Face Recognition from Multiple Images per Subject

Yang Mu
University of Massachusetts
Boston
100 Morrissey Boulevard
Boston, MA, US 02125
yangmu@cs.umb.edu

Henry Z Lo
University of Massachusetts Boston
100 Morrissey Boulevard
Boston, MA, US 02125
henryzlo@cs.umb.edu

Wei Ding[*]
University of Massachusetts Boston
100 Morrissey Boulevard
Boston, MA, US 02125
ding@cs.umb.edu

Dacheng Tao
University of Technology Sydney
235 Jones Street
Ultimo, NSW 2007, Australia
dacheng.tao@uts.edu.au

ABSTRACT

For face recognition, we show that knowing that each subject corresponds to multiple face images can improve classification performance. For domains such as video surveillance, it is easy to deduce which group of images belong to the same subject; in domains such as family album identification, we lose group membership information but there is still a group of images for each subject. We define these two types of problems as *multiple faces per subject*. In this paper, we propose a Bipart framework to take advantage of this group information in the testing set as well as in the training set. From these two sources of information, two models are learned independently and combined to form a unified discriminative distance space. Furthermore, this framework is generalized to allow both subspace learning and distance metric learning methods to take advantage of this group information. Bipart is evaluated on the multiple faces per subject problem using several benchmark datasets, including video and static image data, subjects of various ages, various lighting conditions, and many facial expressions. Comparisons against state-of-the-art distance and subspace learning methods demonstrate much better performance when utilizing group information with the Bipart framework.

Categories and Subject Descriptors

G.4.9 [**Image Precessing and Computer Vision**]: Applications; I.2.6 [**Learning**]: Parameter learning

Keywords

Face recognition, Set classification, Set distance metric learning

1. INTRODUCTION

When recognizing a face present in many images, these images can be thought of as forming a group. Intuitively, using this group information, i.e. knowing what other images have the same face, should improve recognition performance. In this paper, we consider this information in two scenarios.

In the first, what images have the same face is known in the test set. This information is trivially present in the training set. This scenario applies to face recognition in videos, in which object tracking can recognize the same face across frames, but cannot deduce the identity of that face. We call this the *multiple faces with group identity* problem.

In the second scenario, only the existence of face groups is known. For example, in face recognition in a family album, each person corresponds to a group of face images, but we do not know which images constitute this group. We call this *multiple faces with unknown groups*. We show that just knowing that image groups exist can provide valuable information through unsupervised learning.

In both scenarios, we show that by using group information, which is available in both training and test data, facial recognition accuracy can be improved.

Our proposed *Bipart framework* takes advantage of this information. The approach independently learns distance metrics from the training and test sets, then combines them into one distance metric. This framework differs from semi-supervised discriminant analysis (SDA) [2], which does not consider group identity information, and from set classification methods [3, 9], which only work when group identity information is known. The Bipart framework can be applied to both cases [6].

We demonstrate the Bipart framework by using two locally-learned distance metrics. In addition to using our own distance metrics, Bipart can also combine any two projection matrices, including those formed by subspace learning techniques [5, 13], or distance metric learning methods [12, 10]. Furthermore, these matrices can be learned via supervised methods [10] or unsupervised methods [5].

In its supervised form, Bipart can utilize group identity information in the test set to form constraints. In its semi-supervised form, Bipart can deal with situations where this information is not available; instead of utilizing known group identities, these groups are inferred in an unsupervised way.

[*]Corresponding author

MM'14, November 3–7, 2014, Orlando, Florida, USA.
Copyright 2014 ACM 978-1-4503-3063-3/14/11 ...$15.00.
http://dx.doi.org/10.1145/2647868.2655054 .

