DISCRIMINATIVE HESSIAN EIGENMAPS FOR FACE RECOGNITION

Si Si[1], *Dacheng Tao*[2], *Kwok-Ping Chan*[1]

[1]Department of Computer Science, University of Hong Kong, Hong Kong
[2]School of Computer Engineering, Nanyang Technological University, Singapore

ABSTRACT

Dimension reduction algorithms have attracted a lot of attentions in face recognition because they can select a subset of effective and efficient discriminative features in the face images. Most of dimension reduction algorithms can not well model both the intra-class geometry and inter-class discrimination simultaneously. In this paper, we introduce the Discriminative Hessian Eigenmaps (DHE), a novel dimension reduction algorithm to address this problem. DHE will consider encoding the geometric and discriminative information in a local patch by improved Hessian Eigenmaps and margin maximization respectively. Empirical studies on public face database thoroughly demonstrate that DHE is superior to popular algorithms for dimension reduction, e.g., FLDA, LPP, MFA and DLA.

Index Terms—*Dimension Reduction, Manifold Learning, Face Recognition*.

1. INTRODUCTION

Dimension reduction [3],[11] plays an important role in various tasks in computer vision, e.g., face recognition. A key role for face recognition is the distance or similarity between face images which can be solved via dimension reduction, as dimension reduction performs the recognition by enlarging the similarity among the intra-class samples and maximizing the difference among the inter-class samples in a subspace rather than the original feature space.

A dimension reduction algorithm projects the original high-dimensional feature space to a low-dimensional subspace, where specific statistical properties can be well preserved. For example, principle component analysis (PCA) [1], one of the most popular unsupervised dimension reduction algorithms, maximizes the variance of the data in the projected subspace; Fisher's linear discriminative analysis (FLDA) [2], the most traditional supervised dimension reduction algorithm, minimizes the trace ratio between the within class scatter and the between class scatter so that the Gaussian distributed samples can be well separated in the selected subspace; locality preserving projections (LPP) [4] preserves the local geometry of samples by processing an undirected weighted graph that represents the neighbourhood relations of pairwise samples; Marginal Fisher analysis (MFA) [12] considers both the intra-class geometry and interaction of samples from different classes; Discriminative locality alignment (DLA) [5] preserves the discriminative information by maximizing the distance among the inter-class samples and minimizing the distance among the intra-class samples over the local patch of each sample. However the geometric and discriminative information in these dimension reduction algorithms are not well modeled, e.g., LDA does not consider the geometric information; MFA ignores the discriminative information of non-marginal samples from different classes.

By using the patch alignment framework [6], we can model both the intra-class local geometry and the inter-class discriminative information conveniently. In particular, for each sample and its associated patch (neighbours of the sample), it is important to consider the following two properties: 1) the intra-class local geometry can be represented by the local tangent space, which is locally isometric to the manifold of the intra-class nearest samples of the patch; and 2) the inter-class discriminative information can be represented by the margin between the intra-class neighbor samples and the inter-class nearest samples of the patch. Because the method used for local geometry representation is similar to Hessian Eigenmaps [7], the proposed dimension reduction algorithm is termed the Discriminative Hessian Eigenmaps or DHE for short.

The rest of this paper is organized as follows. Section 2 introduces the proposed Discriminative Hessian Eigenmaps (DHE). Section 3 shows the results of thoroughly empirical studies. Section 4 concludes.

2. DISCRIMINATIVE HESSIAN EIGENMAPS

This Section presents the discriminative Hessian Eigenmaps or DHE for short to solve the face recognition tasks. In DHE, we try to find an optimal linear mapping $W \in R^{D\times d}$ so that it can project $\vec{x}_i \in R^D$ to a low-dimensional space as $\vec{y}_i = W^T \vec{x}_i \in R^d$. In this learned low-dimensional space, DHE characterizes two specific properties:

978-1-4244-4296-6/10/$25.00 ©2010 IEEE

1. The local geometry property - nearby samples in the original Euclidean space are close to each other in the learned subspace.
2. The discriminative property - samples from different classes can be well separated in the learned subspace.

