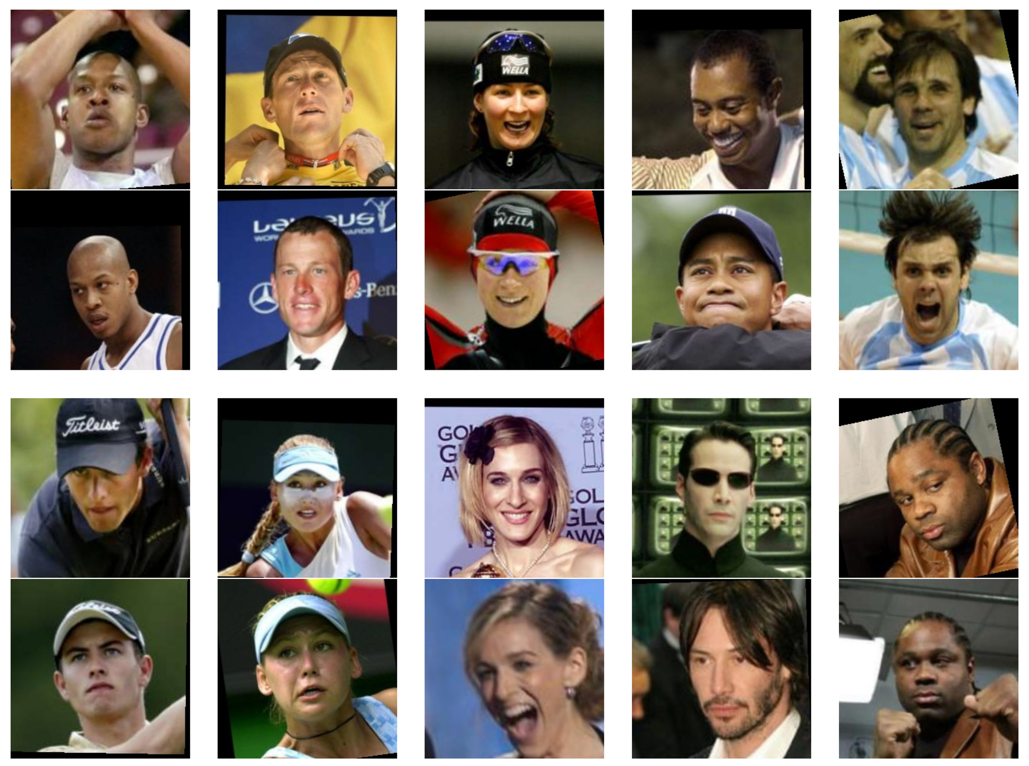
Is that you? Metric Learning Approaches for Face Identification

Matthieu Guillaumin, Jakob Verbeek, Cordelia Schmid
LEAR, INRIA Grenoble, LJK, France

Summary

- Methods for visual identification: are these two images of the same individual ?
- The problem is binary: same class=positive, different class=negative.
- Our approach: use metric learning (ML) to find a Mahalanobis distance that separates positive and negative pairs.
- We introduce a logistic discriminant-based approach to perform metric learning, **LDML**.
- Visual identification is also possible in a k-nearest neighbor approach, leading to Marginalized kNN, **MkNN**.
- We obtain state-of-the-art results on a challenging data set of faces, *Labeled Faces in the Wild* (LFW), consisting of real-world face images from Yahoo! News.



Several example face pairs from the same person from LFW. The top row shows pairs that our method correctly classified, the bottom shows failure cases.

Labeled Faces in the Wild

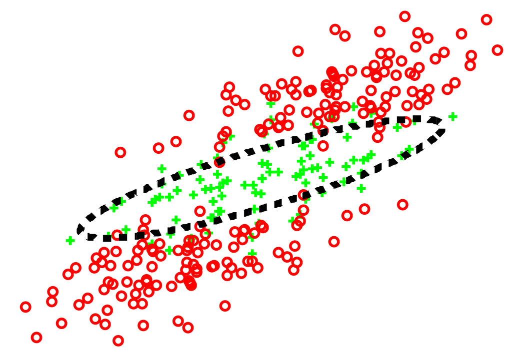
- 13.233 faces of 5749 people – 1680 people with $\geq$2 faces
- 10 independent folds for cross-validation: the people in a fold are not present in the 9 other folds.
- 2 settings for training models:
  restricted (r): 600 fixed *pairs* per fold, no other info.
  unrestricted (u): between 1016 and 1783 *faces* with labels per fold, we can build thousands of pairs.
- `http://vis-www.cs.umass.edu/lfw/`

Logistic Discriminant Metric Learning

We want to learn a Mahalanobis distance $d_{\mathbf{M}}$ that makes images of positive pairs closer than those of negative pairs.

