Bi-directional CRC algorithm using CNN-based features for face classification

Yanan Wang1, Tian Na1, Xiaoning Song1, Guosheng Hu2
1School of Internet of Things Engineering, Jiangnan University, Wuxi, People’s Republic of China
2Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK
E-mail: 6171910011@stu.jiangnan.edu.cn

Abstract: Collaborative representation-based classification (CRC) has become a breakthrough in face classification due to its distinguished collaborative capacity. Nevertheless, insufficient observations of per subject are usually offered by few or even a single gallery image for face classification, which lead to a sensitive response to the variations from the original data set. In this study, the authors present a bi-directional CRC algorithm using convolutional neural network-based features for face classification. They first employ a deep convolutional neural network to extract facial features from the original gallery and query sets, and then develop a fast reverse representation model to obtain the auxiliary residual information between each training sample and the reconstructed one that is achieved from the test sample. Secondly, they offer a new solution to the bi-directional optimisation problem by which the input sample is well represented by the forward linear combination and the reverse one, respectively. The last contribution is to utilise a competitive fusion method for robust face recognition, which weighted reconstructed residuals from the bi-directional representation model. Experimental results obtained from a set of well-known face databases including AR, FERET, and ORL verify the validity of the proposed method, especially in the robustness to small sample size problem.

1 Introduction

It is well known that the amazing breakthrough of sparse representation-based classification (SRC) theory [1–4] facilitates the study of collaborative-representation-based classification (CRC) [5] in signal reconstruction. However, only very few or even a single face image is usually offered to perform face recognition in some practical applications. Despite SRC and CRC have utilised the sparsity properties in the dataset, they are unable to deal with the problem such as pose, illumination, expression, and occlusion changes very well. To solve this problem, many previous works have been presented to construct optimised form of representation, by synthesising virtual training images or enhancing the solution performance of the optimisation problem. For example, new face images of different poses are constructed by an example-based method and the recognition rate is improved in [6, 7]. Meanwhile, face symmetry is very important for fast locating candidate samples, which has achieved success in applications of face detection and classification [8–14]. Recently, Xu et al. integrated the original image and the mirror image for representation-based face classification [15]. Another approach that generates ‘symmetrical’ synthesised face images is also exploited for face recognition by Xu et al. [16].

Convolutional and inverse representation-based linear regression classification (CIRLR) [17] applies the reverse representation into the algorithm for the first time. However, it has exploited a crude complementary representation-based classification (RBC) method; a large number of constructed systems of linear equations may lead to over-fitting because of the redundancy of auxiliary residual between each training sample and the reconstructed one. It also increases the computational cost of the traditional RBC to a great extent. In fact, few attempts have been made to avoid the uncertainty problem in terms of the reverse representation by evaluating the relationship between the forward representation and the reverse model. Uncertainty problem of the optimisation problem is usually derived from inaccurate information, or a lack of knowledge about the efficient key procedure of representation, for instance, the traditional CIRLR model based on the raw feature space may come from the extremely complex data sources that are unreliable. Moreover, it is generally known that a convolutional neural network (CNN) [18, 19] has drawn extensive attention due to its successful applications in a wide range of pattern recognition and computer vision. Also, the robust textural features can be extracted by the CNN network trained on a large number of face images, across a variety of variations such as illumination, pose, occlusion etc. This motivates us to exploit a new bi-directional CRC algorithm using CNN-based features (BCRC-CNN) for robust face classification.

In this study, we develop a BCRC-CNN algorithm to expand a fast hybrid RBC model in the CNN feature space using an efficient bi-directional representation strategy to evaluate the residual between each input sample and the reconstructed one. The contributions to our work are described as 4-fold below:

• We employ a deep CNN, projecting the process of classification from the raw pixel space to the CNN feature space. To this end, we use the pre-trained VGGNet-16 model [20] for feature extraction to alleviate the adverse information from the original gallery and query set, improving the accuracy of the recognition.
• We develop a fast reverse representation model to obtain the auxiliary residual information between each training sample and the reconstructed one that is achieved from the test sample, which is an efficient robust reverse representation mechanism.
• We propose a new solution to the bi-directional optimisation problem by which the input sample is well represented by the forward representation and the reverse one, respectively. The method achieves the residuals from the bi-directional RBC model with the minimum cost.
• We utilise a competitive fusion method to perform face classification based on the BCRC framework by weighting the reconstructed residuals from the bi-directional representation model. Finally, the label of the test sample is estimated by the scores of different subjects.

