ITERATIVE LINEAR REGRESSION CLASSIFICATION FOR IMAGE RECOGNITION

Qingxiang Feng and Yicong Zhou

Department of Computer and Information Science
University of Macau, Macau 999078, China
qxfeng@umac.mo, yicongzhou@umac.mo

ABSTRACT

Traditional linear regression classification (LRC) suffers from a small sample size problem that the limited training samples of each class cannot comprehensively reflect different variations of the class. To address the problem, this paper proposes a novel iterative linear regression classification (ILRC) for image recognition. Different from traditional LRC, ILRC not only generates several new subspaces in each iteration but also uses the discrimination idea to optimize the training-set and testing samples. Extensive experiments on five benchmark databases demonstrate that the proposed ILRC classifier achieves better recognition performance than the traditional LRC and several state-of-the-art methods.

***Index Terms*—** Linear Regression, Object Recognition, Face Recognition.

1. INTRODUCTION

Pattern recognition systems are known to be critically dependent on classifiers. The nearest neighbor (NN) [1] and nearest subspace (NS) [2] methods are well-known approaches in the pattern recognition area. To classify the testing sample, NN uses the best representation of a training sample, while NS is based on the best linear representation of all training samples in each class. Samples from a specific object class are known to lie on a linear subspace [3, 4]. Using this concept, the locally linear regression (LLR) [5] was proposed to specifically address pose variations for face recognition. Using the similar concept, the linear regression classification (LRC) [6] was proposed to develop class-specific models of the registered users and to redefine the face recognition task as a linear regression problem. Since 2010, the LRC-associated approaches [7] have been proposed to prove the recognition performance of LRC under different situations like variable illuminations and facial expressions. These approaches include the kernel-LRC [8, 9], improved-PCA-LRC [10], linear discrimination regression classification (LDRC) [11] and unitary regression classification (URC) [12]. Different from the class-model of LRC, sparse-representation-based classification (SRC) [13, 14] adopts the all-class-model to classify the testing sample. After the SRC classifier being proposed, several improved classifiers were presented for face recognition. For example, Zhang et al. proposed the collaborative-representation-based classification (CRC) [15]. Xu et al. proposed two-phase sparse representation (TPSR) [16] and so on. In this paper, the iterative linear regression classification (ILRC) is proposed for image recognition. ILRC finds the novel class subspace by iteratively removing the unimportant elements, To compare the recognition performance of ILRC with those of the LRC classifier and several state-of-the-art methods, we carry out extensive experiments on the Coil-100 object database, Eth80 object database, NCKU CSIE Robotics Lab face database, GT face database and AR face database.

This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/106/2013/A3 and by the Research Committee at University of Macau under Grants MYRG2014-00003-FST, and MRG017/ZYC/2014/FST.

2. PROPOSED METHOD

After briefly reviewing the traditional LRC, this section first introduces the iterative refinement of the subspace and testing sample. Using these novel subspaces and testing sample, we further propose a novel classifier, called the iterative linear regression classification (ILRC)

2.1. Traditional LRC

Let$X = \{x_i^c\}, c = 1, 2, \cdots, M, i = 1, 2, \cdots, N_c$ denote the prototype set, where x_i^c is the i^{th} prototype belonging to the c^{th} class, M is the number of classes, and N_c is the number of prototype-samples in the c^{th} class.
Because patterns from the same class lie on a linear subspace, the traditional LRC classifier forms a class-model X_c by stacking the q-dimensional image vectors belonging to the same class.

$$X_c = [\ x_1^c \quad x_2^c \quad .... \quad x_{N_c}^c\ ] \in R^{q \times N_c} \tag{1}$$

Let x be an unlabeled test image vector. If x belongs to the c^{th} class, it may be represented as a linear combination of the training images from the same class.

