

Kernel Linear Regression for Low Resolution Face Recognition under Variable Illumination

Shih-Ming Huang, Jar-Ferr Yang

Institute of Computer and Communication Engineering,
Department of Electrical Engineering,
National Cheng Kung University, Tainan, Taiwan
E-mail: smhuang@video5.ee.ncku.edu.tw, jfyang@ee.ncku.edu.tw

ABSTRACT

To improve the limitation of linear regression classification, a class specific kernel linear regression classification is proposed for low resolution face recognition under variable illumination. The nonlinear mapping function enhances the modeling capability for highly nonlinear data distribution. The explicit knowledge of the nonlinear mapping function can be avoided computationally by using the kernel trick. With kernel projection, the class label is also determined by calculating the minimum reconstruction error. Experiments carried out on Yale B facial database in size of 8×8 pixels reveal that the proposed algorithm outperforms the state-of-the-art methods and demonstrates promising abilities against severe illumination variation.

Index Terms—Kernel Linear Regression, Low Resolution Face Recognition, Illumination Variation

1. INTRODUCTION

Automatic face recognition systems [1] are designed to distinguish a specific identity from the unknown objects characterized by face images. Numerous studies have been greatly interested in automatically recognizing faces from still or video images. In realistic situations, face recognition may encounter many great challenges, especially the low resolution problem, which could be caused by environments and capture devices at a distance. Additionally, the low resolution problem might be coupled with other effects such as illumination variation. Our method accounts for low resolution face recognition under illumination variation.

In the literature, numerous researches have been proposed to achieve successful face recognition. These approaches could be categorized into two categories namely *reconstructive* and *discriminative* methods. The reconstructive approaches such as principle component analysis (PCA) [2-4] and independent component analysis (ICA) [5-7] have been reported to be robust for the problem related to noisy pixels. The discriminative approaches such as linear discriminant analysis (LDA) [3], [4] have been known to yield better results in clean conditions. Moreover,

many variants such as kernel PCA (KPCA) [8-10] and kernel LDA (KLDA) [10-12] have been presented to achieve higher performance. These kernel methods enhance the modeling capability by nonlinearly mapping the data from the original space to a very high dimensional feature space, the so-called reproducing kernel Hilbert space (RKHS). Therefore, the KPCA and the KLDA by nonlinearly mapping could utilize high-order statistics, whereas the PCA and LDA only utilize the first and second-order statistics. Thus, for highly nonlinear data distribution, these kernel methods are more suitable.

Recently, a linear regression classification (LRC) algorithm [13] has been proposed for face recognition, which is based on that face images from a specific class are known to lie on a linear subspace [3], [14]. The regression coefficients are estimated by using the least square method, and then the decision is made in favor of the class with the minimum reconstruction error. Experiments reported have shown that the down-sampled image could be used for classification directly. However, as the results reported, the LRC could not withstand severe illumination variations.

To conquer the difficulty of illumination variations in face recognition, a variety of approaches have been proposed to solve the problem [3, 14, 15-20]. For instance, histogram equalization (HE), Gamma correction, logarithm transform, etc. are widely used for illumination normalization. Edge maps, derivatives of the gray-level, Gabor-like filters, and the LDA are well-known illumination invariant feature extraction methods. As to face modeling, the illumination cone method and the spherical harmonic model have been proposed. To the best of our knowledge, these methods could not work for low resolution problem well because of losing the important detail information under low resolution.

In this paper, we propose a novel face recognition algorithm to overcome the limitation of the LRC [13] by applying the kernel method for linear regression. Our method can be treated as an illumination invariant feature extraction method because the proposed approach can

achieve variable lighting face recognition without any preprocessing of illumination normalization and compensation.

The rest of this paper is organized as follows. Section 2 reviews the LRC approaches. Section 3 formulates the proposed kernel linear regression classification (KLRC). Section 4 gives experimental results. Finally, we draw conclusions in Section 5.