In summary, the discriminative information as well as the local geometry will be well modeled in the DHE.

2.1. Modified Hessian Eigenmaps

Empirically, intra-class geometry is useful for classification. Hessian Eigenmaps [7] is a geometry preservation manifold learning method that can recover the underlying parameterization of a manifold M embedded in a high-dimensional space if the manifold M is locally isometric to an open and connected subset of R^d. Because the parameter space need not be convex in Hessian Eigenmaps, it can be applied to model a nonconvex manifold, *e.g.*, an S-curve surface with a hole. Therefore, we adapt Hessian Eigenmaps in DHE to preserve the local geometry for dimension reduction.

Hessian Eignmaps finds the (*d+1*)-dimensional null-space of $H(f)$, where $H(f)$ is the Hessian matrix of a smooth mapping f, *i.e.*, $f: M \mapsto R$. This $H(f)$ can be calculated by using $H(f) = \int_M \left\| H_f(x_i) \right\|_F^2 dx$ wherein $H_f(x_i)$ is the Hessian of f on the patch $X_{H(i)} = \left[x_i, x_{i^1}, \cdots, x_{i^{k_1}} \right]$ and the corresponding output in low-dimensional space is $Y_{H(i)} = \left[y_i, y_{i^1}, \cdots, y_{i^{k_1}} \right]$. The tangent plane $T_{x_i}(M)$, a Euclidean space tangential to M at x_i, is an orthogonal coordinate system. In order to estimate $H_f(x_i)$, we calculate the local coordinate system of $X_{H(i)}$ and each sample in $X_{H(i)}$ has its own local coordinate Π_i on the tangent plane $T_{x_i}(M)$. Afterwards, this $H_f(x_i)$ can be estimated by using Π_i.

However, Hessian Eigenmaps cannot be applied to many practical applications, *e.g.*, face recognition because it requires that $k_1 > d$ where k_1 is the number of the neighbouring samples and d is the dimension of the subspace. It is difficult to guarantee this condition because we have a limited number of samples. We propose to overcome this problem by performing PCA on M at x_i and orthnormalizing the *d*-dimensional representation to obtain the tangent coordinate in $T_{x_i}(M)$. The following steps for the modified Hessian Eigenmaps are similar to those in Hessian Eigenmaps.

Under the patch alignment framework, the objective function for the modified Hessian Eigenmaps to preserve the local geometry on a local patch $Y_{H(i)}$ can be written as

$$H(y_i) = \mathrm{tr}\left(Y_{H(i)} H_f(x_i) H_f^T(x_i) Y_{H(i)}^T \right) = \mathrm{tr}\left(Y_{H(i)} L_{H(i)} Y_{H(i)}^T \right), \quad (1)$$

where $L_{H(i)} = H_f(x_i) H_f^T(x_i)$ encodes the local geometry information of the patch $X_{H(i)}$ and $H(y_i)$ is the local geometry representation. Under the help of $L_{H(i)}$, local geometric information can be further preserved.

2.2. Margin Maximization

As for classification, however, it is insufficient to only retain the local geometry, because no labeling information is taken into account. To further exploit the discriminative power, like the definition of the local geometry, we can define a new margin maximization [13] based scheme for discriminative information preservation over patches. In particular, for each sample x_i associated with a patch $X_{M(i)} = \left[x_i, x_{i^1}, \cdots, x_{i^{k_1}}, x_{i_1}, \cdots, x_{i_{k_2}} \right]$, wherein $x_{i^1}, \ldots, x_{i^{k_1}}$, *i.e.*, the k_1 nearest samples of x_i, are from the same class as x_i, and $x_{i_1}, \ldots, x_{i_{k_2}}$, *i.e.*, the other k_2 nearest samples of x_i, are from different classes against x_i, we define the margin as the average difference between two kinds of distances on this patch. One is called inter-class distance, that is, the distance between x_i and samples taking different labels, *i.e.*, $x_{i_1}, \ldots, x_{i_{k_2}}$; the other is called intra-class distance, that is, the distance between x_i and samples sharing the same label, *i.e.*, $x_{i^1}, \ldots, x_{i^{k_1}}$. Basically, in the patch $X_{M(i)}$'s low-dimensional representation $Y_{M(i)} = \left[y_i, y_{i^1}, \cdots, y_{i^{k_1}}, y_{i_1}, \cdots, y_{i_{k_2}} \right]$, we expect the margin between intra-class and inter-class samples will be maximized as large as possible, *i.e.*,