$$x \approx x^c = X_c \beta_c \tag{2}$$

where the predicted vector x^c is the projection of x onto the c^{th} class subspace and the vector of regression coefficients $\beta_c \in R^{N_c \times 1}$ can be calculated by

$$\beta_c = ({X_c}^T X_c)^{-1} {X_c}^T x \tag{3}$$

LRC computes the distance measure between the predicted vector x^c and the original testing vector x as

$$d_c(x) = ||x - x^c|| \tag{4}$$

where $|| * ||$ means L_2- norm. The rule of LRC is to choose the class with the minimum distance as

$$\min_{c^*} d_c(x) \quad , \quad c = 1, 2, ..., M \tag{5}$$

However, LRC has a small sample size problem that limited training samples of each class cannot comprehensively reflect different variations of the class. To solve this problem, we propose a novel method to optimize the subspace as follows.

2.2. Iterative Refinement of Subspace and Test Sample

Given the c^{th} class subspace determined by N_c prototype sample vectors $\{x_1^c, x_2^c ... x_{N_c}^c\}$. For a testing sample x, the linear regression to the c^{th} class can be described by

$$\mathbf{x} = \sum_{i=1}^{N_c} \alpha_i \mathbf{x}_i^c \tag{6}$$

If the mean of the prototype sample vectors is computed as $\bar{\mathbf{x}}^{c,0} = (1/N_c) \sum_{i=1}^{N_c} \mathbf{x}_i^c$, we can remove the mean and discard a selected unlikely sample $\mathbf{x}_{j*}^c$. After the first round of subspace refinement, the remaining $N_c - 1$ prototype samples can be represented as,

$$.\mathbf{x}_i^{c,1} = \mathbf{x}_i^c - \bar{\mathbf{x}}^{c,0}, i = 1, 2, ..., N_c - 1 \tag{7}$$

It can also be described by a newly updated subspace $\{\mathbf{x}_1^{c,1}$, $\mathbf{x}_2^{c,1}, ..., \mathbf{x}_{N_c-1}^{c,1}\}$. Accordingly, the testing sample also needs to remove the prototype mean as

$$\mathbf{x}^{c,1} = \mathbf{x} - \bar{\mathbf{x}}^{c,0} \tag{8}$$

Similarly, after the k^{th} iteration (for $k > 1$), the $N_c - k$ prototype sample vectors are given by

$$\mathbf{x}_i^{c,k} = \mathbf{x}_i^{c,k-1} - \bar{\mathbf{x}}^{c,k-1} \tag{9}$$

and the k^{th} updated testing sample is expressed by

$$\mathbf{x}^{c,k} = \mathbf{x}^{c,k-1} - \bar{\mathbf{x}}^{c,k-1} \tag{10}$$

where

$$\bar{\mathbf{x}}^{c,k-1} = \frac{1}{(N_c - k + 1)} \sum_{i=1}^{N_c - k + 1} \mathbf{x}_i^{c,k-1} \tag{11}$$

After the k^{th} iteration process of discarding a selected unlikely sample, the updated $N_c - k$ prototype samples can be represented by the k^{th} refined subspace $\{\mathbf{x}_1^{c,k}, \mathbf{x}_2^{c,k}, ..., \mathbf{x}_{N_c-k}^{c,k}\}$. The selection of the unlikely samples in each iteration will be discussed in Section 2.3.

2.3. Proposed ILRC

First, stacking the q-dimensional image vectors, the class-specific model X_c is given as

$$\mathbf{X}_c = [\ \mathbf{x}_1^c \quad \mathbf{x}_2^c \quad \cdots \quad \mathbf{x}_{N_c}^c\ ] \in R^{q \times N_c} \tag{12}$$

By iteratively discarding a selected unlikely sample, the updated $N_c - k$ prototype samples at the k^{th} iteration can be represented by

$$\mathbf{X}_{c,k} = \{\mathbf{x}_1^{c,k}, \mathbf{x}_2^{c,k}, ..., \mathbf{x}_{N_c-k}^{c,k}\} \tag{13}$$

