 3 reports the 1-recall@n performance for the two descriptors, for different depths of the tree and for different amounts of distractors.

Searching the faces in only one leaf of the tree affects the overall performance: misclassifications that can happen at any level of the tree cannot be recovered later. This loss increases with the depth of the tree. We can observe this by comparing the performance of the global metric used without and with the tree structure (“MLBoost” in Table 3) (methods whose names include “MLBoost-Tree-d*-Global”). The loss is of about 1-2% at $n=1$ with FV.

We can also observe that the local metric (the lines denoted as “MLBoost-Tree-d*”) improves the performance of the tree-based structure for the small values of $n$, not only compensating entirely the loss induced by the tree but also giving a significant improvement of more than 6% ($n=1$, without distractors, with FV). Comparing our method with the state-of-the-art method of [2] (lines denoted as “PCCA-Tree-d*”) shows that our methods can perform up to 30% better for some configurations. We can also see that the FV signature provides significant gain compared to the LBP signature.

Table 3: Performance of the hierarchical tree structure and comparison with PCCA [2].

<table>
<thead>
<tr>
<th>Sign.</th>
<th>Method</th>
<th>Dim.</th>
<th>$n= 1 $</th>
<th>$n= 10 $</th>
<th>$n= 20 $</th>
<th>$n= 1 $</th>
<th>$n= 10 $</th>
<th>$n= 20 $</th>
<th>$n= 1 $</th>
<th>$n= 10 $</th>
<th>$n= 20 $</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>MLBoost</td>
<td>32</td>
<td>30.7</td>
<td>53.7</td>
<td>62.2</td>
<td>26.2</td>
<td>43.0</td>
<td>48.7</td>
<td>21.5</td>
<td>34.5</td>
<td>38.5</td>
</tr>
<tr>
<td>LBP</td>
<td>MLBoost-Tree-d3-Global</td>
<td>32</td>
<td>27.9</td>
<td>50.4</td>
<td>58.6</td>
<td>24.1</td>
<td>39.5</td>
<td>46.1</td>
<td>20.1</td>
<td>31.4</td>
<td>36.2</td>
</tr>
<tr>
<td>LBP</td>
<td>MLBoost-Tree-d3</td>
<td>32</td>
<td>29.6</td>
<td>54.6</td>
<td>62.2</td>
<td>23.6</td>
<td>39.5</td>
<td>45.2</td>
<td>17.7</td>
<td>31.0</td>
<td>35.2</td>
</tr>
<tr>
<td>LBP</td>
<td>PCCA-Tree-d3 [2]</td>
<td>32</td>
<td>-</td>
<td>33.3</td>
<td>40.9</td>
<td>-</td>
<td>27.4</td>
<td>32.2</td>
<td>-</td>
<td>18.5</td>
<td>23.9</td>
</tr>
<tr>
<td>LBP</td>
<td>MLBoost-Tree-d4-Global</td>
<td>32</td>
<td>27.0</td>
<td>49.2</td>
<td>56.7</td>
<td>22.7</td>
<td>37.4</td>
<td>42.3</td>
<td>18.9</td>
<td>30.5</td>
<td>33.6</td>
</tr>
<tr>
<td>LBP</td>
<td>MLBoost-Tree-d4</td>
<td>32</td>
<td>28.6</td>
<td>52.0</td>
<td>59.3</td>
<td>21.0</td>
<td>35.5</td>
<td>40.9</td>
<td>17.3</td>
<td>26.7</td>
<td>30.5</td>
</tr>
<tr>
<td>LBP</td>
<td>PCCA-Tree-d4 [2]</td>
<td>32</td>
<td>-</td>
<td>36.2</td>
<td>41.6</td>
<td>-</td>
<td>23.4</td>
<td>36.6</td>
<td>-</td>
<td>30.3</td>
<td>31.7</td>
</tr>
<tr>
<td>FV</td>
<td>MLBoost</td>
<td>32</td>
<td>44.7</td>
<td>69.0</td>
<td>78.0</td>
<td>28.6</td>
<td>48.0</td>
<td>52.5</td>
<td>20.1</td>
<td>34.0</td>
<td>38.3</td>
</tr>
<tr>
<td>FV</td>
<td>MLBoost-Tree-d3-Global</td>
<td>32</td>
<td>42.6</td>
<td>65.7</td>
<td>75.7</td>
<td>27.7</td>
<td>45.9</td>
<td>48.9</td>
<td>19.1</td>
<td>33.3</td>
<td>37.1</td>
</tr>
<tr>
<td>FV</td>
<td>MLBoost-Tree-d3</td>
<td>32</td>
<td>48.0</td>
<td>69.0</td>
<td>75.2</td>
<td>30.3</td>
<td>42.8</td>
<td>45.4</td>
<td>22.0</td>
<td>33.1</td>
<td>36.4</td>
</tr>
<tr>
<td>FV</td>
<td>MLBoost-Tree-d4-Global</td>
<td>32</td>
<td>43.0</td>
<td>66.0</td>
<td>75.2</td>
<td>27.6</td>
<td>45.4</td>
<td>48.9</td>
<td>19.1</td>
<td>33.3</td>
<td>36.9</td>
</tr>
<tr>
<td>FV</td>
<td>MLBoost-Tree-d4</td>
<td>32</td>
<td>51.1</td>
<td>70.2</td>
<td>73.8</td>
<td>27.0</td>
<td>39.2</td>
<td>42.3</td>
<td>22.5</td>
<td>30.3</td>
<td>31.7</td>
</tr>
</tbody>
</table>

5 Conclusions

This paper has introduced a boosted ML method called MLBoost that, in contrast with most of ML methods, does not require any hyper-parameters to be tuned and is not subject to overfitting. The paper also shows how to use the proposed MLBoost algorithm for computing hierarchical organizations of face databases allowing to reduce the computational cost of identity-based face retrieval. The proposed experimental validation has demonstrated the robustness of the method w.r.t. overfitting. In addition, its performance is significantly better than PCCA, which is one of the state-of-the-art ML method. Finally, local ML has been shown to provide better performance than global ML in this context.
Acknowledgments

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-12-SECU-0005 (project PHYSIONOMIE).

References