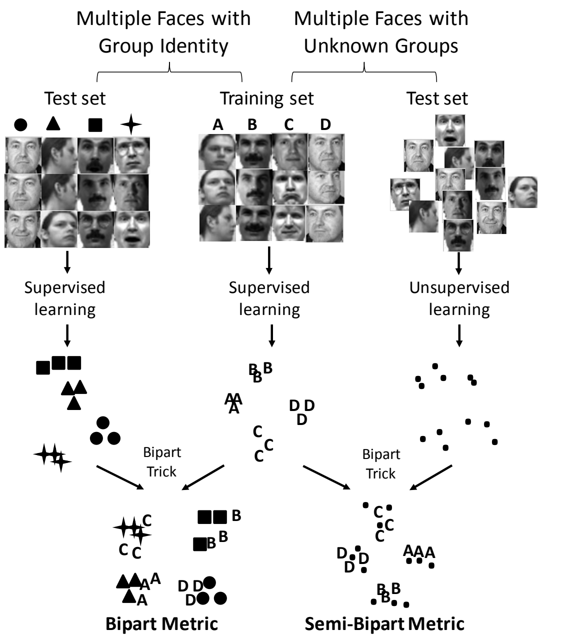

Figure 1: The Bipart framework in its supervised and semi-supervised forms. Both learn two distance metrics from the training and test sets. When the test set has group identity information, its metric can be learned with labeled groups, resulting in supervised Bipart. When this information is not available, the test metric must be learned unsupervised, resulting in semi-supervised Bipart.

It is assumed that group identity information is always available in the training set; see Figure 1 for a visual description.

2. BIPART DISTANCE LEARNING

We propose the Bipart method, which uses two distance metrics: one learned from the training set, and the other from the test set.

2.1 Bipart Trick

With any two images $\mathbf{x}_i$ and $\mathbf{x}_j$ in the data set, the distance metric $\mathbf{A}$ is in the form of

$$d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{A} (\mathbf{x}_i - \mathbf{x}_j)}, \quad (1)$$

where $\mathbf{A}$ is positive semi-definite, and parameterizes a family of Mahalanobis distances. Replacing $\mathbf{A}$ with $\mathbf{W}^T\mathbf{W}$ in Equation (1) using Cholesky decomposition, we get:

$$\begin{aligned} d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) &= \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{W}\mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j)} \\ &= \left\| \mathbf{W}^T (\mathbf{x}_i - \mathbf{x}_j) \right\|. \end{aligned} \quad (2)$$

The Bipart trick fuses two distance metrics by replacing $\mathbf{W}$ with $\mathbf{W}_1\mathbf{W}_2$. Therefore, the Bipart distance metric is:

$$d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = \left\| \mathbf{W}_2^T \mathbf{W}_1^T (\mathbf{x}_i - \mathbf{x}_j) \right\|, \quad (3)$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ correspond to the distance metrics learned from test and training data respectively. Note the similarity between Equations (2) and (3).

Equation (3) is equivalent to projecting the original face images to $\mathbb{R}^{d_1}$ space using $\mathbf{W}_1$, then to $\mathbb{R}^{d_2}$ space using $\mathbf{W}_2$. $\mathbf{W}_2$ plays a more important role, as it preserves the previously projected information. Which set (testing or training) we use to learn which matrix is discussed in Section 3.

2.2 Bipart Distance Metric Objective

To demonstrate how $\mathbf{W}_1$ and $\mathbf{W}_2$ in Equation (3) are learned, we show $\mathbf{W}_1$. $\mathbf{W}_2$ follows the same procedure.

For each image $\mathbf{x}_i$, we minimize the distance from the other images in the block $\mathbf{B}_i^s$ containing $\mathbf{x}_i$, where $\mathbf{B}_i^s$ is the group of face images corresponding to the same subject:

$$\arg\min_{\mathbf{A}_1} \sum_{i=1}^{n_1} \sum_{p=1}^{k_i^s} d^2_{\mathbf{A}_1}(\mathbf{x}_i, \mathbf{x}_p^{\mathbf{B}_i^s}), \quad (4)$$