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j),$$

where $\mathbf{M}$ is a symmetric semi-definite positive matrix.



This is equivalent to finding a separating ellipsoid in the space of data differences.

We model the probability that pair (i, j) is positive with

$$p(y_i = y_j | \mathbf{x}_i, \mathbf{x}_j; \mathbf{M}, b) = \sigma(b - d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j)),$$

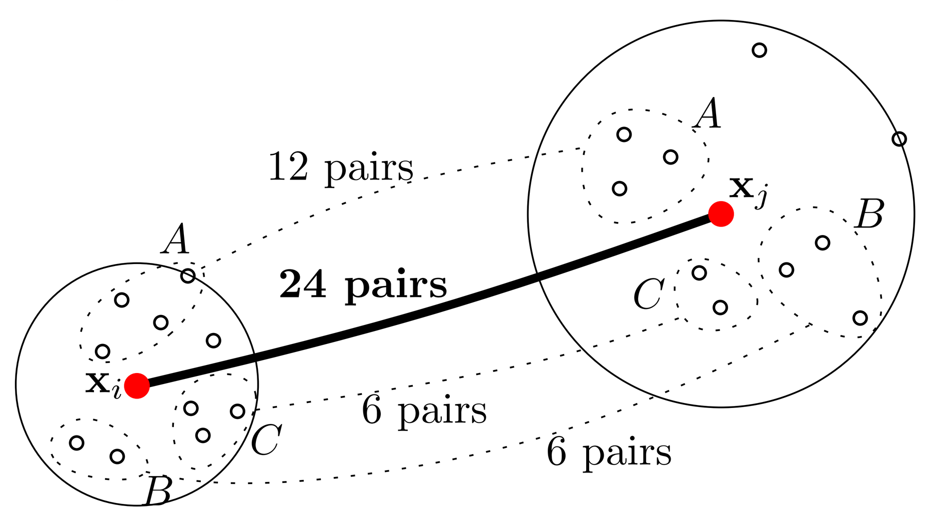
where $\sigma(z) = (1 + \exp(-z))^{-1}$ is the sigmoid function and b a bias term (the optimal distance threshold).
This is a standard logistic regression model (linear in $\mathbf{M}$) on which we perform maximum likelihood estimation via projected gradient descent, to enforce convex constraints on $\mathbf{M}$ (diagonality, SDP, ...).

Marginalized kNN

We adapt kNN classification to classify pairs of unknown classes by marginalizing kNN classification over classes:

$$p(y_i = y_j | \mathbf{x}_i, \mathbf{x}_j) = \sum_c p(y_i = c|\mathbf{x}_i) p(y_j = c|\mathbf{x}_j) = k^{-2} \sum_c n_c^i n_c^j.$$

where n_c^i is the number of neighbors of $\mathbf{x}_i$ with class c.
The classification score for MkNN is the number of positive pairs that can be formed from the neighborhoods of the two images.



The neighborhoods of $\mathbf{x}_i$ and $\mathbf{x}_j$ share three classes A, B and C, from which we can form 24 positive pairs out of 100 possible pairs.

Metric learning is also used, albeit one that optimizes kNN classification (LMNN [1]). Contrary to LDML, this setup requires training data with class labels, and not only labeled pairs.

