Some faces are more equal than others: Hierarchical organization for accurate and efficient large-scale identity-based face retrieval

Binod Bhattarai, Gaurav Sharma, Frédéric Jurie, Patrick Pérez

Some faces are more equal than others: Hierarchical organization for accurate and efficient large-scale identity-based face retrieval

Binod Bhattarai[1], Gaurav Sharma[2], Frédéric Jurie[1], Patrick Pérez[2]

GREYC, CNRS UMR 6072, Université de Caen Basse-Normandie, France[1]
Technicolor, Rennes, France[2]

Abstract. This paper presents a novel method for hierarchically organizing large face databases, with application to efficient identity-based face retrieval. The method relies on metric learning with local binary pattern (LBP) features. On one hand, LBP features have proved to be highly resilient to various appearance changes due to illumination and contrast variations while being extremely efficient to calculate. On the other hand, metric learning (ML) approaches have been proved very successful for face verification 'in the wild', *i.e.* in uncontrolled face images with large amounts of variations in pose, expression, appearances, lighting, *etc*. While such ML based approaches compress high dimensional features into low dimensional spaces using discriminatively learned projections, the complexity of retrieval is still significant for large scale databases (with millions of faces). The present paper shows that learning such discriminative projections locally while organizing the database hierarchically leads to a more accurate and efficient system. The proposed method is validated on the standard Labeled Faces in the Wild (LFW) benchmark dataset with millions of additional distracting face images collected from photos on the internet.

1 Introduction

In the present paper, we address the task of identity-based face retrieval: given a query face image, retrieve the face(s) of the same person from a large database of known faces with large changes in face appearances due to pose, expression, illumination, etc. This task finds numerous applications, particularly in indexing and searching large video archives and surveillance videos and in controlling access to resources.

Many appearance features, based on highly localized pixel neighborhoods, have been proposed in the recent literature [1–4]. All of them attempt to capture the statistics of local pixel neighborhoods using either histograms [1, 2, 4] or with higher order statistics [3]. While the more expressive features add some extra performance, Local Binary Patterns (LBP) are attractive because of their extreme computational efficiency. Such efficiency is especially desirable in the case of limited computational capability *e.g.* embedded systems (see comparisons for LBP computation times on different architectures [5]), or in that of very

large datasets *e.g.* millions of faces. In the present paper, we propose to use LBP features as our feature descriptor for the task of large scale identity based face retrieval.

Metric learning based approaches [6–8] have shown that learned low dimensional discriminative projections can be applied for the task of comparing faces with good performances. Such metric learning can be seen as a *global* approach where a single linear projection is learned to discriminate all types of faces. Recently, the SVM-KNN method of Zhang *et al.* [9] has demonstrated (for visual classification task) that learning a collection of *local* (linear) discriminative models leads to better performance. Also, recent Kumar *et al.*'s attribute-based works on facial analysis [10, 11] hint towards the presence of local modes in the (attribute transformed) space of faces. In the same way, Verma *et al.* [12] proposed a novel framework to learn similarity metrics using class taxonomies, showing that nearest neighbor classifiers using the learned metrics get improved performance over the best discriminative methods. Inspired by these previous works, we propose to organize large face databases hierarchically using locally and discriminatively learned projections. More concretely, we propose a semi-supervised hierarchical clustering algorithm, alternating between the two steps of (i) learning local projections and (ii) clustering for splitting the faces into sets of more localized regions in face space. Intuitively, we expect such a hierarchical setup to capture coarse differences, *e.g.* gender, at the top levels and then specialize the different projections at the bottom levels to finer differences between the faces. Fig. 1 gives an overview of our approach in contrast to traditional metric learning. One big difference with [10, 11] or [12] is that our approach does not need any face taxonomy nor predefined set of attributes. Both are automatically discovered.

In the following, we set the context for our work in §2 and then describe our approach in detail in §3. We discuss our approach in relation to the most closely related works in §3.1. We then give qualitative and quantitative experimental results validating our approach in §4 and conclude the paper in §5.

2   Context and related works

Comparing face images of different persons with large variations in appearance, pose, illumination, *etc.*, is a challenging problem. Locally computed features like Local Binary Patterns (LBP), Local Ternary Patterns (LTP) and Local quantized patterns (LQP) have been quite successful to address these kinds of problems [13, 2, 14]. One of the recent state-of-art methods [15] on Labeled Faces in the Wild (LFW) [16], the most challenging face verfication dataset, computes very high dimensional LBP (of dimension as high as 100k). In the recent years, several other variants of LBP have been introduced for different computer vision tasks (*e.g.* [17–20]). In this paper, we use the standard LBP descriptor for a good efficiency and performance trade-off.