$$\sum_{p=1}^{k_2} \left\| y_i - y_{i_p} \right\|^2 \frac{1}{k_2} - \sum_{j=1}^{k_1} \left\| y_i - y_{i^j} \right\|^2 \frac{1}{k_1}. \quad (2)$$

On the other hand, based on (2), we try to minimize the bellowing objective function:

$$\begin{aligned} M(y_i) &= \sum_{j=1}^{k_1} \left\| y_i - y_{i^j} \right\|^2 \frac{1}{k_1} - \sum_{p=1}^{k_2} \left\| y_i - y_{i_p} \right\|^2 \frac{1}{k_2} \\ &= \mathrm{tr}\left(Y_{M(i)} \begin{bmatrix} -e_{k_1+k_2}^{\ T} \\ I_{k_1+k_2} \end{bmatrix} \mathrm{diag}(w_i) \left[-e_{k_1+k_2}, I_{k_1+k_2} \right] Y_{M(i)}^T \right) \quad (3) \\ &= \mathrm{tr}\left(Y_{M(i)} L_{M(i)} Y_{M(i)}^T \right), \end{aligned}$$

where $w_i = \left[\overbrace{1/k_1, \ldots, 1/k_1}^{k_1}, \overbrace{-1/k_2, \ldots, -1/k_2}^{k_2} \right]^T$; $I_{k_1+k_2}$ is the

$(k_1+k_2)\times(k_1+k_2)$ identity matrix ; $e_{k_1+k_2}=[1,...,1]^T\in R^{k_1+k_2}$; $L_{M(i)}=\begin{bmatrix}\sum_{j=1}^{k_1+k_2}(w_i)_j & -w_i^T\\ -w_i & \mathrm{diag}(w_i)\end{bmatrix}$ and $M(y_i)$ is the margin information representation.

2.3. Discriminative Hessian Eigenmaps (DHE)

By using the results obtained from the previous subsections, we can obtain the optimization framework to learn the projection matrix W, which can utilize both the local geometry and the discriminative information. Because the margin representation $M(y_i)$ and the local geometry representation $H(y_i)$ are defined over patches, and each patch has its own coordinate system, alignment strategy is adopted here to build a global coordinate for all patches defined for the training samples. As a consequence, the objective function for DHE to solve the dimension reduction problem is given by

$$W=\underset{W\in R^{D\times d}}{\arg\min}\sum_{i=1}^{l}\left(M(y_i)+\beta H(y_i)\right), \tag{4}$$

where β is the tuning parameter. If we define two selection matrixes $S_{H(i)}$ and $S_{M(i)}$, which select samples in the i^{th} patch from all the training samples $Y_L=[y_1,y_2,\cdots,y_l]$ for constructing $M(y_i)$ and $H(y_i)$, respectively. Therefore, $Y_{H(i)}=Y_LS_{H(i)}$ and $Y_{M(i)}=Y_LS_{M(i)}$ with $Y_{H(i)}$ representing the patch for the local geometry preservation and $Y_{M(i)}$ denoting the patch for margin maximization. After plugging (1) and (3), the objective function in (4) will turn to