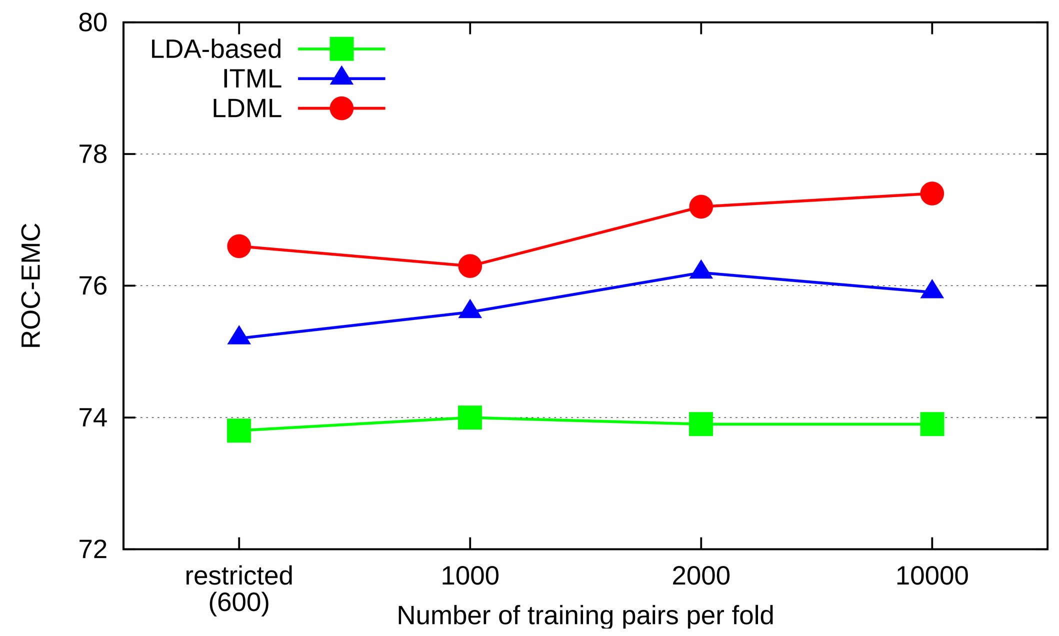
Experiments and results

Features are extracted at 9 positions on the face using SIFT descriptors at 3 different scales: 3456-D face descriptor (**SFD**). PCA is used to reduce dimensionality.

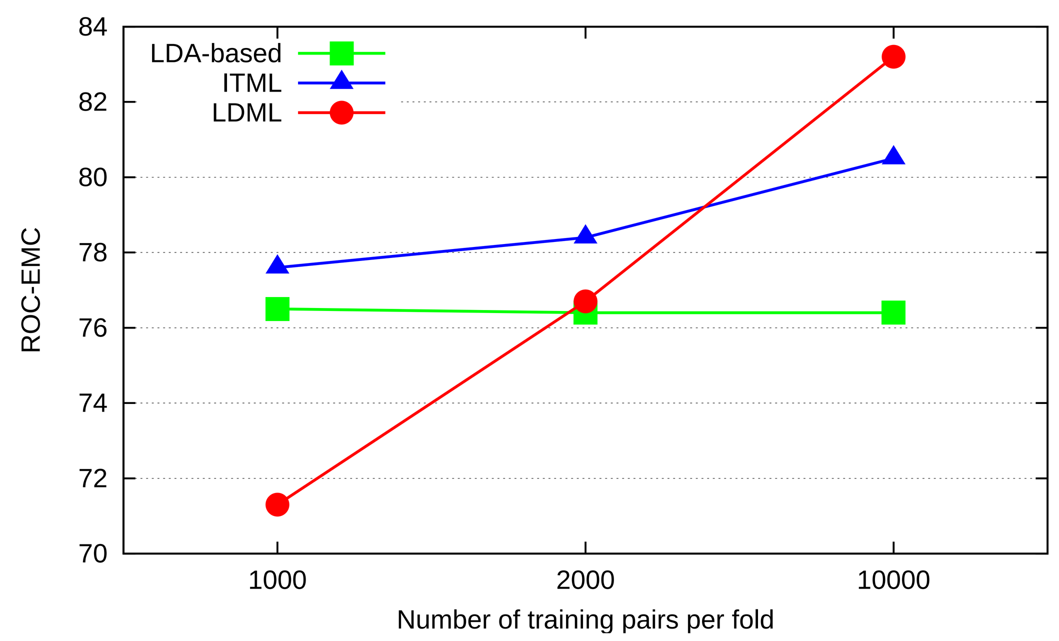
Face identification on LFW

Reported performances are ROC-EMC (best operating point of ROC using equal misclassification cost) and accuracy (as defined by the LFW protocol).