Many other recent papers address the problem with novel approaches, *e.g.* discriminative part-based approach by Berg and Belhumeur [21], probabilistic
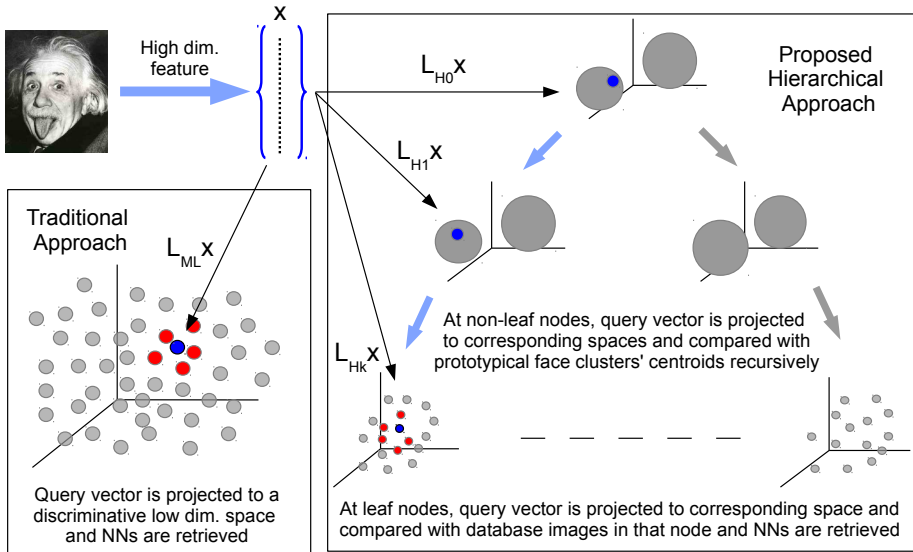
Fig. 1. Principle of the proposed method, in contrast with the traditional metric learning based approach. While the traditional approach learns a single projection (L_{ML}) the proposed approach works hierarchically and learns different projection matrices (L_{H_n}) for different nodes. See §3 for details.

elastic model by Li *et al.* [22], Fisher vectors with metric learning by Simonyan *et al.* [7], novel regularization for similarity metric learning by Cao *et al.* [23], fusion of many descriptors using multiple metric learning by Cui *et al.* [24], deep learning by Sun *et al.* [25], method using fast high dimensional vector multiplication by Barkan *et al.* [26] or robust feature set matching for partial face recognition by Weng *et al.* [27]. Many of the most competitive approaches on LFW combine different features, *e.g.* [28–30] and/or use external data, *e.g.* [10, 31].

Metric learning has been recently shown to give good results on very diverse computer vision tasks [32–36]. We refer the reader to Bellet *et al.* [37] for an excellent survey on Metric Learning. More specifically, methods based on metric learning have been reported to improve accuracy for face verification, either on static images [23, 28, 8, 7] or on videos [38]. The key idea is to learn a Mahalanobis like metric of the form $D^2_M(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top M(\mathbf{x}_i - \mathbf{x}_j)$, parametrized by the symmetric positive semi-definite (PSD) matrix M, to compare any two faces (described with some features) $\mathbf{x}_i$ and $\mathbf{x}_j$. The learning is based on optimizing some loss function which penalizes high distance between positives and small distance between negative pairs (see [37] for a survey of different metric learning methods/objectives). Since maintaining M as PSD is usually computationally expensive, M is often factorized as $M = L^\top L$. Then the problem can be seen as a linear embedding problem where the features are embedded in the row space of L and compared with the Euclidean distance there:

$$D^2_L(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top L^\top L(\mathbf{x}_i - \mathbf{x}_j) = \|L\mathbf{x}_i - L\mathbf{x}_j\|_2^2. \tag{1}$$

Local metric learning, *e.g.* learning a metric as a function of input vector, has also been studied [39]. However, this is expensive, specially in large scale as comparison with every instance will require projecting the query with a different matrix *vs.* only one projection in the case of a global metric.