Likewise, $\mathbf{B}_i^d$, which contains images similar to $\mathbf{x}_i$ but corresponds to different subject with $\mathbf{x}_i$, form the dissimilarity constraints:

$$\arg\max_{\mathbf{A}_1} \sum_{i=1}^{n_1} \sum_{q=1}^{k_i^d} d^2_{\mathbf{A}_1}(\mathbf{x}_i, \mathbf{x}_q^{\mathbf{B}_i^d}), \quad (5)$$

where $A_1 = \mathbf{W}_1^T\mathbf{W}_1$ is the distance metric learned; n_1, k_i^s and k_i^d are the numbers of examples in the test set blocks $\mathbf{B}_i^s$ and $\mathbf{B}_i^d$ respectively; $\mathbf{x}_p^{\mathbf{B}_i^s}$ and $\mathbf{x}_q^{\mathbf{B}_i^d}$ are the p^{th} and q^{th} images in $\mathbf{B}_i^s$ and $\mathbf{B}_i^d$.

Equations (4) and (5) combine to form the objective function:

$$\arg\min_{\mathbf{A}_1} \sum_{i=1}^{n_1} \left(\sum_{p=1}^{k_i^s} d^2_{\mathbf{A}_1}(\mathbf{x}_i, \mathbf{x}_p^{\mathbf{B}_i^s}) - \beta \sum_{q=1}^{k_i^d} d^2_{\mathbf{A}_1}(\mathbf{x}_i, \mathbf{x}_q^{\mathbf{B}_i^d}) \right), \quad (6)$$

where β quantifies the relative importance of the two types of constraints.

The distance metric $\mathbf{A}_1$ as well as $\mathbf{W}_1$ can be solved from Equation (6). Under $\mathbf{A}_1$, the distance between any two examples is equivalent to the Euclidean distance projected using the projection matrix $\mathbf{W}_1$ [7]. $\mathbf{W}_2$ is learned from the training set. With $\mathbf{W}_1$ and $\mathbf{W}_2$, we can obtain the final distance metric $\mathbf{A}$ by Equation (3).

3. BIPART FRAMEWORK

Bipart bridges subspace learning and distance metric learning. In Equation (3), we can use a projection matrix $\mathbf{W}$, as done in subspace learning; conversely, we can use Equation (4), replacing $\mathbf{A}_1$ with any metric $\mathbf{A}$ used in distance learning.

The objectives of linearized subspace learning algorithms can be transformed into the following form:

$$\arg\min_{\mathbf{W}} \operatorname{tr}(\mathbf{W}^T \mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{W}). \quad (7)$$

For example, LPP [5] adds the constraint $\mathbf{W}^T\mathbf{X}\mathbf{X}^T\mathbf{W} = \mathbf{I}_d$, and PCA, LDA, ONPP add the constraint $\mathbf{W}^T\mathbf{W} = \mathbf{I}_d$ on Equation (7). The differences in constructing the matrix $\mathbf{L}$ embody the various motivations of spectral analysis based subspace learning methods.

On the other hand, distance metric approaches can be formulated as follows:

$$\begin{aligned} &\arg\min_{\mathbf{A}} \textstyle\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in S} d^2_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j), \\ &s.t. \textstyle\sum_{(\mathbf{x}_i, \mathbf{x}_j) \in D} d^2_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) \geq \theta \end{aligned} \quad (8)$$

where S contains sample pairs with the same class label, while D contains sample pairs from different classes. Almost all distance metric approaches minimize the similarity constraint as their main objective. The differences lie in optimizing the dissimilarity constraints. For example, Xing's method [12] sets $\theta = 1$, while LMNN[10] defines $\theta = 1 + \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in S} d^2_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j)$ in Equation (8). Our Bipart method simply maximizes θ when optimizing the similarity objective function, which is defined in Equations (4) and (5).

To incorporate and improve subspace learning and distance metric learning methods, Bipart can be used as either a part of a supervised or semi-supervised framework.

Supervised Form: Supervised Bipart can be used to deal with multiple faces with known groups. In this form, group identity in the test set is known and used to train a distance metric. Though $\mathbf{W}_1$ and $\mathbf{W}_2$ can learned from either the test or training set, $\mathbf{W}_1$ should be learned from the set with more information. In our case, we learn $\mathbf{W}_1$ from the test set, since there are more testing images.