$$\begin{aligned}W&=\underset{W\in R^{D\times d}}{\arg\min}\sum_{i=1}^{l}\begin{pmatrix}\mathrm{tr}\left(Y_{M(i)}L_{M(i)}Y_{M(i)}^T\right)\\+\beta\,\mathrm{tr}\left(Y_{H(i)}L_{H(i)}Y_{H(i)}^T\right)\end{pmatrix}\\
&=\underset{W\in R^{D\times d}}{\arg\min}\sum_{i=1}^{l}\begin{pmatrix}\mathrm{tr}\left(Y_LS_{M(i)}L_{M(i)}\left(Y_LS_{M(i)}\right)^T\right)\\+\beta\,\mathrm{tr}\left(Y_LS_{H(i)}L_{H(i)}\left(Y_LS_{H(i)}\right)^T\right)\end{pmatrix}\\
&=\underset{W\in R^{D\times d}}{\arg\min}\,\mathrm{tr}\left(Y_L\sum_{i=1}^{l}\begin{pmatrix}S_{M(i)}L_{M(i)}S_{M(i)}^T\\+\beta S_{H(i)}L_{H(i)}S_{H(i)}^T\end{pmatrix}Y_L^T\right)\\
&=\underset{W\in R^{D\times d}}{\arg\min}\,\mathrm{tr}\left(Y_LLY_L^T\right),\end{aligned} \tag{5}$$

where $L=\sum_{i=1}^{l}\left(S_{M(i)}L_{M(i)}S_{M(i)}^T+\beta S_{H(i)}L_{H(i)}S_{H(i)}^T\right)$ is the alignment matrix encoding both the local geometry and the discriminative information.

For linearization, $Y_L=W^TX_L$ is usually considered, where W is the projection matrix. We can impose different constraints, *e.g.*, $Y_L^TY=I$ or $W^TW=I$, to uniquely determine Y_L. The constraint $W^TW=I$ will be adopted throughout the paper. Under this constraint and $Y_L=W^TX_L$, the solution of (5) can be obtained by using the conventional Lagrangian multiplier method [10] or the generalized eigenvalue decomposition [8].

3. EXPERIMENTS

In this Section, we justify our proposed DHE algorithm with four representative dimension reduction algorithms, which are the Fisher's linear discriminant analysis (FLDA) [2], the locality preservation projections (LPP) [4] with the supervised setting, the marginal Fisher's analysis (MFA) [12] and discriminative locality alignment (DLA) [6] for face recognition based on a public database: CMU-PIE dataset [9].

Figure 1. Sample images from CMU-PIE database

The CMU-PIE dataset contains 41,368 images of 68 people under 13 different poses, 43 different illumination conditions, and 4 different expressions, and we randomly select 10 images per individual in the CMU-PIE dataset in this experiment. Example face images from the CMU-PIE database are shown in Figure 1. The images from CMU-PIE used for our experiments are of size 32x32 in raw pixel.

In the training stage, we learn the projection matrix W from each involved algorithm on the training samples. In the testing stage, each testing sample will be projected into the low-dimensional space by W and after that nearest-neighbor rule (NN) is applied to predict label of the test image in the selected subspace.

We randomly select p (= 4, 5, 6) images per individual for training in the database, and use the remaining images for testing. All trials are repeated ten times, and then the average recognition rates are calculated. Figure 2 shows the results of DHE against FLDA, LPP, MFA and DLA with regard to face recognition accuracy under different dimensions. Table 1 provides the best recognition rate for each algorithm. It also provides the optimal values of k_1, k_2, and β for DHE which are tuned by the cross validation.

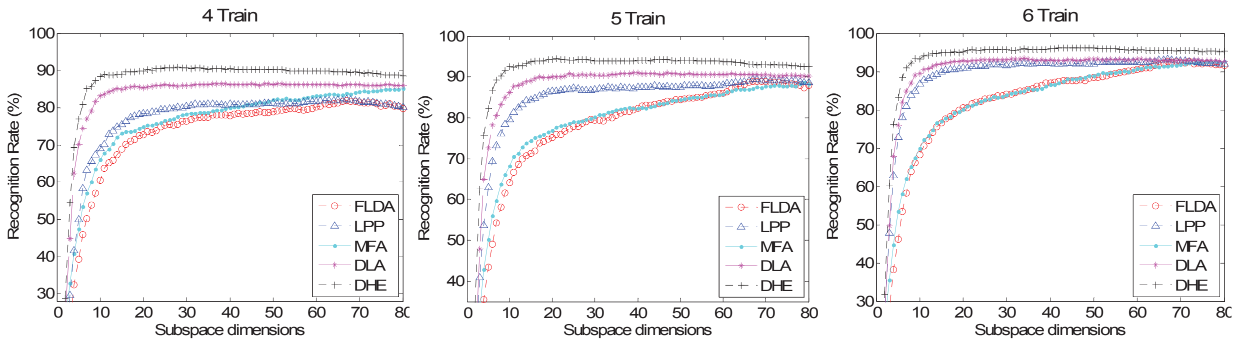

Figure 2. Recognition rate vs. dimension reduction on the CMU-PIE database under different splits.