Closely related to our work, hierarchically organized (metric) learning systems have also been explored in the past, *e.g.* the works by Hwang *et al.* [40], Deng *et al.* [41], Zheng *et al.* [42], Verma *et al.* [12]. However, they assume the presence of a taxonomy (most often a natural semantic taxonomy), while here we do not assume any such information. Our method is also related to clustering in general and with side information in particular [43–46], the side information here being in the form of (sparse) pairwise *must-link* and *must-not-link* constraints. The goal of many of these works is to learn a metric to improve the performance of clustering with an implicit assumption that the constraints relate directly to the clusters. While in the current work, the metric learning with constraints relates to a first level of embedding which can be thought of a person identity space and then the clustering is done in such identity space. So, unlike previous works, it will be normal in our approach that two must-not-link vectors (faces of different persons) get assigned to same cluster as long as these different people share similar facial traits.

We are interested in the problem of comparing faces using learned metrics. In particular, we are interested in identity-based face retrieval with a focus on accuracy and efficiency of the setup for large-scale scenarios, *i.e.* with hundreds of thousands of distractors. As such, in addition to the above mentioned works on facial analysis, our method is also related to the SVM-KNN method of Zhang *et al.* [9] and to works on large scale image retrieval using product quantization of Jégou *et al.* [47]. We postpone discussing our method in the context of these methods to §3.1, after describing our method in the next section.

3   Approach

We work in the semi-supervised scenario where we have some annotated training pairs $\mathcal{A} = \{(\mathbf{x}_i, \mathbf{x}_j), y_{ij}\}$ with $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^D$ being features for face examples i, j resp. (*e.g.*, Local Binary Patterns [1]) and $y_{ij} = 1$ if the image pairs are of the same person and $y_{ij} = -1$ otherwise. We propose to learn a hierarchical organization of the faces for efficient face retrieval. Note that we assume the annotations are sparse, in the sense that only a very small fraction of pairs in the database is annotated.

We aim at exploiting the similarities between faces of different persons. In our hierarchical layout, we would like to first split the faces into groups based on coarse appearance similarities, *e.g.* gender, and then, at finer level, we would like to learn to discriminate between finer details in coarsely similar faces. We now discuss the case of a binary tree but the method could be applied for arbitrary k-ary trees. We start by taking all the faces into one node and learn a discriminative subspace using margin maximizing metric learning: we minimize a logistic loss function using the recently proposed Pairwise Constrained Prin-

cipal Components (PCCA) [8] approach. In particular, we solve the following optimization,

$$\min_{L} E(L) = \sum_{\{(i,j)\}} \ell_\beta \left(y_{i,j}(\mathcal{D}_L^2(\mathbf{x}_i, \mathbf{x}_j) - 1) \right), \tag{2}$$

where $\ell_\beta(x) = \frac{1}{\beta} \log(1 + e^{\beta x})$ is the generalized logistic loss,

$$\mathcal{D}_L^2(\mathbf{x}_i, \mathbf{x}_j) = \|L(\mathbf{x}_i - \mathbf{x}_j)\|_2^2 \tag{3}$$

is the distance in the row space of the projection matrix L and sum is taken over all labeled face pairs. The intuition of such metric learning formulation is that we would like to find a subspace (parametrized by the projection matrix L) where the distance between the positive pairs is small and that between the negative pairs is large.

We then obtain the projected features $X_p = LX$, where $X = [\mathbf{x}_1, \dots, \mathbf{x}_N]$ is the matrix of all face features in the database, and use k-means to cluster X_p into two clusters in the projected space. By doing this we hope to cluster the faces based on relatively coarse similarities. Once we have the clustering, we create two child nodes of the root containing only the faces from the two clusters respectively. We then repeat the process at each of the child nodes, working with faces in the current node only. At each node we save the indices of the faces which belong to the node along with the current projection matrix and cluster centroids (for the non-leaf nodes). We continue the process until a certain depth, which is a free parameter, is achieved. Algorithm 1 gives the pseudocode for the learning algorithm.

Once the hierarchical structure is built, the retrieval for a new query face is done by traversing the tree with the following decision rule at each node: if it is a non-leaf node, project the face into its subspace and compare with the centroids and move to the closest child node (recall there is a child node for every cluster). If it is a leaf node, then project the face to its subspace and compare with all the faces in that node (projected onto the same subspace) and return the list of the nearest neighbors. Fig. 1 gives an illustration of the retrieval process.