2. LINEAR REGRESSION CLASSIFICATION (LRC)

Assume we have N subjects with p_i training images from the i^{th} class, $i = 1, 2, \dots, N$. Each gray scale training image is in size of $a \times b$ and is represented as $\mathbf{v}_{i,j} \in \mathbb{R}^{a \times b}$, $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, p_i$. Each training image is down-sampled to an order of $c \times d$ and $\mathbf{v}_{i,m} \in \mathbb{R}^{a \times b}$ is transformed to column vector $\mathbf{w}_{i,j} \in \mathbb{R}^{q \times 1}$, where $q = cd$, and $ab > cd$. Each column vector is normalized so that maximum pixel value is 1. In order to apply linear regression to estimate class specific model, we have to stack the column vectors $\mathbf{w}_{i,j}$ regarding the class-membership. Hence, for the i^{th} class, we have

$$\mathbf{W}_i = [\mathbf{w}_{i,1}, \dots, \mathbf{w}_{i,j}, \dots, \mathbf{w}_{i,p_i}] \in \mathbb{R}^{q \times p_i}, \quad (1)$$

where each vector $\mathbf{w}_{i,j}$ is a column vector of \mathbf{W}_i . Thus, in the training phase, the i^{th} class is represented by a vector space \mathbf{W}_i , which is called the regressor or predictor for each subject.

If \mathbf{y} belongs to the i^{th} class, it can be represented as a linear combination of the training images from the i^{th} class and can be defined as

$$\mathbf{y} = \mathbf{W}_i \boldsymbol{\beta}_i + \mathbf{e}, \quad i = 1, 2, \dots, N. \quad (2)$$

where $\boldsymbol{\beta}_i \in \mathbb{R}^{p_i \times 1}$ is the vector of regression parameters and \mathbf{e} is an error vector that is an i.i.d. random variable with zero mean and variance σ^2 . The goal of regression is to find $\tilde{\boldsymbol{\beta}}_i$, which minimizes the residual errors as

$$\tilde{\boldsymbol{\beta}}_i = \arg \min_{\boldsymbol{\beta}_i} \|\mathbf{W}_i \boldsymbol{\beta}_i - \mathbf{y}\|_2^2, \quad i = 1, 2, \dots, N. \quad (3)$$

The regression coefficients can be solved through the least-square estimation and can be written as a matrix form as

$$\tilde{\boldsymbol{\beta}}_i = (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \mathbf{W}_i^T \mathbf{y}, \quad i = 1, 2, \dots, N. \quad (4)$$

The estimated vector of parameters $\tilde{\boldsymbol{\beta}}_i$ and predictors \mathbf{W}_i are used to predict the response vector $\tilde{\mathbf{y}}_i$ for the i^{th} class as

$$\tilde{\mathbf{y}}_i = \mathbf{W}_i \tilde{\boldsymbol{\beta}}_i, \quad i = 1, 2, \dots, N. \quad (5)$$

By substituting (4) for $\tilde{\boldsymbol{\beta}}_i$ in (5), we have

$$\tilde{\mathbf{y}}_i = \mathbf{W}_i (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \mathbf{W}_i^T \mathbf{y}, \quad i = 1, 2, \dots, N. \quad (6)$$

Therefore, we can get a class specific projection matrix as [21],

$$\tilde{\mathbf{y}}_i = \mathbf{H}_i \mathbf{y}, \quad i = 1, 2, \dots, N, \quad (7)$$

where $\tilde{\mathbf{y}}_i$ is the projection of \mathbf{y} onto the subspace of the i^{th} class by the projection matrix, $\mathbf{H}_i = \mathbf{W}_i (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \mathbf{W}_i^T$. It is

noted that the projection matrix is a symmetric matrix and also idempotent.