Influence of training set size for low dimensionality PCA – 35 PCA dimensions kept



Influence of training set size for high dimensionality PCA – 100 PCA dimensions kept



Comparison in the unrestricted setting
Using SFD alone, best ROC-EMC from best PCA dim.

ITML [2]	LDA-based	LMNN	LDML	MkNN
80.5 ± 0.5	79.3 ± 0.3	80.5 ± 0.5	$\mathbf{83.2 \pm 0.4}$	**83.1** $\pm$ **0.5**

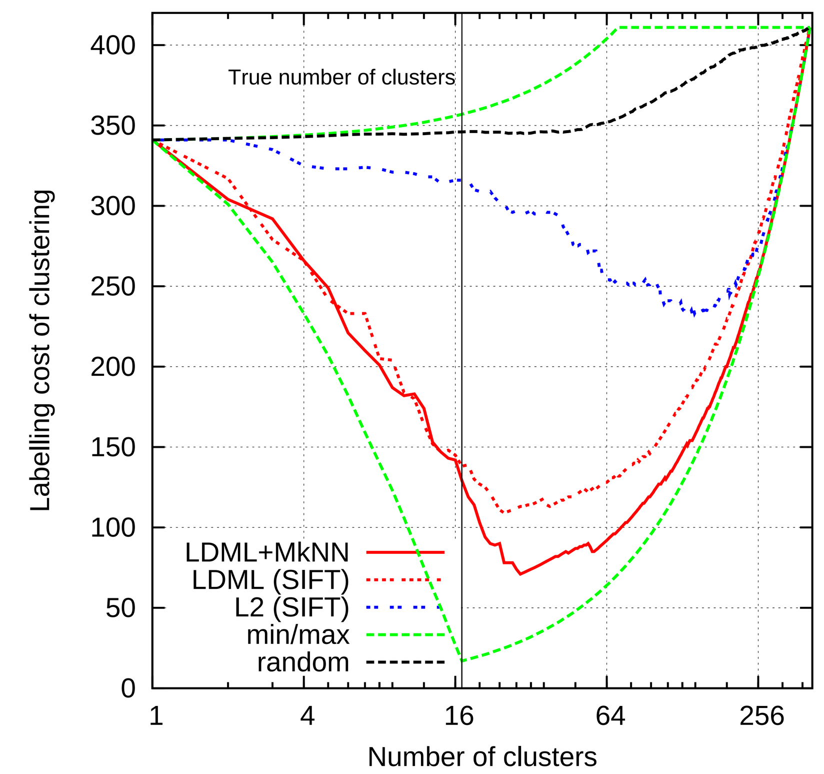
Comparison to state-of-the-art
Linear combination of several descriptors

Method	[3] (r)	LDML (r)	LDML+MkNN (u)
Accuracy	78.47 ± 0.5	79.27 ± 0.6	$\mathbf{87.50 \pm 0.4}$

Clustering

Metrics learnt on 9 folds, applied to the held-out fold. We focus on the 17 most frequent persons ($\geq$10 images). The 17 people account for 411 faces. Clustering is performed using hierarchical agglomeration.

Example of a typical cluster obtained with LDML+MkNN at minimum labeling cost (see graph below). It is pure except for the last two faces.



Labeling cost is the minimum number of clicks a user has to perform to label the data: +1 to name a cluster, +1 to rename a single image.

Recognition from one example

Example face (a), correctly recognised (b), incorrectly rejected (c) and incorrectly accepted (d) faces for 7 of the 17 persons of interest.

Metric or method	L2	LDML	LDML+MkNN
Precision at equal error	14.0%	38.8%	53.3%

References

[1] K. Weinberger, J. Blitzer and L. Saul. Distance metric learning for large margin nearest neighbor classification. NIPS, 2006
[2] J. V. Davis, B. Kulis, P. Jain, S. Sra and I. S. Dhillon. Information-Theoretic Metric Learning. ICML, 2007.
[3] L. Wolf, T. Hassner and Y. Taigma. Descriptor based methods in the wild. ECCV Workshop, 2008.