3.1   Relation with closely related works

Recently, Zhang *et al.* [9] proposed the SVM-KNN method, which for a test example creates on-the-fly a local discriminative support vector machine (SVM) classifier, based on its nearest neighbors. The motivation is that a complex non-linear decision boundary could be approximated with a piece-wise linear decision boundary. Also recently, many works based on 'local' comparisons, *e.g.* attribute based works of Kumar *et al.* [10, 11] where the faces are represented as vectors of confidences for the presence of attribute like long hairs, open mouth, *etc.*, have been shown to be important. We could imagine that the faces with such attributes would occupy a local region (or perhaps manifold) in the full face space and, thus, the success of such facial analysis system motivates us to work locally in the face space. Also, the success of SVM-KNN reassures us of the merit

---

Algorithm 1 Learning local metrics and organizing face database hierarchically.

1: Input: (i) Set of face features $X = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$, (ii) Sparse pairwise annotation $\mathcal{A}$, (iii) Height of the tree h, (iv) Dimensions of local projection subspaces at different depths/levels $\{D_0, \ldots, D_h\}$
2: Initialize: $n \leftarrow 0$, idxs$\leftarrow (1, \ldots, N)$, tree$\leftarrow \emptyset$
3: queue.add(n, idxs) {Tree construction in a breadth-first manner}
4: **while** $n < 2^h - 1$ **do**
5:    n, idxs $\leftarrow$ queue.pop()
6:    $\ell \leftarrow \lceil \log_2 n \rceil$ {Current level/depth}
7:    $L_n \leftarrow$ learn_metric($X[:, \text{idxs}], \mathcal{A}[\text{idxs}], D_\ell$)
8:    **if** $\ell < h$ **then**
9:      $C_1, C_2 \leftarrow$ cluster($LX[:, \text{idxs}], 2$)
10:     idxs$_1$, idxs$_2 \leftarrow$ cluster_assign($X[:, \text{idxs}], C_1, C_2$)
11:     queue.add($n + 1$, idxs$_1$)
12:     queue.add($n + 2$, idxs$_2$)
13:    **else**
14:     $C_1, C_2 \leftarrow \emptyset$
15:    **end if**
16:    tree.add_node($\{n, L_n, \text{idxs}, C_1, C_2\}$)
17: **end while**

---

of a local strategy. In our case, such locality is automatically discovered in a data driven way. In the upper levels of the tree, the Vonoroi cells, corresponding to the clustering in the respective discriminative spaces of the nodes, can be interpreted as such local regions where the faces are similar in a coarse way, *e.g.* one node could be of female faces *vs.* another of that of males. While as we go down the levels we expect such differences to become more and more subtle. We show later that qualitative results support our intuition. Hence, we could hope that concentrating on a local region (towards the bottom of the tree) where faces differ very slightly could help us discriminate better, perhaps even at a cheaper cost.

Another closely related but complementary stream of work is that of product quantization by Jégou *et al.* [47]. They propose to learn, in an unsupervised fashion, very compact binary codes to represent images and do very fast nearest neighbor retrieval at large scale. The key point is that they assume/expect the feature space to be Euclidean. However, face retrieval by directly comparing the image representations with Euclidean distance is not optimal and learning a Mahalanobis metric or equivalently a projection is required. Upon projecting the faces to such a space, Euclidean distance can be used and hence product quantization can be applied. As we have already discussed before, the proposed method can be seen as learning different projections for different local regions, we could use different product quantizations in corresponding different local regions found by the proposed method. Hence, the proposed method and product quantization are complementary to each other.

Finally, it worth comparing our approach to the recent work of Verma *et al.* [12], who proposed a framework for learning hierarchical similarity metrics

using class taxonomies. Interestingly, they show that nearest neighbor classifiers using the learned metrics get improved performance over Euclidean distance-based k-NN and over discriminative methods. Our approach bears similarity with [12] as we also learn a hierarchy of similarity metrics. However, a notable difference is that our approach does not require any taxonomy. This is a big advantage as defining a taxonomy of faces would be more than challenging. Providing sufficient training annotations (*i.e.* sufficient number of faces for each level of the hierarchy) would be another complication.

4   Experimental results

Metric Used. We are interested in the task of identity based face retrieval, *i.e.* given a query face images, retrieving face(s) of the same person from a large database of known face images. Our objective is to find the same person and so, for us, it suffices if at least one of the retrieved faces is of the same person. In the ideal case, the top ranked retrieved face would be of the same person, but it would make a practical system if the correct face is ranked in the top n images, for a small value of n, as they can be manually verified by an operator. Hence, we propose to evaluate the method for k-call@n [48] (with $k = 1$): the metric is 1 if at least k of the top n retrieved results are relevant. We average this over our queries and report the mean 1-call@n.