The LRC is developed based on the minimum reconstruction error. In other words, if the original vector belongs to the subspace of class i , the predicted response vector $\tilde{\mathbf{y}}_i$ will be the closest vector to the original vector. The identity i^* could be determined by calculating the Euclidean distance measure between the predicted response vectors and the original vector as

$$\begin{aligned} i^* &= \arg \min_i \|\tilde{\mathbf{y}}_i - \mathbf{y}\|, \quad i = 1, 2, \dots, N \\ &= \arg \min_i \|\mathbf{H}_i \mathbf{y} - \mathbf{y}\| \end{aligned} \quad (8)$$

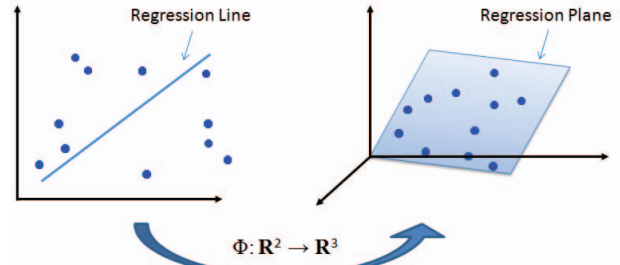


Figure 1. Illustration of a mapping from \mathbb{R}^2 to \mathbb{R}^3 . The left figure shows it is difficult to fit the data by a regression line because of nonlinear data distribution, whereas the right figure shows it is easy to fit the data by a regression plane because of linear data distribution.

3. KERNEL LINEAR REGRESSION CLASSIFICATION (KLRC)

The LRC has been demonstrated that it could achieve good performance for certain conditions, but not for severe illumination variations. This is because illumination variation makes the data distribution more complicated. Assume the original input space can always be mapped to some higher dimensional feature space where the data set is distributed linearly. As depicted in Fig. 1, the left figure shows it is hard to fit the data by a regression line because of nonlinear data distribution, whereas the right figure shows it is easy to fit the data by a regression plane because of linear data distribution by a mapping from \mathbb{R}^2 to \mathbb{R}^3 . Thus, it can be expected that a nonlinear mapping prior to linear regression may improve the limitation of the LRC.

So, the key methodology of the KLRC is to apply a nonlinear mapping function to the input column vector $\Phi(\mathbf{w}_{i,j}) : \mathbb{R}^q \rightarrow \mathbb{R}^f$ and then solve for LRC in the resulting feature space \mathbb{R}^f . Note that the dimensionality of the space \mathbb{R}^f is arbitrarily large.

Specifically, each column vector $\mathbf{w}_{i,j}$ is projected from the original space \mathbb{R}^q to a high dimensional space \mathbb{R}^f by a nonlinear mapping function $\Phi : \mathbb{R}^q \rightarrow \mathbb{R}^f$, $f > n$. Therefore, \mathbb{R}^f is now the space spanned by $\Phi(\mathbf{w}_{i,j})$. Therefore, the projected vectors can be used for regression as

$$\mathbf{y} = \Phi(\mathbf{W}_i) \boldsymbol{\beta}_i, \quad i = 1, 2, \dots, N. \quad (9)$$

Because of the increase in dimensionality, the mapping $\Phi(\mathbf{w}_{i,j})$ is made implicitly by the use of kernel function satisfying Mercer's theorem. By using dual representation $\beta_i = \Phi(\mathbf{W}_i)^T \alpha_i$, the projection stated in (9) becomes

$$\begin{aligned} \mathbf{y} &= \Phi(\mathbf{W}_i) \Phi(\mathbf{W}_i)^T \alpha_i, \quad i=1,2,\dots,N \\ &= \mathbf{K}_i \alpha_i, \end{aligned} \quad (10)$$

The kernel matrix \mathbf{K}_i is positive semi-definite if it is constructed using Mercer kernel. Typically, kernels include polynomial kernel and Gaussian kernel, all of which satisfy Mercer's theorem. By using singular value decomposition (SVD), we can discard the eigenvalues, which are too small, to reduce noise and prevent over fitting. Then, the low rank-r singular value factorization of \mathbf{K}_i is \mathbf{K}_i^r . In this paper, we discard 50% singular values. Then, we have

$$\mathbf{y} \approx \mathbf{K}_i^r \alpha_i. \quad (11)$$

The goal of regression is to minimize the residual errors as

$$\tilde{\alpha}_i = \arg \min_{\alpha_i} \|\mathbf{K}_i^r \alpha_i - \mathbf{y}\|_2^2. \quad (12)$$