The rest of the paper is organised from the following four sections: Section 2 introduces the presentation of the VGGNet-16 model and the outline of CIRLR. The proposed method is presented in Section 3 in detail. Section 4 evaluates the effectiveness of BCRC-CNN by experimenting on several common
In this section, we will show the outline of the CIRLRC method. We suppose that there are \( p \) classes and each class has \( M \) training samples. Let \( x^i \) stand for the \( i \)th training sample of the \( k \)th class.

In CIRLRC, the mirror test sample \( y^j \in \mathbb{R}^{w \times b} \) for the original test sample \( y \in \mathbb{R}^{w \times b} \) is produced by

\[
y^j(p, q) = y(p, b - q + 1),
\]

where \( p = 1, \ldots, a \), \( q = 1, \ldots, b \) and \( a, b \) stand for the numbers of rows and columns of the face image matrix, respectively, as well as \( y^j(p, q) \) and \( y(p, q) \) denote the pixels located in the \( p \)th row and \( q \)th column of \( y^j \) and \( y \), respectively. Hence, the virtual training samples \( x^1, \ldots, x^M \) are also generated in the same way. Then \( y^j \) and \( x^i \) are converted into column vectors and the same denotations are used to stand for them.

Let \( X^i = \{x^1, \ldots, x^i, \ldots, x^M\} \), for the \( i \)th class, a test sample \( y \) is approximated using as a linear combination of the training samples from the \( i \)th class

\[
y = \hat{X}a_i, \quad i = 1, \ldots, K,
\]

where \( a_i \in \mathbb{R}^{2 \times M} \) is the coefficient vector of the \( i \)th class. Thus, \( a_i \) can be estimated using the least-square estimation

\[
a_i = (\hat{X}^T\hat{X} + \lambda I)^{-1}\hat{X}^T y,
\]

where \( \lambda \) is a small positive constant. Then the score between the test sample and the \( i \)th class can be obtained by

\[
s_i = \| y - \hat{X}a_i \|.
\]

For the \( j \)th class, let \( Z_j = \{\bar{x}_1, \ldots, \bar{x}_j, \ldots, \bar{x}_K, y, y'\} \), and the linear representation of each native training sample \( \{x_j, m = 1, \ldots, M\} \) from the \( j \)th class can be written as

\[
x_j = \beta_j y, \quad j = 1, \ldots, M
\]

where \( \beta_j \) is calculated using \( \beta_j = Z_j^T Z_j + \lambda I \) and the deviation between the test sample and the \( j \)th training sample of the \( j \)th class is obtained by

\[
d_j = \| y - \beta_j y', j = 1, \ldots, M
\]

where \( \beta_j \) and \( \beta_j \) are the \((K-1)*M + 1\)th and the \((K-1)*M + 2\)th components of \( \beta_j \), respectively. The distance between the test sample and the \( j \)th class can be solved using

\[
d_j = \frac{1}{M} \sum_{i=1}^{M} d_j.
\]