Database and query set. We use the aligned version [29] of the Labeled Faces in the Wild (LFW) database by Huang *et al.* [16]. The dataset has more than 13000 images of over 4000 persons. In addition to LFW, for large-scale experiments, we add up to one million distractor faces that were obtained by crawling Flickr.com and retaining face detection with high confidences. We select the persons/identities in LFW which have at least five example images and randomly sample one image each from them to use as our query set. We use all the LFW images except the query set to learn our system. The results are reported as the mean performance (1-call@n) over all the queries. All the evaluation is done with LFW annotations and, as the distractor images are from personal image collections from the internet while LFW images are that of well-known/celibrities, it is assumed that the distractors do not have the same identities as the query images.

Image description. To describe the images we use the Local Binary Pattern (LBP) descriptors of Ojala *et al.* [1]. We use grayscale images and centre crop them to size 170×100 pixels and do not do any other preprocessing. We use the publicly available `vlfeat` [49] library for LBP, with cell size parameter set to 10, of dimension 9860 for a face image.

Baseline parameter. To set the dimension of the baseline projection matrix we did preliminary experiments, with the standard protocol of LFW dataset, with values in $\{16, 32, 64, 128\}$ and found the performance (verification on LFW test set) saturated for d greater than 32. Hence we fixed the projection dimension to 32.

Tree parameters. We fixed the learned tree to be a binary tree and also fixed the dimension of the projection at successive levels to differ by a multiplicative factor of 2. Thus, the two parameters for the proposed hierarchical organization are the tree depth and the starting projection dimension. We report experiments with depths of 3 and 4, and with starting projection dimension of 128 and 256, leading to leaf nodes with dimensions 32 (same as baseline) in two cases and 16 (half of baseline) in one case. We discuss further in the §4.2.

4.1   Qualitative Results

Fig. 2 shows some example images from the 16 nodes obtained with a tree of depth 4. The clusters shown correspond to the ordering of the leaf nodes at the bottom, *i.e.* every odd cluster and its next neighbor were grouped together in the previous level in the tree and so on. We can note how similar faces are grouped together successively in the different levels of the tree. Cluster 1–12 are predominantly male faces, cluster 13–16 are females. Cluster 15 seems to specialize to females with bangs (hair over the forehead) and 14 on short hair and smiling females. Cluster 2 seems to have bald (or with very little hair) males who wear glasses while cluster 11 has males with smiling faces. With such semantically interpretable visual qualitative results, we conclude that the method seems to perform an attribute-based clustering.

4.2   Quantitative Results

Fig. 3 shows the performances of the baseline *vs.* the proposed method for three different configurations of (i) starting projection dimension 128 with tree depth 3, denoted '128-d3', (ii) starting projection dimension 128 with tree depth 4, denoted '128-d4', and (iii) starting projection dimension 256 with tree depth 4, denoted '256-d4'.

We note that the different configurations of the proposed method give different time complexities. The 128-d3 and 256-d4 trees have leaf node projection dimensions of 32 (same as baseline) with 4 and 8 leaf nodes respectively while the 128-d4 tree has a projection dimension of 16 with 8 nodes. The time complexity for the proposed method depends on (i) projection and Euclidean distance computation with two centroids at non-leaf nodes (repeated $(h-1)$ times, where h is the height of the tree) and (ii) projection and Euclidean distance computation with all the database vectors in leaf nodes. The leaf nodes have about the same number of database vectors and hence a tree with same leaf node projection dimension (of 32) as baseline but with 4 (8) nodes is expected to be 4× (8×) faster than baseline as the bottleneck in large-scale scenario is the computation of Euclidean distances with a large number of (compressed) database vectors.