Let x be an unlabeled testing image and $\mathbf{x}^{c,0} = \mathbf{x}$ for each c. Using the traditional LRC, the k^{th} updated testing sample vector in the c^{th} class $\mathbf{x}^{c,k}$ can be represent by

$$\mathbf{x}^{c,k} = \sum_{i=1}^{N_c-k} \beta_i^{c,k} \mathbf{x}_i^{c,k} \tag{14}$$

where the optimal regression coefficient vector $\beta^{c,k} \in R^{(N_c-k)\times 1}$ can be calculated as

$$\beta^{c,k} = (\mathbf{X}_{c,k}^T \mathbf{X}_{c,k})^{-1} \mathbf{X}_{c,k}^T \mathbf{x}^{c,k}, foreachc \tag{15}$$

The k^{th} updated distance between the testing sample $x^{c,k}$ and the subspace $X_{c,k}$ can be computed as

$$d^{c,k} = ||x_c - \mathbf{X}_{c,k}^T \beta^{c,k}|| \tag{16}$$

If $\beta_{j*}^{c,k}$ has the minimum magnitude among the elements of $\beta^{c,k}$ as

$$j^c* = \min_i \left| \beta_i^{c,k} \right| \tag{17}$$

Based on the linear regression defined in Eqn.(14), the corresponding prototype sample $\mathbf{x}_{j^c*}^{c,k}$ will be the most unlikely one. Hence, in the k^{th} iteration, $\mathbf{X}_{c,k} \in R^{q \times (N_c - k)}$ becomes the class subspace by removing the mean and the most unlikely prototype sample, $\mathbf{x}_{j^c*}^{c,k}$ as Eqns. (9) and (11). The k^{th} updated testing sample in (14) is iteratively computed by Eqn. (10). This new iterative class subspace will be continuously updated until the number of samples of the new class subspace is equal to one. The distance between the testing sample and the c^{th} class is described as

$$d_c = \min(d^{c,k}), k = 1, 2, ..., N_c, for, all, c \tag{18}$$

Algorithm 1: Iterative Linear Regression Classification
Input: The original testing sample x and original prototype-set $X = \{x_i^c, c = 1, 2, \cdots, M, i = 1, 2, \cdots, N_c\}$ **Output:**The index of x.
Step 1: Construct the M class-models X_c using Eqn. (12).
Step 2: For the $k^{th}(k = 1, 2, ..., N_c)$ iteration, do
Step 2.1: Obtain the novel class-model $X_{c,k}$ and novel test vector $x^{c,k}$ by (13) and (14), respectively
Step 2.2: Use Eqn. (16) to compute the k^{th} distance $d^{c,k}$ between the prototype subspace $X_{c,k}$ and test sample $x^{c,k}$ in the c^{th} class.
Step 2.3: Select the minimum distance in the iterations by Eqn. (18),d_c , which will be treated as the distance between the test sample x and the c^{th} class subspace.
Step 3: The classification rule of ILRC is to select the class with the minimum distance according to Eqn. (19).

Finally, the ILRC classifier selects the class with the minimum one as given by

$$\min_{c*} d_c, c = 1, 2, ..., M \tag{19}$$

Algorithm 1 shows the detailed classification procedures of ILRC.

2.4. ILRC vs LRC

Because each testing image can be represented as a linear combination of all training images of a specific class, ILRC is a statistical method similar to LRC. However, ILRC is different from LRC in following aspects.
(1) LRC utilizes the original training-set to solve the least square error for classification. However, ILRC generates new training images in each iteration. This allows ILRC to have strongly representational capability and thus a better recognition rate than LRC.
(2) Motivated by LDA [17] and LDRC [11], ILRC increases the discriminative information in the training set and testing sample in each iteration. This is helpful for classification. But LRC fails to consider it.