This can be solved through the least-square estimation. Because \mathbf{K}_i^r does not have full rank, the pseudoinverse can be used. So, we get

$$\tilde{\alpha}_i = (\mathbf{K}_i^r)^+ \mathbf{y}, \quad (13)$$

and because of low rank approximation, $\mathbf{K}_i^r (\mathbf{K}_i^r)^+ \neq \mathbf{I}$. The response vector $\tilde{\mathbf{y}}_i$ for the i^{th} class can be predicted by

$$\tilde{\mathbf{y}}_i = \mathbf{K}_i^r \tilde{\alpha}_i. \quad (14)$$

By substituting (13) in (14), therefore, we can obtain

$$\tilde{\mathbf{y}}_i = \mathbf{P}_i \mathbf{y} \quad (15)$$

and get a class specific kernel projection matrix as

$$\mathbf{P}_i = \mathbf{K}_i^r (\mathbf{K}_i^r)^+, \quad (16)$$

where $\tilde{\mathbf{y}}_i$ is the projection of \mathbf{y} onto the kernel subspace of the i^{th} class by the class specific kernel projection matrix, \mathbf{P}_i .

The KLRC is developed based on the minimum reconstruction error. In recognition phase, the identity i^* could be determined by calculating the Euclidean distance measure between the predicted response vectors and the original vector as

$$\begin{aligned} i^* &= \arg \min_i \|\tilde{\mathbf{y}}_i - \mathbf{y}\|, \quad i=1,2,\dots,N \\ &= \arg \min_i \|\mathbf{P}_i \mathbf{y} - \mathbf{y}\| \end{aligned} \quad (17)$$

4. EXPERIMENTAL RESULTS

We have examined our algorithm on publicly available face databases: the Yale Face Database B (Yale B) [18]. Also, we evaluate the proposed method against low resolution problem. All experimental results only reported the top 1 recognition accuracy (%).

The Yale B contains images of 10 individuals with 9 poses and 64 illuminations per pose. The frontal face images of all subjects, each with 64 different illuminations are used for evaluation. All images are cropped and resized to 8x8 pixels, as shown in Fig. 2. The Yale B is divided into five subsets based on the angle of the light source directions. As a result, there are total 640 images: 70 (7 images per person), 120 (12 images per person), 120 (12 images per person), 140 (14 images per person), and 190 (19 images per person) images in Subsets 1 to 5, respectively. We follow the evaluation protocol as reported in [18]. Training is conducted using Subset 1 and the remaining subsets (Subsets 2 to 5) are used for testing.

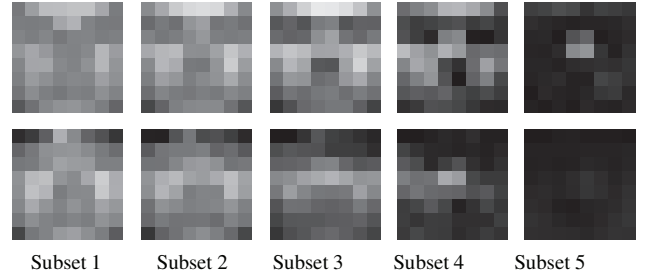


Figure 2. Ten samples in size of 8x8 pixels from each subset of two persons from the Yale B face database.

In the experiments, we compare the proposed methods, KLRC(p) and KLRC(g), with LRC, PCA, KPCA(p), KPCA(g), LDA, KLDA(p), KLDA(g), and ICA, where p and g denote the polynomial kernel and Gaussian kernel, respectively. The results are depicted in Fig. 3, which reflects that the KLRC attains higher recognition rate than the LRC, PCA, KPCA, LDA, KLDA, and ICA for low resolution face recognition under variable lighting. Particularly, in Subsets 4 and 5, the KLRC has gained improvement significantly.

It is noted that although it is widely accepted that the discriminant based approaches offer high robustness to lighting variation, they still cannot withstand severe illumination variation.

5. CONCLUSIONS

In this paper, we propose a class specific kernel linear regression classification for low resolution face recognition under variable lighting. Using the nonlinear mapping function enhances the modeling capability for highly nonlinear data distribution such as illumination variation. We have demonstrated that the proposed KLRC performs better than the LRC, PCA, KPCA, LDA, and KLDA for low resolution face recognition under variable lighting. In summary, the KLRC improves the limitation of the LRC dramatically for face recognition under severe illumination variations and could be an illumination invariant feature extraction method.