After learning $\mathbf{W}_1$ and $\mathbf{W}_2$, we combine the two distance metrics using the Bipart trick. This sequential projection results in a single projection which contains information from both the training and test sets.

A similar method for combining two matrices is LFDA [8], which combines and optimizes both LPP and LDA into one objective function. However, unlike LFDA which only uses training set information, Bipart sequentially combines data from both the training set and test set. In addition, Bipart does not require Laplacian matrices, and thus is applicable to any method which can be formulated as either Equation (7) or (8) [12, 10].

Performance for the Bipart metric $\mathbf{W}_1\mathbf{W}_2$ will be no worse than the performance either $\mathbf{W}_1$ or $\mathbf{W}_2$. When $\mathbf{W}_2$ is full rank, the performance of $\mathbf{W}_1\mathbf{W}_2$ will always be the same as $\mathbf{W}_1$, since there is no dimensionality reduction from $\mathbf{W}_2$. If $\mathbf{W}_2$ does not contribute to the overall performance, it can be set as an identity matrix.

Though we use the distance matrices described in the previous section, Bipart can be generalized to a supervised framework by replacing $\mathbf{W}_1$ and $\mathbf{W}_2$ with any projection matrix learned using supervised subspace learning or distance metric learning methods, as mentioned in Equations (7) and (8).

Semi-Supervised Form: Semi-supervised Bipart (semi-Bipart) can be used for face recognition problems with unknown groups. In this case, $\mathbf{W}_1$ is learned from the training set, and $\mathbf{W}_2$ from the test set. Without known group information, the test set must be learned without supervision, using only the objective function in Equation (4), with $\mathbf{B}^s_i$ being formed via k-nearest-neighbors. Many unsupervised dimensionality reduction methods are ready to be plugged into the semi-Bipart framework, e.g. locally linear embedding (LLE) [7] and many variants of spectral clustering. Unlike the semi-supervised methods discussed in [2] which use a linear tuning parameter to balance the weights between supervised and unsupervised parts, Bipart smoothly combines both parts without needing a tuning parameter.

4. EXPERIMENTS

4.1 Data

We used four face recognition datasets: 11 face images each from 15 individuals in the Yale face database, with various facial expressions; 564 total face images from 20 individuals in the UMIST face database, with various poses; 560 total face images from 80 subjects in the ChokePoint surveillance video dataset [11], with variations in illumination, pose, sharpness, and alignment; 11 images each from 66 individuals in FG-NET dataset [1], where subject age is ranges from newborn to 69. Figure 2 shows some sample images from these datasets.

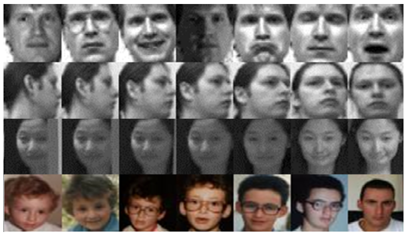

Figure 2: Images from our data sets. From the top row to the bottom: YALE, UMIST, ChokePoint, and FG-NET.

YALE, UMIST, and ChokePoint images were aligned and cropped according to eyes, scaled to 40×40, and normalized to have gray scale values between 0 and 255. In FG-NET, each image contains 68 labeled points describing shape features; Active Appearance Model (AAM) features [4] were extracted from these points. In order to capture 98% of the variance of the whole data, 458 model parameters were extracted according to the AAM to represent each individual.

4.2 Methods

We compared Bipart with the standard subspace learning methods LDA (or Fisherface) and LPP (or Laplacianface) [5]; a semi-supervised method, semi-supervised discriminant analysis (SDA) [2]; a popular set classification method, linear affine hull based image set distance (L-AHISD) [3]; and a state-of-the-art distance learning method, large margin nearest-neighbor (LMNN) [10].