3. EXPERIMENTAL RESULTS

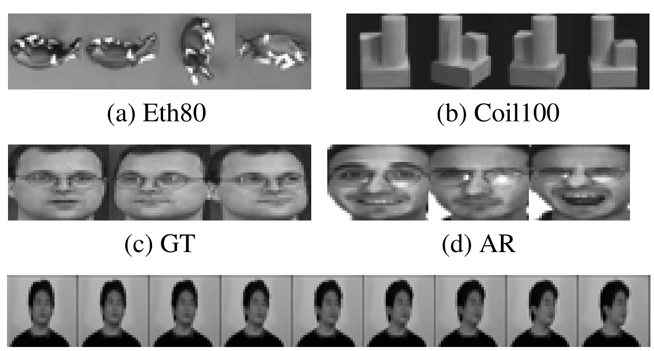
To verify the performance of the proposed algorithm for object and face recognition, ILRC is compared with several existing methods on five well-known databases. These methods include the LRC [6], URC [12], LDRC [11], SVM [18], LDA [17], SRC [14], CRC [15], TPSR [16], SFR [19], M-RC [20] and DLRC [21]. The optimization parameter is set 0.01 in the experiments.

Table I: RECOGNITION RATES OF SEVERAL CLASSIFIERS ON THE ETH80 AND COIL100 OBJECT DATABASES (%)

	Eth80		Coil100	
	3	4	3	4
SVM	9.14	10.57	58.22	58.13
LDA	9.97	9.73	63.44	58.75
CRC	11.25	12.80	65.56	65.00
SRC	12.86	14.97	68.89	67.75
TPSR	9.24	10.51	67.00	64.88
LRC	15.46	18.31	74.89	73.75
URC	15.53	18.11	74.89	73.75
SFR	6.74	8.34	60.11	58.25
DLRC	8.91	10.77	64.44	64.88
ILRC	19.57	22.53	76.56	74.88

3.1. Experimental setups of five databases

1): The eth80 object database [22]: In our experimental, all images are resized to 32 × 32 grayscale images.
2): The Coil-100 dataset [23]: Our experiments use 12 different view angles per object (0, 30, 60,..., 330). The subset of Coil-100 dataset contains 1200 images downsampled into size of 32 × 32.
3): NCKU CSIE Robotics Lab face database [24]: It contains 6660 images of 90 subjects. In our experimental, we use the part A containing 3330 images of 90 subjects. Each image is cropped into size of 40 × 40.
4): Georgia Tech face database [25]: It contains images of 50 people. Each image is manually cropped into size of 30 × 40.
5): The AR database [26]: To test the performance of ILRC with expression variations, a subset of the AR database includes 600 face images of 100 individuals with different expressions (smile, anger, scream). All images are cropped into 50 × 40 pixels.



(a) Eth80 (b) Coil100

(c) GT (d) AR

(e) NCKU

Fig. 1: Sample images from the (a) eth80 object database, (b) Coil100 object database, (c) GT face database, (d) AR face database, and (e) NCKU CSIE Robotics Lab face database.

Table II: RECOGNITION RATES OF SEVERAL CLASSIFIERS ON THE AR FACE DATABASE (%)

Classifier	Scream	Anger	Smile	Mean
SVM	63.50	82.50	88.50	78.17
LDA	44.00	77.00	73.50	64.83
CRC	74.50	96.50	97.00	89.33
SRC	80.50	95.00	96.50	90.67
TPSR	76.00	96.00	97.50	89.83
LRC	78.50	95.50	93.50	89.17
LDRC	74.00	95.00	98.00	73.75
SFR	58.00	90.00	89.50	79.17
DLRC	66.00	93.50	94.00	84.50
ILRC	80.50	98.00	97.00	91.83

Table III: THE RECOGNITION RATE OF SEVERAL CLASSIFIERS ON THE GT FACE DATABASE (%)

Classifier	3	4	5	Mean
SVM	31.00	36.00	40.40	35.80
LDA	40.83	33.82	31.00	35.22
CRC	46.00	49.64	54.60	50.08
TPSR	53.17	55.82	59.20	56.06
LRC	51.83	56.36	59.20	55.80
URC	51.83	56.36	59.20	55.80
LDRC	41.33	41.82	44.20	42.45
SFR	40.83	44.55	50.00	45.13
DLRC	45.50	52.00	56.80	51.43
ILRC	53.83	58.18	59.80	57.27