Then the scores and distances of the test sample \( y \) with respect to all classes are normalised by

\[
s_j' = (s_j - s_{\text{min}})/(s_{\text{max}} - s_{\text{min}}),
\]

\[
d_j' = (d_j - d_{\text{min}})/(d_{\text{max}} - d_{\text{min}})
\]

where \( s_{\text{min}} = \min(s_1, s_2, \ldots, s_K) \), \( s_{\text{max}} = \max(s_1, s_2, \ldots, s_K) \), \( d_{\text{min}} = \min(d_1, \ldots, d_K) \), \( d_{\text{max}} = \max(d_1, \ldots, d_K) \). Finally

\[
t_j = w_1 s_j' + w_2 d_j'
\]

is used to calculate the ultimate score with respect to the \( j \)th class where \( w_1 \) and \( w_2 \) are the weights. Let \( w_1 \) be smaller than \( w_1 \) and \( w_1 + w_2 = 1 \). Hence, the label of the test sample is estimated using the index of the smallest values over all \( t_j \)

\[
\text{Label}(y) = \arg \min_j t_j.
\]

2 Related works

2.1 Presentation of the VGGNet-16 model

The VGGNet [20] is originally developed by the Visual Geometry Group. In this study, we adopt the pre-trained VGGNet-16 model for feature extraction. The architecture of the VGGNet-16 network is shown in Fig. 1. It has five convolutional segments; each contains two or three convolution layers. Also, a maximum pooling layer is connected to reduce the size of images. Dropout is also used after the first and second full connection layer to avoid over-fitting.

2.2 Outline of CIRLRC

In this section, we will show the outline of the CIRLRC method. We suppose that there are \( K \) classes and each class has \( M \) training samples. Let \( x_i \) stand for the \( m \)th training sample of the \( k \)th subject.

In CIRLRC, the mirror test sample \( y^j \in \mathbb{R}^{w \times b} \) for the original test sample \( y \in \mathbb{R}^{w \times b} \) is produced by

\[
y^j(p, q) = y(p, b - q + 1),
\]

where \( p = 1, \ldots, a \), \( q = 1, \ldots, b \) and \( a, b \) stand for the numbers of rows and columns of the face image matrix, respectively, as well as \( y^j(p, q) \) and \( y(p, q) \) denote the pixels located in the \( p \)th row and \( q \)th column of \( y^j \) and \( y \), respectively. Hence, the virtual training samples \( x^1, \ldots, x^M \) are also generated in the same way. Then \( y^j \) and \( x_i \) are converted into column vectors and the same denotations are used to stand for them.

Let \( X^i = \{x^1, \ldots, x^i, \ldots, x^M\} \), for the \( i \)th class, a test sample \( y \) is approximated using as a linear combination of the training samples from the \( i \)th class

\[
y = \hat{X}a_i, \quad i = 1, \ldots, K,
\]

where \( a_i \in \mathbb{R}^{2 \times M} \) is the coefficient vector of the \( i \)th class. Thus, \( a_i \) can be estimated using the least-square estimation

\[
a_i = (\hat{X}^T\hat{X} + \lambda I)^{-1}\hat{X}^T y,
\]

where \( \lambda \) is a small positive constant. Then the score between the test sample and the \( i \)th class can be obtained by

\[
s_i = \| y - \hat{X}a_i \|.
\]
Let $f$ equal the objective function, then we have

\[
\frac{\partial f}{\partial A} = -2X^T \gamma Q + 2\beta A^T Q + 2\mu A.
\]

\[(16)\]

Finally, the optimal solutions to (14) are

\[A = (X^T X' + \beta Q^T Q + \mu I)^{-1} X^T \gamma',\]

\[Q = (\alpha y^T y + \beta A^T A + \lambda)^{-1} y^T X'.\]

(17)

(18)

Then the coefficient vectors $A$ and $Q$ can be iteratively computed by (17) and (18) until the predefined termination condition is satisfied.

