SEMI-SUPERVISED FACE RECOGNITION WITH LDA SELF-TRAINING

Xuran Zhao, Nicholas Evans and Jean-Luc Dugelay

Multimedia Communications Department, EURECOM
2229 Route des Crêtes , BP 193, F-06560 Sophia-Antipolis Cedex, France
{zhaox, evans, dugelay}@eurecom.fr

ABSTRACT

Face recognition algorithms based on linear discriminant analysis (LDA) generally give satisfactory performance but tend to require a relatively high number of samples in order to learn reliable projections. In many practical applications of face recognition there is only a small number of labelled face images and in this case LDA-based algorithms generally lead to poor performance. The contributions in this paper relate to a new semi-supervised, self-training LDA-based algorithm which is used to augment a manually labelled training set with new data from an unlabelled, auxiliary set and hence to improve recognition performance. Without the cost of manual labelling such auxiliary data is often easily acquired but is not normally useful for learning. We report face recognition experiments on 3 independent databases which demonstrate a constant improvement of our baseline, supervised LDA system. The performance of our algorithm is also shown to significantly outperform other semi-supervised learning algorithms.

***Index Terms*—** face recognition, LDA, self-training, semi-supervised learning,

1. INTRODUCTION

For more than a decade automatic face recognition (AFR) has been one of the most active research topics in computer vision, machine learning and biometrics. In addition to established applications in access control, surveillance and general security, relatively new applications in digital content structuring, search and retrieval are fast gaining popularity. For example, Google's Picasa[1] application utilizes AFR to label faces detected within a photograph so that queries can be performed to return all the pictures containing a particular person. The extension of such algorithms to the wider Internet has already been reported [1].

Many practical AFR applications are characterized by the weak training of templates or models involving only a small number of labelled training data. In these cases AFR performance is generally not robust to inter-session variation in illumination, occlusion, pose and expression since such variation is not well represented in the template or model. Meanwhile, a large pool of unlabeled auxiliary data is generally easily obtained since its collection does not entail costly manual labelling. Images acquired during testing and general operation may be more representative of inter-session variation and may be used to enhance the template or model via appropriate adaptive or self-training approaches. By iteratively augmenting the training set with more and more images, inter-session variation may be incorporated into the template or model and thus better performance can be expected.

Semi-supervised learning refers to a general class of machine learning techniques that make use of both labelled and unlabelled data for training, typically a small amount of labelled data and a larger amount of unlabelled data [2]. Roli and Marcialis [3] proposed an original semi-supervised face recognition algorithm whereby a PCA-based classifier is initially weakly trained with a small number of manually labelled examples before it is used to classify unlabelled auxiliary data to augment the training set. In related work, also applied to PCA-based classifiers, Roli [4] proposed a variation in which 3 independent classifiers were used. In this work unlabelled auxiliary data are added to augment the labelled dataset only if more than two classifiers agree on the classification result. Neither of the approaches, however, embraces the discriminant power of linear discriminant analysis (LDA). LDA is one of the most popular linear projection techniques for feature extraction, and it is a powerful tool for face recognition when sufficient and representative training examples are available [5]. Over-fitting can occur, however, when the training data is limited and in this case performance can be drastically reduced [6]. To this end, Cai et al. [7] proposed a semi-supervised LDA (SDA) approach which aims to discover the geometrical structure of the data manifold from the unlabeled data but this work did not consider self-training.

In this paper, we propose a new semi-supervised face recognition approach based on LDA and self-training. In contrast to the work in [7], the principal objective is to use automatically labelled, auxiliary data to improve the performance of a classifier that is weakly trained on a small amount of manually labelled data. To our knowledge, it is the first work to couple semi-supervised self-training with an LDA-based approach to face recognition.

[1] http://picasa.google.com/

The remainder of this paper is organized as follows. The new LDA self-training algorithm in described in Section 2. Experiments and results are detailed in Section 3 before our conclusions are presented in Section 4.