The future research directions include the robustness issues related to expression and pose variations, and the reliable estimates of regression coefficients from a single training sample per class. On the other hand, the selection of the optimal kernel remains open.

6. REFERENCES

- [1] W. Zhao, R. Chellappa, A. Rosenfeld, P.J. Phillips, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.
- [2] M. Turk and A. Pentland. "Eigenfaces for Recognition." *Journal of Cognitive Science*, pp.71-86, 1991.
- [3] P. Belhumeur, J. Hespanha, and D. Kriegman, "EigenFaces vs. Fisherfaces: Recognition using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, 1997.
- [4] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, pp. 228-233, 2001.
- [5] P.C. Yuen and J.H. Lai, "Face Representation Using Independent Component Analysis," *Pattern Recognition*, vol. 35, no. 6, pp. 1247-1257, 2002.
- [6] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans Neural Networks*, vol. 13, no. 6, pp. 1450-1464, 2002.
- [7] J. Yang, D. Zhang, and J. Yang, "Constructing PCA Baseline Algorithms to Reevaluate ICA-Based Face-Recognition Performance," *IEEE Trans. Systems, Man, and Cybernetics, Part B*, pp.1015-1021, 2007.
- [8] B. Moghaddam, A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696-710, 1997.
- [9] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299-1319, 1998.
- [10] M. H. Yang, "Kernel Eignefaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," *Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 215-220, May 2002.
- [11] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, pp. 2385-2404, 2000.
- [12] S. Mika *et al.*, "Fisher discriminant analysis with kernels," *Neural Networks for Signal Proc. IX*, pp. 41-48, 1999.
- [13] Imran Naseem, Roberto Togneri, Mohammed Bennamoun, "Linear Regression for Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106-2112, July 2010.
- [14] R Barsi and D Jacobs, "Lambertian Reflection and Linear Subspaces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 3, pp. 218-233, 2003.
- [15] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 721-732, Jul. 1997.
- [16] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible illumination conditions," *Int. J. Comput. Vis.*, vol. 28, no. 3, pp. 245-260, Jul. 1998.
- [17] H. F. Chen, P. N. Belhumeur, and D. J. Kriegman, "In search of illumination invariants," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 13-15.
- [18] A. S. Georghiades, P. N. Belhumeur, and D. W. Jacobs, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 630-660, Jun. 2001.
- [19] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," in *Proc. IEEE Workshop on AMFG*, 2003, pp. 157-164.
- [20] M. Savvides and V. Kumar, "Illumination normalization using logarithm transforms for face authentication," in *Proc. IAPR AVBPA*, 2003, pp.549-556.
- [21] David C. Hoaglin, Roy E. Welsch, "The Hat Matrix in Regression and ANOVA," *The American Statistician*, vol. 32, no. 1, pp. 17-22, February 1978.

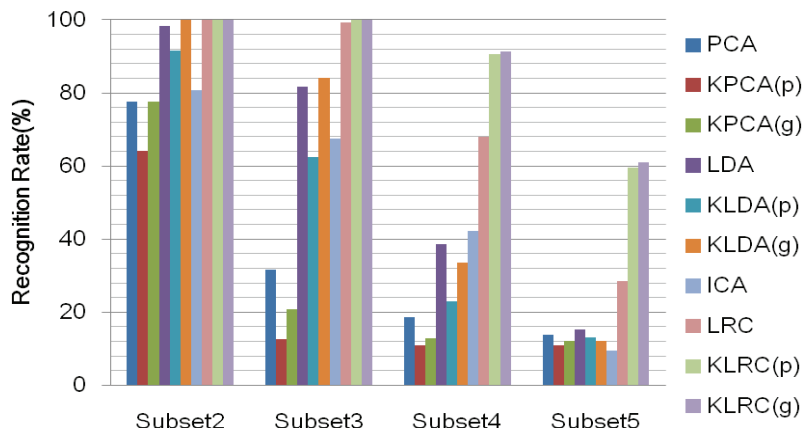


Figure 3. Performance comparison on Yale B in size of 8×8 pixels.