Default parameter settings for all methods were used, and only the dimension of the resulting subspaces were tuned. We set $k^d_i = t + 1$ for Bipart, $k^d_i = 0$ for semi-Bipart, and $k^s_i = t-1$ for both, where t is the number of training samples. β was set to 0.4 for all the experiments. The dimension d_1 was automatically determined by using k-nearest-neighbors on the training set. The final dimension d_2 was chosen to be the best performing value. After projection, nearest neighbor was used to classify faces.

4.3 Experimental Design

In these experiments, one group of face images is provided for each subject. In the training set, subject labels were given for these groups, while in the test set, they were not. The goal was to correctly identify the subject of the group.

This group identity information was used directly in the supervised Bipart framework and L-AHISD; in the other methods, voting was applied between the images in the group to determine the identity of the group.

Unknown group experiments. As in the group identity experiments, each subject was assumed to have multiple images. However, the image to subject groupings were not given for the test set. We used semi-Bipart for this problem, and constructed group information in the test set using k-nearest-neighbors. L-AHISD was not applicable in this

Multiple Faces with Unknown Groups

Dataset	t	**Semi-Bipart**	**LMNN**	**L-AHISD**	**LDA**	**SDA**	**LPP**
YALE	2	$59.78 \pm 4.04(20)$	$55.33 \pm 5.04(119)$	n/a	$54.15 \pm 5.44(14)$	$52.52 \pm 4.06(14)$	$56.67 \pm 3.72(18)$
	4	$76.38 \pm 4.04(30)$	$74.29 \pm 3.08(42)$	n/a	$73.81 \pm 4.07(14)$	$68.48 \pm 4.15(14)$	$74.48 \pm 3.26(21)$
UMIST	2	$83.01 \pm 3.81(12)$	$78.45 \pm 4.42(125)$	n/a	$66.58 \pm 3.02(18)$	$68.86 \pm 3.52(18)$	$66.84 \pm 3.67(26)$
	4	$92.97 \pm 2.55(16)$	$92.10 \pm 2.53(15)$	n/a	$81.37 \pm 4.04(19)$	$84.61 \pm 3.82(19)$	$76.53 \pm 20.53(48)$
ChokePoint	2	$57.77 \pm 2.23(109)$	$62.38 \pm 1.91(35)$	n/a	$54.75 \pm 2.31(74)$	$51.40 \pm 2.36(79)$	$56.17 \pm 2.31(105)$
	4	$77.37 \pm 2.62(53)$	$77.50 \pm 2.14(58)$	n/a	$68.63 \pm 2.99(69)$	$64.00 \pm 3.48(80)$	$69.04 \pm 2.75(144)$
FG-NET	2	$46.57 \pm 2.04(91)$	$41.62 \pm 2.37(304)$	n/a	$26.52 \pm 1.48(63)$	$32.96 \pm 2.12(65)$	$19.36 \pm 1.95(85)$
	4	$64.29 \pm 1.22(186)$	$58.79 \pm 1.60(413)$	n/a	$53.90 \pm 1.47(60)$	$53.03 \pm 1.85(58)$	$21.04 \pm 1.56(79)$

Multiple Faces with Group Identity

Dataset	t	**Bipart**	**LMNN**	**L-AHISD**	**LDA**	**SDA**	**LPP**
YALE	2	$97.33 \pm 4.66(14)$	$82.67 \pm 6.44(119)$	80.00 ± 12.57	$86.67 \pm 8.31(14)$	$80.67 \pm 9.66(14)$	$87.33 \pm 5.84(18)$
	4	$99.33 \pm 2.11(41)$	$94.00 \pm 3.78(42)$	95.33 ± 6.32	$96.00 \pm 4.66(14)$	$92.00 \pm 8.78(14)$	$96.67 \pm 5.67(21)$
UMIST	2	$97.72 \pm 3.40(12)$	$93.94 \pm 5.32(125)$	100.00 ± 0.00	$93.61 \pm 4.92(18)$	$88.43 \pm 7.16(18)$	$93.91 \pm 5.92(26)$
	4	$99.70 \pm 0.96(14)$	$98.48 \pm 2.60(15)$	100.00 ± 0.00	$96.79 \pm 4.34(19)$	$95.41 \pm 3.97(19)$	$92.79 \pm 13.78(48)$
ChokePoint	2	$84.75 \pm 3.48(54)$	$75.38 \pm 4.08(35)$	84.12 ± 3.59	$67.13 \pm 5.68(74)$	$61.88 \pm 5.31(79)$	$71.00 \pm 4.03(105)$
	4	$93.88 \pm 1.61(47)$	$85.00 \pm 2.50(58)$	90.50 ± 2.65	$76.62 \pm 5.47(69)$	$71.50 \pm 4.16(80)$	$77.62 \pm 4.43(144)$
FG-NET	2	$85.30 \pm 3.57(248)$	$73.79 \pm 3.35(304)$	62.88 ± 4.36	$52.73 \pm 6.69(63)$	$56.52 \pm 5.25(65)$	$39.55 \pm 4.76(85)$
	4	$91.97 \pm 2.48(360)$	$90.00 \pm 2.04(413)$	79.85 ± 3.71	$87.12 \pm 3.99(60)$	$85.30 \pm 4.58(58)$	$41.82 \pm 5.16(79)$