3.2. Object Recognition

Here, we use the eth80-cropped-close128 and Coil-100 databases to verify the performance of ILRC for object recognition. In the experiments, we choose 3 and 4 images of each object from two databases as the prototype set and the rest images form the testing set. We evaluate the performance of several classifiers for object recognition. From the experiment results in Table I, we can conclude that (1) the performance of all classifiers changes dramatically in different databases. This is because object recognition is a complicated and challenging task. (2) ILRC obtains the best performance among all compared classifiers for object recognition.

3.3. Face Recognition with Expression Variations

In this subsection, the AR face database is used to evaluate the performance of ILRC for face recognition with expression variations. We first test the recognition rate of the ’smile’ expression where all smile face images are used as the testing set and the rest face images (anger and scream) form the training set. Similar experiment setups conduct for evaluating expressions: ’anger’ and ’scream’. The results are shown in Table II. As can be seen, all methods have different recognition performances for different expressions. For example, all methods have lower recognition rates on ’scream’ than those on ’smile’. ILRC performs the best in these compared methods for face recognition with expression variation.

3.4. Face Recognition with Gesture Variation

This subsection uses the Georgia Tech and NCKU CSIE Robotics Lab face databases to verify the performance of ILRC for face recognition with gesture variations. In the experiments, we choose 3, 4 and 5 images of each person from two databases as the prototype set and the rest images are used as the testing set. We compare ILRC with several existing classifiers for face recognition with gesture variations. The experimental results are shown in Tables III and Table IV.

Table IV: THE RECOGNITION RATE OF SEVERAL CLASSIFIERS ON THE NCKU CSIE ROBOTICS LAB FACE DATABASE (%)

Classifier	3	4	5	Mean
SVM	40.65	41.41	42.92	41.03
CRC	45.49	45.56	46.63	45.53
SRC	50.44	50.56	51.31	50.50
TPSR	45.39	44.28	46.49	44.84
LRC	49.61	49.02	49.76	49.31
URC	51.54	51.62	52.88	51.58
SFR	40.92	41.82	43.37	41.37
DLRC	42.29	42.16	44.86	42.23
ILRC	54.15	54.24	55.21	54.20

We can observe that the sparse-based and linear-regression-based methods obtain unsatisfactory recognition performance for this task. However, ILRC outperforms other existing classifiers for face recognition with gesture variations.

4. CONCLUSION

In this paper, iterative linear regression classification (ILRC) has been proposed for image recognition. ILRC uses the novel iterative class subspaces to increase the number of prototype samples such that the class subspace have better representative capability for classification and recognition. Experiment results on five benchmark datasets have demonstrated that the proposed ILRC achieves the best recognition rates compared with the LRC classifier and several state-of-the-art methods.What is more, we know that the object recognition is a complicated and challenging task so that the performance of the classifiers changes dramatically in different database. Besides, we know that the sparse-based and linear-regression-based methods are sensitive to face recognition with gesture variations.

5. REFERENCES

[1] Thomas M Cover and Peter E Hart, "Nearest neighbor pattern classification," *Information Theory, IEEE Transactions on*, vol. 13, no. 1, pp. 21–27, 1967.

[2] Jason Ho, Ming-Hsuan Yang, Jongwoo Lim, Kuang-Chih Lee, and David Kriegman, "Clustering appearances of objects under varying illumination conditions," in *CVPR*. IEEE, 2003, vol. 1, pp. I–11.

[3] Peter N Belhumeur, João P Hespanha, and David J Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *T-PAMI*, vol. 19, no. 7, pp. 711–720, 1997.

[4] Ronen Basri and David W Jacobs, "Lambertian reflectance and linear subspaces," *T-PAMI*, vol. 25, no. 2, pp. 218–233, 2003.