2. LDA SELF-TRAINING ALGORITHM

Here we describe our baseline LDA-based AFR system and then a semi-supervised variant based on self-training.

2.1 Baseline System

Linear subspace analysis has been used for AFR over many years and is now a well-known simple, efficient and proven approach. LDA is a supervised algorithm which, according to an optimised projection W_{opt}, projects data vectors x_i in a new space where the ratio between the inter-class (or between, S_B) and intra-class (or within, S_W) scatter is maximized. S_W and S_B are determined according to:

$$S_W = \sum_{j=1}^{c} \sum_{i=1}^{l_j} \left(x_i^j - \mu_j\right)\left(x_i^j - \mu_j\right)^T, \quad (1)$$

and

$$S_B = \sum_{j=1}^{c} l_j \left(\mu_j - \mu\right)\left(\mu_j - \mu\right)^T, \quad (2)$$

where x_i^j is the i^{th} sample of class j, μ_j is the mean of class j, c is the number of classes, and l_j is the number of samples in class j. The global mean, subsuming all classes, is denoted by μ. We define the total scatter according to:

$$S_T = \sum_{i=1}^{l} (x_i - \mu)(x_i - \mu)^T \quad (3)$$

where l is the total number of samples such that $S_T = S_B + S_w$. W_{opt} is obtained according to the objective function:

$$W_{opt} = \arg\max_W \frac{W^T S_B W}{W^T S_T W} = [\boldsymbol{w}_1 \ldots \ \boldsymbol{w}_m] \quad (4)$$

where $\{\boldsymbol{w}_i | i = 1, \ldots, m\}$ are the eigenvectors of S_B and S_T which correspond to the m largest generalized eigenvalues according to:

$$S_B \boldsymbol{w}_i = \lambda_i S_T \boldsymbol{w}_i, \ i = 1, \ldots, m \quad (5)$$

Note that there are at most c - 1 nonzero generalized eigenvalues, so m is upper-bounded by c - 1. Since S_W is often singular it is common to first apply principal component analysis (PCA) to reduce the t-dimensional image vector to a g–dimensional vector, where $t > g > c - 1$, before LDA is used to obtain (c–1)-dimensional vectors.

This is the well-known Fisherface algorithm [5] which generally outperforms the Eigenface approach [8] when sufficient quantities of labelled data are available. When the quantity of data is low S_W in particular can be noisy which leads to unreliable projections and poor performance [7].

2.2 LDA self-training algorithm

A possible solution to deal with insufficient training examples involves semi-supervised learning, which learns from both labelled and unlabelled examples. The semi-supervised PCA-based self-training AFR algorithm proposed in [3] is applied to improve classifiers that are weakly trained using a small labelled dataset $\mathbf{D}_l$. This classifier is then used to automatically label an auxiliary dataset $\mathbf{D}_u$. A fraction of the data with which the system is most confident is then reassigned to $\mathbf{D}_l$ and the classifier is re-trained using the augmented dataset. When repeated iteratively the labelled dataset is steadily enlarged and thus the recogniser is potentially more robust.

However, PCA is an unsupervised approach to dimension reduction. Self-training approaches can thus only help to update the templates for each subject rather than to improve the PCA projection itself. With LDA, in contrast, automatically labelled data not only serve to update templates, but also to increase the amount of data for learning and hence to improve the projection. In this paper, we demonstrate how a standard LDA-based AFR system can be enhanced through the power of self-learning.