We observe that as more and more distractors are added the proposed method performs better. In the presence of large number of distractors, 100 nearest neighbor are expected to lie in a smaller region around the query points and hence an explanation for the better performance of the method could be that it is better adapted to local neighborhood. In the zero distractor case, we observe that the
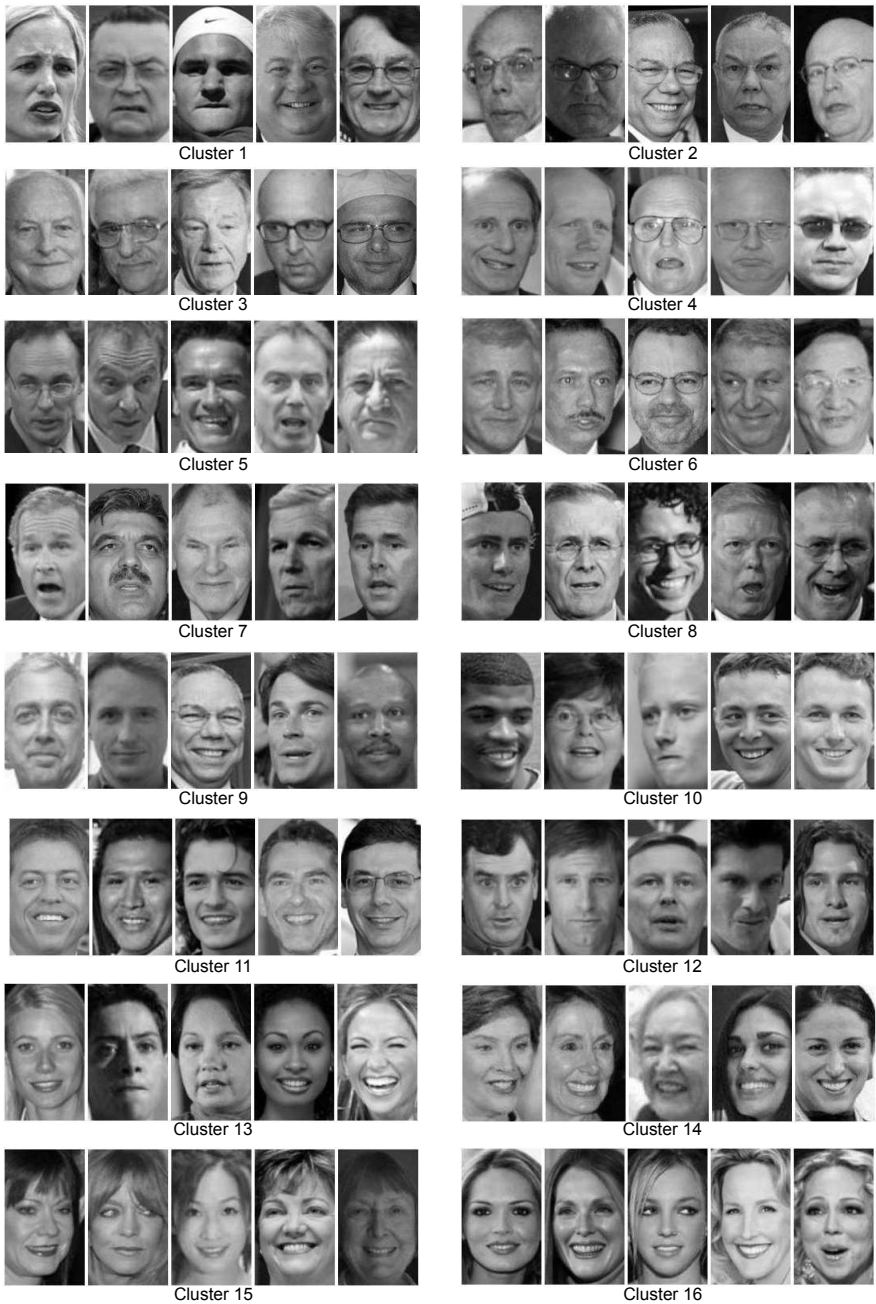
Fig. 2. Visualization the clustering obtained at leaf nodes for a tree of depth 4. The clusters are ordered from left to right and top to bottom, *i.e.* top eight (bottom eight) clusters together form the left (right) node at the first split. Images are randomly selected.
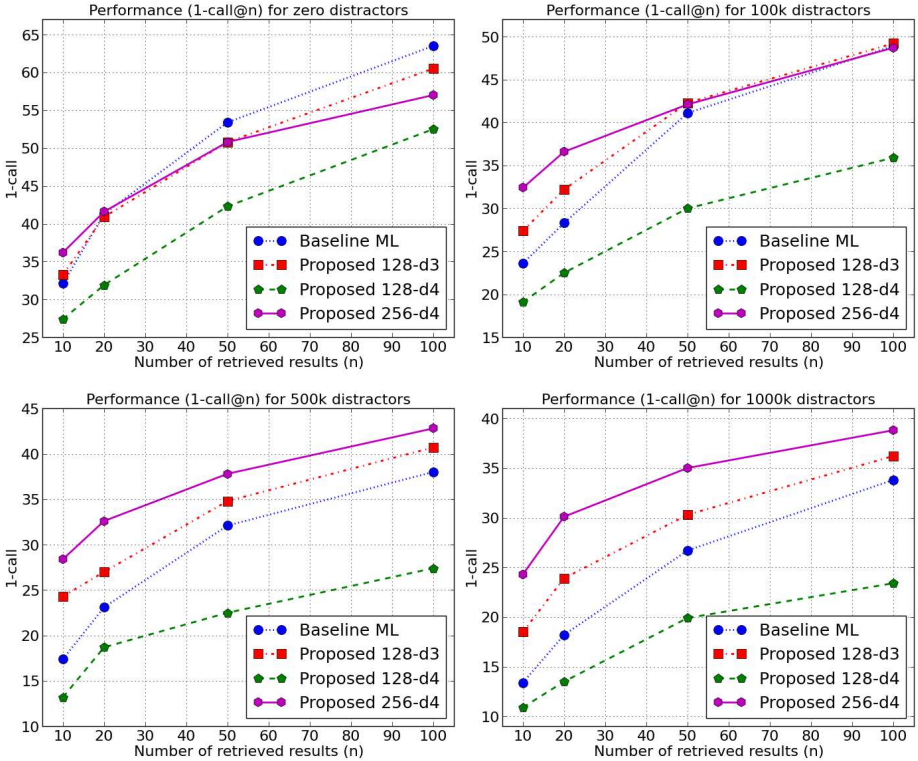
Fig. 3. The performance of the baseline method and that of the proposed method for three different combinations of parameters (starting projection dimension and tree depth) for different numbers of distractors (0, 100k, 500k and 1m) at different operating points.