[5] Xiujuan Chai, Shiguang Shan, Xilin Chen, and Wen Gao, "Locally linear regression for pose-invariant face recognition," *T-IP*, vol. 16, no. 7, pp. 1716–1725, 2007.

[6] Imran Naseem, Roberto Togneri, and Mohammed Bennamoun, "Linear regression for face recognition," *T-PAMI*, vol. 32, no. 11, pp. 2106–2112, 2010.

[7] Jeng-Shyang Pan, Qingxiang Feng, Lijun Yan, and Jar-Ferr Yang, "Neighborhood feature line segment for image classification," *T-CSVT*, vol. 25, no. 3, pp. 387–398, 2015.

[8] Shih-Ming Huang and Jar-Ferr Yang, "Kernel linear regression for low resolution face recognition under variable illumination," in *ICASSP*. IEEE, 2012, pp. 1945–1948.

[9] Qingxiang Feng, Chun Yuan, Jiawen Huang, and Weifeng Li, "Center-based weighted kernel linear regression for image classification," in *ICIP*. IEEE, 2015, pp. 3630–3634.

[10] Shih-Ming Huang and Jar-Ferr Yang, "Improved principal component regression for face recognition under illumination variations," *Signal Processing Letters, IEEE*, vol. 19, no. 4, pp. 179–182, 2012.

[11] Shih-Ming Huang and Jar-Ferr Yang, "Linear discriminant regression classification for face recognition," *SPL, IEEE*, vol. 20, no. 1, pp. 91–94, 2013.

[12] Shih-Ming Huang and Jar-Ferr Yang, "Unitary regression classification with total minimum projection error for face recognition," *Signal Processing Letters, IEEE*, vol. 20, no. 5, pp. 443–446, 2013.

[13] John Wright, Yi Ma, Julien Mairal, Guillermo Sapiro, Thomas S Huang, and Shuicheng Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.

[14] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma, "Robust face recognition via sparse representation," *T-PAMI*, vol. 31, no. 2, pp. 210–227, 2009.

[15] Lei Zhang, Meng Yang, and Xiangchu Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *ICCV*. IEEE, 2011, pp. 471–478.

[16] Yong Xu, David Zhang, Jian Yang, and Jing-Yu Yang, "A two-phase test sample sparse representation method for use with face recognition," *T-CSVT*, vol. 21, no. 9, pp. 1255–1262, 2011.

[17] Aleix M Martínez and Avinash C Kak, "Pca versus lda," *T-PAMI*, vol. 23, no. 2, pp. 228–233, 2001.

[18] Bernd Heisele, Purdy Ho, and Tomaso Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *ICCV*. IEEE, 2001, vol. 2, pp. 688–694.

[19] Yong Xu and Qi Zhu, "A simple and fast representation-based face recognition method," *NCA*, vol. 22, no. 7-8, pp. 1543–1549, 2013.

[20] Jie Xu and Jian Yang, "Mean representation based classifier with its applications," *Electronics letters*, vol. 47, no. 18, pp. 1024–1026, 2011.

[21] Qingxiang Feng, Qi Zhu, Lin-Lin Tang, and Jeng-Shyang Pan, "Double linear regression classification for face recognition," *Journal of Modern Optics*, vol. 62, no. 4, pp. 288–295, 2015.

[22] Bastian Leibe and Bernt Schiele, "Analyzing appearance and contour based methods for object categorization," in *CVPR*. IEEE, 2003, vol. 2, pp. II–409.

[23] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al., "Columbia object image library (coil-20)," Tech. Rep., Technical Report CUCS-005-96, 1996.

[24] Yaser Arianpour, Sedigheh Ghofrani, and Hamidreza Amindavar, "Locally nonlinear regression based on kernel for pose-invariant face recognition," in *ISSPA*. IEEE, 2012, pp. 401–406.

[25] Ara V Nefian, "Embedded bayesian networks for face recognition," in *ICME*. IEEE, 2002, vol. 2, pp. 133–136.

[26] Aleix M Martinez, "The ar face database," *CVC Technical Report*, vol. 24, 1998.