The algorithm is summarized in Table 1. The input to the system is a labelled dataset $\mathbf{D}_l$ and a larger unlabelled auxiliary dataset $\mathbf{D}_u$. First a supervised Fisherface algorithm is applied to reduce the t–dimensional image vectors to a g-dimensional vector through PCA and then to a (c-1)-dimensional vector through LDA. A template is calculated for each class by calculating the projected mean. The set of unlabelled samples $\mathbf{D}_u$ is then automatically assigned the label of its nearest template, using the Euclidean distance. Then, for each class, the single example which is nearest to the corresponding template is removed from $\mathbf{D}_u$ and added to the labelled set $\mathbf{D}_l$. If, for any given class, there are no corresponding examples in $\mathbf{D}_u$ then the corresponding labelled set in $\mathbf{D}_l$ is left unchanged. The PCA and LDA projections are relearned and the templates are recalculated. The process is repeated iteratively until $\mathbf{D}_u$ is empty. A less conservative strategy can also be used whereby, upon each iteration, more than one automatically labelled example is added to the training data for each class. This results in a faster algorithm but one which does not capitalise on all the additional training data when each individual sample is selected. Improved computational efficiency thus comes at the cost of reduced performance. The algorithm can work both in a *transductive* or *semi-supervised* configuration. A *transductive* configuration refers to the situation where both the training and testing set are available in the learning process, which reflects an application similar to the automatic labelling of photos in a digital album; A *semi-supervised* configuration refers to the situation where the testing set is not available during the learning process, and reflects a video security application, for example.

Finally we note that, to avoid S_B and S_T being identical, the LDA algorithm needs at least 2 initial training examples per class. When only a single labelled image is available this restriction can be easily overcome by acquiring a second image through Eigenface recognition, so that LDA may then be applied in the normal way.

3. EXPERIMENTAL RESULTS

In this section we report experiments that aim to assess the LDA self-training algorithm and to compare its performance to that of other semi-supervised learning methods. Our

- Given:
  - D_l, a set of labelled examples from *c* classes;
  - D_u, a set of unlabelled examples;
  - *g*, PCA inter-media space dimension in Fisherface algorithm;
- Apply Supervised Fisherface initialization using D_l:
  - Use PCA to project D_l into a *g*-dimension inter-space, then use LDA to further project into *c*-1 dimension feature space;
  - A template is created by calculating the projected mean of each class;
- Apply iterative of self-training:
  - Project the D_u into the PCA inter-space then into the LDA feature space;
  - Label each example in D_u according to the nearest template;
  - For each class, the *n* examples closest to the template are removed from D_u and added to D_l;
  - Relearn the PCA and LDA projections with new D_l, and update the recalculate the templates;
  - *Iterate until D_u is empty;*

Table 1: LDA self-training algorithm

experiments were performed on three standard, independent datasets: the Olivetti Research Lab (ORL) database [9], the AR database [10] and the CMU PIE database [11].

Experiments with the ORL database were performed with a transductive configuration while those with the AR database were performed in a *semi-supervised* configuration. The aim is to show that our method is beneficial in both cases. Experiments conducted with the CMU PIE database relate to single training images. Here we aim to show the benefit of our algorithm over that reported in [7] which was assessed on the same database.

The PCA inter-space dimension *g* was seen to have a strong influence on performance but behaviour was observed to be consistent across the three different datasets. For all experiments reported in this paper *g* was set equal to 1.5 times the number of classes (persons).

3.1 Transductive configuration

The ORL database contains images from 40 subjects with 10 images per subject, including pose and expression variations. Original images contain 92×112 pixels but, for computational efficiency, all images were down sampled to 23×28 pixels. Results reported below indicate that our algorithm works well with such low-resolution images.

For any single trial, a template is derived for each subject using between i = 1 to 5 labelled training images which are randomly selected according to the ground-truth reference. The remaining images are used either as unlabelled examples for self-training or as test data. Fig. 2 shows the average recognition rate observed from 20 trials. The horizontal axis represents the number of self-training iterations, while the vertical axis is the recognition rate. Profiles are illustrated for each value of *i* and confirm that recognition accuracy increases when a greater number of images is used for training (55% for *i*=1 cf. 96% for *i*=5, without self-training). All profiles are further shown to rise as more training images are acquired through self-training (55% without self-training cf. 82% with 9 iterations, for *i*=1). Note that the maximum number of self-training iterations decreases with increasing *i* since there are then fewer unlabelled images available.