The last step is to weigh the reconstructed residuals from the bi-directional representation model for the ultimate classification. Let $s_i = \|y' - X'A\|_1^2 + a\|y'Q - X'\|_1^2$ and $d_i = \|X_i' - y'Q\|_1 (i = 1, \ldots, K)$ denote the forward score and the reverse distance between $y$ and the $i$th subject, respectively. Then normalise them by $s'_i = \min (s_1, \ldots, s_K)$, $s''_i = \max (s_1, \ldots, s_K)$, $d'_i = \min (d_1, \ldots, d_K)$, $d''_i = \max (d_1, \ldots, d_K)$, $s_i' = (s_i - s'_i)/(s''_i - s'_i)$, $d_i' = (d_i - d'_i)/(d''_i - d'_i)$ and combine them by $t_i = w_i s_i' + w_i d_i'$ to calculate the ultimate score. $w_i$ and $w_i$ are the weights and $w_i + w_i = 1$. As forward representation being more reliable than inverse representation in evaluating the dissimilarity, the proposed algorithm assigns a larger value to $w_i$ than $w_i$. Finally, the label of $y$ is estimated using

\[
\text{Label}(y) = \arg \min_i t_i.
\]

(19)

4 Experimental results

In this section, we will evaluate the effectiveness of the BCRC-CNN algorithm and compare it with some state-of-the-art algorithms by experimenting on publicly available datasets, including AR [21], FERET [22] and ORL [23]. Images of faces in all three databases are varied with illumination, expression, and pose. In all the experiments, we set the value of $w_i$ to 0.6, 0.7 and 0.8 to compare and expect better experimental results. The parameters $\alpha$, $\beta$, $\mu$ and $\lambda$ were the small positive constants which were set as 0.01 because it had less influence than forward expression, whose abbreviator coefficient was 1. $\beta$ was equal to 0.1, which was used to balance the sparsity between forward representation and inverse representation. $\mu$ and $\lambda$ were the small positive constants which were set as 0.01 to have only little influence on the classification performance.) Moreover, we repeated each experiment 20 times and took the average result.

The AR database [21] contains about 1860 images of 120 individuals, including faces with different facial expressions, illuminations, and disguises. Each image was resized to $50 \times 40$. Fig. 3 shows some images from AR.

The FERET database [22] is a benchmark database for evaluating face classification algorithms. We select a subset of FERET for experimenting that includes 200 individuals, each of which provided seven images. We used the down-sampling algorithm to resize each image to $80 \times 80$. Fig. 4 shows some normalised images from the FERET database.

The ORL database [23] is composed of 400 images from 40 different persons. Some individuals were taken images at different times, with varying lighting, facial expressions, and facial details.
We then resize each image to $46 \times 56$. Fig. 5 shows some normalised images from ORL.

### 4.1 Experiment on the AR

For AR, the first one, two, and three images per class are, respectively, selected for training. Table 1 shows the average recognition rates (%) of linear regression classification (LRC) [24], CRC [5], SRC [4], nearest neighbour (NN) [25], CIRLRC [17] and BCRC-CNN under different weights. In most cases, BCRC-CNN classifies more accurately than the other methods. For instance, when the number of training samples of each subject is three, the classification accuracies of LRC, CRC, SRC, CIRLRC and the proposed method with $w_1 = 0.6$ are 68.71, 78.94, 80.30, 76.82, and 88.94%, respectively.

To better evaluate our methods under the effect of variations in appearance in the AR database, we divide the samples from the same subject into three situations: (i) the first and eighth images are selected as the natural faces taken in two periods. (ii) the second–fourth as well as ninth–11th samples are selected for expression changes. (iii) the fifth–seventh and 12th–14th samples are selected for illumination changes. In each round of the experiment, we select the corresponding images per class for testing samples and the rest are training samples.

Fig. 6 shows the classification results by means of CIRLRC and BCRC-CNN with $w_1 = 0.6$ over the variations on the AR database. As we can see from Fig. 6a that the proposed method shows the robustness to the expression and illumination changes. Also, the recognition rates of BCRC-CNN under different situations are all above 90%. Furthermore, the great experimental result of BCRC-CNN for illumination classification indicates that the proposed bi-directional representation learning model can improve the classification performance. The running time(s) of both methods is also given in Fig. 6b. It is necessary to emphasise CIRLRC took more time to complete the process. For example, when there are illumination changes, the consuming time (s) of CIRLRC is 18,989.16 s, while the proposed method only needs 733.09 s which is 96.13% faster than the previous algorithm. Since the proposed method develops an efficient bi-directional representation model using both the forward representation and the reverse modelling based on CNN features instead of calculating the score and the distance such as CIRLRC. As analysed above, these figures strongly prove the effectiveness of BCRC-CNN.