Table 1: Averaged accuracy and standard deviations for various methods and datasets over ten-fold cross-validation. t is the number of training samples per subject. Numbers in the parentheses are the dimensions of the final projection space.

setting. SDA and semi-Bipart used both the training and the test set; all other methods only learned from the training set.

We used $t = 2, 4$ samples per person for training; all remaining samples are used for testing. For each t, test and training sets were generated using random splits 10 times.

4.4 Results

Averaged accuracies and standard deviations are shown in table 1. In the unknown group experiments, semi-Bipart performed the best, followed by LMNN, which as shown in Equation 8 has a similar objective function. However, semi-Bipart has stronger constraints on dissimilar faces, and exploits discriminative information in the test set. LDA, SDA and LPP found lower-dimensional subspaces, but were not as accurate.

In the group identity experiments, supervised Bipart and L-AHISD achieve better performance than other methods which only use group identity information for voting. Bipart, by using information in both training and test sets, outperforms L-AHISD on most datasets.

Performance for all methods improved with more training data (higher t in table 1). However, more training data in our experimental design means less testing data. Since Bipart uses test data to augment the learned projection when training data is scarce, the performance gap between Bipart and other methods is expected to shrink with higher t.

5. ACKNOWLEDGEMENTS

The work was partially funded by the University of Massachusetts Boston Proposal Development Grant 2014 and the Australian Research Council Projects FT-130101457 and DP-140102164.

6. REFERENCES

[1] Fg-net database: http://www.fgnet.rsunit.com.

[2] D. Cai, X. He, and J. Han. Semi-supervised discriminant analysis. *ICCV*, pages 1–7, 2007.

[3] H. Cevikalp and B. Triggs. Face recognition based on image sets. In *CVPR*, pages 2567–2573. IEEE, 2010.

[4] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE TPAMI*, 23:681–685, 2001.

[5] X. He and P. Niyogi. Locality preserving projections. In *NIPS*, 2003.

[6] Y. Mu, H. Lo, W. Ding, K. Amaral, and S. Crouter. Bipart: Learning block structure for activity detection. *IEEE Transactions on Knowledge and Data Engineering*, 99:1, 2014.

[7] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[8] M. Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *JMLR*, 8:1027–1061, 2007.

[9] R. Wang, S. Shan, X. Chen, Q. Dai, and W. Gao. Manifold-manifold distance and its application to face recognition with image sets. *IEEE Transactions on Image Processing*, 21(10):4466–4479, 2012.

[10] K. Weinberger and L. Saul. Distance metric learning for large margin nearest neighbor classification. *JMLR*, 10:207–244, June 2009.

[11] Y. Wong and S. C. et al. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In *CVPRW*, 2011.

[12] E. Xing and A. N. et al. Distance metric learning, with application to clustering with side-information. In *NIPS*, pages 505–512, 2002.

[13] T. Zhang, D. Tao, and J. Yang. Discriminative locality alignment. In *ECCV*, pages 725–738, 2008.