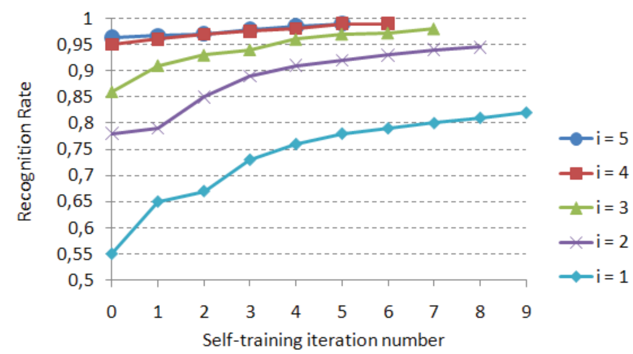

Figure 2. Recognition rate as a function of the number of self-training iterations

Fig. 3 illustrates comparative results for alternative semi-supervised AFR algorithms, namely *PCA-self training* [3], semi-supervised LDA (SDA) [7] in addition to profiles for *supervised Eigenface* [8] *and Fisherface* [5] algorithms. All systems are our own implementations except for the SDA algorithm which comes from the source code provided by the authors of [7]. In all cases results are averaged over 20 trials. Results show that LDA self-training outperforms all other algorithms by a significant margin and serve to demonstrate the merit in combining a discriminant classifier with self-training.

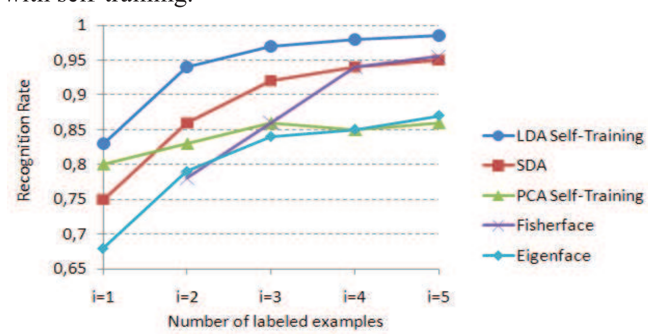

Figure 3. Recognition accuracies on ORL database

3.2 Semi-supervised configuration

The second set of experiments aim to evaluate the self-training LDA algorithm in a semi-supervised configuration, where test data is not available during the learning process. Here experiments where performed on the AR database which contains over 4,000 face images from 126 people, and includes expression, illumination and occlusion variations. We first purged the dataset of occluded images and randomly selected 100 subjects (50 male, 50 female). The resulting subset contained 14 images per subject. All images were manually cropped to focus on the face and resized to 32×32 pixels. 3 images per subject were randomly selected as test images. For any one trial i = 1 to 5 images were labelled according to the ground-truth reference and used for template learning. The others are used as unlabelled images for self-training. Results for the five different algorithms are illustrated in Fig. 4 and again show that the self-training algorithm outperforms the other algorithms.

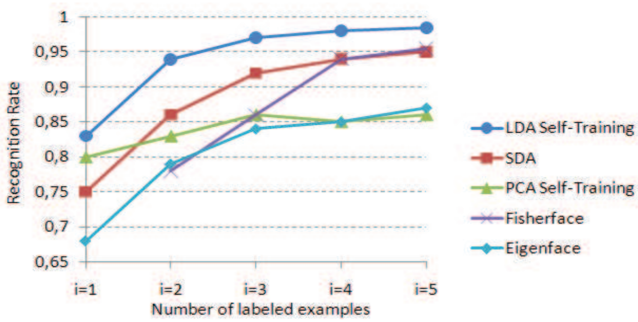

Figure 4. Recognition accuracies on AR database

3.3 Single training image test

The CMU PIE face database contains 68 subjects with 41,368 face images captured with varying pose, illumination and expressions. Each image contains 32×32 pixels. For all experiments reported here we used only frontal pose images which correspond to 43 per subject from which 30 were randomly selected as training data. For any single trial, a single training image is randomly selected for each subject and the remaining 29 images are left unlabelled and are pooled for subsequent self-training. As before results are averaged over 20 trials. From the results illustrated in Table 2 we can see that although the LDA self-training algorithm exhibits larger standard deviation among different trials, it nevertheless achieves the best performance among all the other algorithms, with a significant margin.