| Method | Number of training samples |
|--------|----------------------------|
|        | 1       | 2       | 3       |
| LRC    | 72.95   | 70.63   | 68.71   |
| CRC    | 80.51   | 79.65   | 78.94   |
| SRC    | 79.87   | 80.90   | 80.30   |
| NN     | 72.95   | 70.69   | 69.70   |
| CIRLRC | 78.91   | 77.92   | 76.82   |
| BCRC_CNN ($w_1 = 0.6$) | **81.73** | 86.88 | 88.94 |
| BCRC_CNN ($w_1 = 0.7$) | 81.67 | 86.94 | 89.32 |
| BCRC_CNN ($w_1 = 0.8$) | 81.54 | **87.15** | **89.62** |

**Fig. 6** Comparisons of classification results by means of CIRLRC and BCRC-CNN with $w_1 = 0.6$ over three variations on the AR database

(a) Recognition rates, (b) Running time

We then resize each image to $46 \times 56$. Fig. 5 shows some normalised images from ORL.
CIRLRC through a different number of training samples. The classification. The strength of the novel scheme in this study lies in Meanwhile, the residuals from the forward representation and the samples, implying effectiveness and robustness of BCRC-CNN. Furthermore, it is worth stressing that the curve of BCRC- and the recognition rates of BCRC-CNN are nearly 100% in most CNN is always flat as the increase in the number of training samples to perform representation based two-step face recognition',  

**6 Acknowledgments**

This work was partially supported by the National Key Research and Development Program of China (Grant No. 2017YFC1601800, 2016YFD0401204), the National Natural Science Foundation of China (Grant No. 61876072), the Natural Science Foundation of Jiangsu Province (Grant No. BK20161135), China Postdoctoral Science Foundation (Grant No. 2018T110441, 2016M590407) and Six Talent Peaks Project in Jiangsu Province (Grant No. XYDX-012).

**7 References**

[1] Gu, G., Hou, Z., Chen, C., et al.: ‘A dimensionality reduction method based on structured sparse representation for face recognition’, Artif. Intell. Rev., 2016, 46, (4), pp. 1–13

[2] Ma, X., Zhuang, W., Li, Y., et al.: ‘Sparse representation based undersampled face recognition with shared prototype auxiliary dictionaries’, Neurocomputing, 2017, 239, pp. 58–68

[3] Chang, X.P., Zheng, Z.L., Xie, C.M.: ‘Sparse representation-based face recognition for single sample’, Comput. Eng., 2010, 36, (21), pp. 175–177

[4] Wright, J., Yang, A.Y., Ganesh, A., et al.: ‘Robust face recognition via sparse representation’, IEEE Trans. Pattern Anal. Mach. Intell., 2009, 31, (2), pp. 210–227

[5] Zhang, L., Yang, M., Feng, X.: ‘Sparse representation or collaborative representation: which helps face recognition’. Proc. IEEE Int. Conf. on Computer Vision, Barcelona, Spain, November 2011, pp. 471–478

[6] Beymer, D., Poggio, T.: ‘Face recognition from one example view’. Int. Conf. on Computer Vision (ICCV), Cambridge, USA, June 1995

[7] Vetter, T.: ‘Synthesis of novel views from a single face image’, Int. J. Comput. Vis. 1998, 28, (2), pp. 102–116

[8] Elhamifar, E., Vidal, R.: ‘Robust clustering using structured sparse representation’. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, USA, June 2011

[9] Zhuang, L.S., Yang, A.Y., Zhou, Z.H., et al.: ‘Single-sample face recognition with image corruption and misalignment via sparse illumination transfer’. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Portland, USA, June 2013