Table 1. Best recognition rates (%) on CMU-PIE database. The numbers in the parentheses are the subspace dimensions. For DHE, The numbers in the parentheses from left to right are the subspace dimensions, k_1, k_2 and β respectively.

	FLDA	LPP	MFA	DLA	DHE
4 Train	81.79(67)	82.33(68)	88.58(78)	86.50(39)	**90.86**(29,3,6,1)
5 Train	88.94(67)	89.38(67)	90.29(80)	90.94(62)	**94.44**(47,3,6,5)
6 Train	92.58(69)	93.17(67)	92.67(80)	93.46(34)	**96.18**(47,3,6,10)

As shown in Figure 2 and Table 1, DHE outperforms conventional algorithms or at least can obtain a comparable performance in comparing with the conventional algorithms, because DHE can precisely model both the intra-class geometry and the inter-class discriminative information in the local patch.

4. CONCLUSION

In this paper, we have proposed a novel linear dimension reduction algorithm, termed Discriminative Hessian Eigenmaps (DHE). DHE is superior to the conventional dimensionality reduction algorithms because it focuses on accurately modeling both the intra-class geometry and inter-class discrimination in the local patch. Empirical studies on face recognition tasks demonstrate that DHE is more effective than conventional algorithms.

5. ACKNOWLEDGEMENT

S. Si and K.-P. Chan thank the support from HKU-SPF Grant (project number 10400016). D. Tao thanks support from the Nanyang SUG Grant (project number M58020010) and Microsoft Operations PTE LTD-NTU Joint R&D (project number M48020065).

6. REFERENCES

[1] H. Hotelling, “Analysis of a Complex of Statistical Variables into Principal Components,” *Journal of Educational Psychology*, vol. 24, pp. 417-441, 1933.

[2] R. A. Fisher, “The Use of Multiple Measurements in Taxonomic Problems,” *Annals of Eugenics*, vol. 7, pp. 179-188, 1936.

[3] D. Tao et al., “General Tensor Discriminant Analysis and Gabor Features for Gait Recognition,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, 2007.

[4] X. He and P. Niyogi, “Locality Preserving Projections,” *NIPS*, vol. 16, Vancouver, Canada, 2003.

[5] T. Zhang et al., “Discriminative Locality Alignment,” *Proc. European Conf. Computer Vision*, vol. 1, pp. 725–738, 2008.

[6] T. Zhang et al., “Patch alignment for dimensionality reduction,” *IEEE Trans. Knowledge and Data Engineering*, 2009.

[7] D. L. Donoho and C. Grimes, “Hessian Eigenmaps: new locally linear embedding techniques for high-dimensional data,” *Proc. National Academy of Arts and Sciences*, vol. 100, pp. 5591–5596, 2003.

[8] T. Zhang et al., “A unifying framework for spectral analysis based dimensionality reduction,” *Int’l Joint Conf. Neural Networks*, pp. 1670–1677, 2008.

[9] T. Sim et al., “The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces,” Technical Report CMU-RI-TR-01-02, Carnegie Mellon University, 2001

[10] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier (Optimization and Neural Computation Series)*, Athena Scientific, 1996.

[11] D. Tao et al., “Geometric Mean for Subspace Selection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 260–274, 2009.

[12] D. Xu et al., “Marginal Fisher Analysis and Its Variants for Human Gait Recognition and Content Based Image Retrieval,” *IEEE Trans. Image Processing*, vol. 16, no. 11, pp. 2811-2821, 2007.

[13] W. Liu et al., “Transductive Component Analysis,” *Proc. IEEE Int’l Conf. Data Mining*, pp. 433–442, 2008.