proposed method is better in the case of small n, *i.e.* it is able to do relatively better retrieval when smaller neighborhoods are considered, while the baseline performs better when n is large and hence larger neighborhoods are considered. The success of the method in the presence of a large number of disctractors underlines the need for locally adapted metrics for identity based face retrieval, especially in a large scale scenario.

Time complexity. The proposed method is expected to be faster in the large scale setting where the number of vectors in the database is greater than the feature dimension. In that case the cost of projecting the query becomes negligible compared to the cost of computing the nearest neighbors in the projected space. Assuming the database vectors uniformly occupy the leaf nodes, a tree with N leaves is then expected to give an N fold speed-up. We carried out all our experiments on a computer with Intel Xeon 2.8 GHz CPU running linux. Empirically we obtain speedups of about 2.8×, 5.9× and 10.2× for trees with

4, 8 and 16 nodes respectively, with our unoptimized Python implementation for the experiments with one million distractors, with all computations being timed with data in RAM.

5   Conclusions

We presented a method for accurate and efficient identity based face retrieval, which relies on a hierarchical organization of the face database. The method is motivated by the recent works on local learning of discriminative decision boundaries and of metrics, and works based on attributes. We showed quantitatively that organizing faces hierarchically, with automatically learned hierarchy, leads to an attribute based clustering of faces. Further, we showed quantitatively that the method is capable of better retrieval at a better time complexity compared to the baseline method in large-scale setting.

References

1. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. PAMI **24**(7) (July 2002) 971–987
2. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Transactions on Image Processing **19**(6) (2010) 1635–1650
3. Sharma, G., ul Hussain, S., Jurie, F.: Local higher-order statistics (lhs) for texture categorization and facial analysis. In: ECCV. (2012)
4. Hussain, S.U., Triggs, B.: Feature sets and dimensionality reduction for visual object detection. In: BMVC. (2010)
5. López, M.B., Nieto, A., Boutellier, J., Hannuksela, J., Silvén, O.: Evaluation of real-time LBP computing in multiple architectures. Journal of Real-Time Image Processing (2014) 1–22
6. Guillaumin, M., Verbeek, J., Schmid, C.: Is that you? Metric learning approaches for face identification. In: ICCV. (2009)
7. Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A.: Fisher vector faces in the wild. In: BMVC. (2013)
8. Mignon, A., Jurie, F.: PCCA: A new approach for distance learning from sparse pairwise constraints. In: CVPR. (2012)
9. Zhang, H., Berg, A.C., Maire, M., Malik, J.: SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. In: CVPR. (2006)
10. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: ICCV. (2009)
11. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Describable visual attributes for face verification and image search. PAMI **33**(10) (2011) 1962–1977
12. Verma, N., Mahajan, D., Sellamanickam, S., Nair, V.: Learning hierarchical similarity metrics. In: CVPR. (2012)

13. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: ECCV. (2004)
14. Hussain, S.U., Napoléon, T., Jurie, F., et al.: Face recognition using local quantized patterns. In: BMVC. (2012)
15. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: CVPR. (2013)
16. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst (October 2007)
17. Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Computer vision using local binary patterns. Volume 40. Springer (2011)
18. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with center-symmetric local binary patterns. In: Computer Vision, Graphics and Image Processing. Springer (2006) 58–69
19. Xie, X.: A review of recent advances in surface defect detection using texture analysis techniques. Electronic Letters on Computer Vision and Image Analysis **7**(3) (2008) 1–22
20. Ojansivu, V.: Blur invariant pattern recognition and registration in the Fourier domain. PhD thesis (2009)
21. Berg, T., Belhumeur, P.N.: POOF: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation. In: CVPR. (2013)
22. Li, H., Hua, G., Lin, Z., Brandt, J., Yang, J.: Probabilistic elastic matching for pose variant face verification. In: CVPR. (2013)
23. Cao, Q., Ying, Y., Li, P.: Similarity metric learning for face recognition. In: ICCV. (2013)
24. Cui, Z., Li, W., Xu, D., Shan, S., Chen, X.: Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In: CVPR. (2013)
25. Sun, Y., Wang, X., Tang, X.: Hybrid deep learning for face verification. In: ICCV. (2013)
26. Barkan, O., Weill, J., Wolf, L., Aronowitz, H.: Fast high dimensional vector multiplication face recognition. In: ICCV. (2013)
27. Weng, R., Lu, J., Hu, J., Yang, G., Tan, Y.P.: Robust feature set matching for partial face recognition. In: ICCV. (December 2013)
28. Guillaumin, M., Verbeek, J., Schmid, C.: Is that you? Metric learning approaches for face identification. In: ICCV. (2009)
29. Wolf, L., Hassner, T., Taigman, Y.: Similarity scores based on background samples. In: ACCV. (2009)
30. Nguyen, H.V., Bai, L.: Cosine similarity metric learning for face verification. In: ACCV. (2010)
31. Berg, T., Belhumeur, P.N.: Tom-vs-pete classifiers and identity-preserving alignment for face verification. In: BMVC. (2012)
32. Bar-Hillel, A., Hertz, T., Shental, N., Weinshall, D., Ridgeway, G.: Learning a Mahalanobis metric from equivalence constraints. JMLR **6**(6) (2005)
33. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: ICML. (2007)
34. Frome, A., Singer, Y., Sha, F., Malik, J.: Learning globally-consistent local distance functions for shape-based image retrieval and classification. In: CVPR. (2007)
35. Weinberger, K., Blitzer, J., Saul, L.: Distance metric learning for large margin nearest neighbor classification. In: NIPS. (2006)
36. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning with application to clustering with side-information. In: NIPS. (2003)

37. Bellet, A., Habrard, A., Sebban, M.: A survey on metric learning for feature vectors and structured data. arXiv.org (2013)
38. Cinbis, R.G., Verbeek, J., Schmid, C.: Unsupervised metric learning for face identification in TV video. In: ICCV. (2011)
39. Wang, J., Kalousis, A., Woznica, A.: Parametric local metric learning for nearest neighbor classification. In: NIPS. (2012)
40. Hwang, S.J., Grauman, K., Sha, F.: Semantic kernel forests from multiple taxonomies. In: NIPS. (2012)
41. Deng, J., Berg, A.C., Fei-Fei, L.: Hierarchical semantic indexing for large scale image retrieval. In: CVPR. (2011)
42. Zheng, L., Li, T.: Semi-supervised hierarchical clustering. In: ICDM. (2011)
43. Zeng, H., Song, A., Cheung, Y.M.: Improving clustering with pairwise constraints: a discriminative approach. Knowledge and information systems **36**(2) (2013) 489–515
44. Sublemontier, J., Martin, L., Cleuziou, G., Exbrayat, M.: Integrating pairwise constraints into clustering algorithms: optimization-based approaches. In: ICDMW. (2011)
45. Wang, X., Davidson, I.: Flexible constrained spectral clustering. In: SIGKDD. (2010)
46. Jain, A.K.: Data clustering: 50 years beyond k-means. PRL **31**(8) (2010) 651–666
47. Jegou, H., Douze, M., Schmid, C.: Product quantization for nearest neighbor search. PAMI **33**(1) (2011) 117–128
48. Chen, H., Karger, D.R.: Less is more: probabilistic models for retrieving fewer relevant documents. In: Special Interest Group in Information Retrieval. (2006)
49. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/ (2008)