	Unlabeled Set	*Test Set*
Eigenface [8]	25.3±1.7	25.3±1.6
Laplacianface [12]	56.1±2.3	56.4±2.4
Consistency [10]	52.0±1.8	--
LapSVM [14]	56.5±1.6	56.9±2.6
LapRLS [14]	57.5±1.6	57.9±2.6
SDA [7]	59.0±2.0	59.5±2.7
LDA self-training	**84.5±9.5**	**71.3±6.5**

Table 2. Recognition rate on CMU PIE database (mean ± std-dev %)

4. CONCLUSIONS

This paper presents a new semi-supervised face recognition algorithm based on LDA self-training. Despite its simplicity it successfully exploits both labelled and un-labelled data for template learning and delivers superior performance than existing approaches. Training data is augmented with automatically labelled, auxiliary data that is often easily obtained without the cost of manual labelling. Experiments on three independent datasets show that the new algorithm is robust to variations in illumination, pose and expression and that it outperforms related approaches in both transductive and semi-supervised configurations. These observations indicate that the new self-training algorithm is successful in overcoming the over-fitting problems which typify LDA-based approaches to automatic face recognition and that they warrant further attention.

ACKNOWLEDGEMENT

This work was partially supported by the TABULA RASA project funded under the 7th Framework Programme of the European Union (EU), grant agreement number 257289.

REFERENCES

[1] N. Kumar, P. N. Belhumeur, and S. K. Nayar. "Face-Tracer: A Search Engine for Large Collections of Images with Faces" in Proc. ECCV 2008, pp. 340--353, October 2008.

[2] X. Zhu, "Semi-supervised learning literature survey", Technical report, Univ. Wisconsin, Madison, USA, Jan. 2006.

[3] F. Roli, G.L. Marcialis, "Semi-supervised PCA-based face recognition using self training," in Proc. Joint IAPR Int. Work. on Structural and Syntactical Pattern Recognition and Statistical Techniques in Pattern Recognition S+SSPR06, Springer LNCS 4109: 560-568, 2006.

[4] F. Roli. "Semi-supervised multiple classifier systems: Background and research directions," In Proc. of the 6th International Workshop on Multiple Classifier Systems, pp. 1–11, 2005.

[5] P. N. Belhumeur, J. P. Hepanha, and D. J. Kriegman. "Eigen-faces vs. fisherfaces: recognition using class specific linear projection," IEEE Transactions on PAMI, 19(7):711–720, 1997.

[6] A.M. Martinez and A.C. Kak, "PCA versus LDA", in IEEE Transaction on PAMI, vol. 23(2), PP. 228-233, Feb. 2001.

[7] D. Cai, X. He, and J. Han. "Semi-supervised discriminant analysis," In Proc. ICCV, 2007

[8] M. Turk and A. Pentland. "Eigenfaces for recognition", Journal of Cognitive Neuroscience, 3(1):71–86, 1991.

[9] http://www.cam-orl.co.uk

[10] A.M. Martinez and R. Benavente, "The AR Face Data-base," CVC Technical Report, no. 24, June 1998.

[11] T. Sim, S. Baker, and M.Bsat. "The CMU pose, illumination, and expression database", IEEE Transactions on PAMI, vol. 25(12), pp. 1615-1618, 2003

[12] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang. "Face recognition using laplacianfaces", IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(3):328–340, 2005.

[13] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Sch¨olkopf. "Learning with local and global consistency", In Advances in Neural Information Processing Systems 16, 2003.

[14] M. Belkin, P. Niyogi, and V. Sindhwani. "Manifold regu-larization: A geometric framework for learning from examples", Journal of Machine Learning Research, 2006.