[10] Deng, W.H., Hu, J.N., Guo, J., et al.: ‘Robust, accurate and efficient face recognition from a single training image: a uniform pursuit approach’, Pattern Recognit., 2010, 43, pp. 1748–1762

[11] Deng, W., Hu, J., Guo, J.: ‘Extended SRC: undersampled face recognition via intraclass variant dictionary’, Pattern Anal. Mach. Intell., 2012, 34, pp. 1864–1870

[12] Su, M.C., Chou, C.H.: ‘Application of associative memory in human face detection’. Int. Joint Conf. on Neural Networks (IJCNN), Washington, DC, USA, July 1999

[13] Saha, S., Bandypadhyay, S.: ‘A symmetry based face detection technique’. WIE National Symp. on Emerging Technology (WioNSET), Calcutta, India, June 2007

[14] Saber, E., Tekalp, A.M.: ‘Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions’, Pattern Recognit. Lett., 1998, 19, (8), pp. 669–680

[15] Xu, Y., Li, X., Yang, J., et al.: ‘Integrate the original face image and its mirror image for face recognition’, Neurocomputing, 2014, 131, (7), pp. 191–199

[16] Xu, Y., Zhu, X., Li, Z., et al.: ‘Using the original and ‘symmetrical face’ training samples to perform representation based two-step face recognition’, Pattern Recognit., 2013, 46, (4), pp. 1151–1158

[17] Xu, Y., Li, X., Yang, J., et al.: ‘Integrating conventional and inverse representation for face recognition’, IEEE Trans. Cybern., 2014, 44, (10), pp. 1738–1746

**Table 2** Comparison of accuracies (%) and running time (s) on the FERET database

| Method         | Accuracy 2 | Time 2 | Accuracy 3 | Time 3 | Accuracy 4 | Time 4 |
|----------------|------------|--------|------------|--------|------------|--------|
| LRC            | 61.70      | 63.00  | 58.13      | 58.34  | 74.00      | 55.49  |
| CRC            | 53.30      | 57.09  | 44.00      | 72.65  | 53.67      | 76.07  |
| SRC            | 60.10      | 1033.93| 56.25      | 933.21 | 69.83      | 1420.70|
| NN             | 54.50      | 88.61  | 50.38      | 112.38 | 67.67      | 129.66 |
| TPTSR          | 58.40      | 271.10 | 50.75      | 394.68 | 60.83      | 457.18 |
| CIRLRC (w1=0.6)| 65.23      | 9382.10| 60.23      | 14921.45| 69.50      | 49985.24|
| BCRC_CNN (w1=0.7) | 93.40 | 131.66 | 93.75 | 210.69 | 93.33 | 376.34 |
| BCRC_CNN (w1=0.8) | 93.30 | 131.66 | 93.75 | 210.69 | 93.30 | 376.34 |
| BCRC_CNN (w1=0.9) | 93.60 | 131.66 | 93.75 | 210.69 | 93.83 | 376.34 |

**Fig. 7** Comparisons of recognition rates by means of BCRC-CNN with w1 = 0.6 and other classification methods on ORL

**4.2 Experiment on the FERET**

For FERET, θ (θ = 2, 3, 4) images of each class are randomly selected as training samples and treat the remaining ones for testing with respect to 200 different persons. Table 2 shows the comparison of BCRC-CNN with LRC [24], CRC [5], SRC [4], NN [25], two-phase test sample representation (TPTSR) [26] and CIRLRC [17] through a different number of training samples. The best recognition rates of BCRC-CNN are 93.60, 93.75, and 93.83% with θ (θ = 2, 3, 4) training samples. From the experimental results, we can observe our method achieved higher classification rates but consumed less running time.

**4.3 Experiment on the ORL**

For ORL, θ (θ = 1, 2, …, 9) images of each class are randomly selected as training samples and treat the remaining ones for testing with respect to 40 different persons. Fig. 7 illustrates the comparison results of BCRC-CNN and LRC [24], CRC [5], SRC [4], NN [25], CIRLRC [17] with different number of training samples. The proposed algorithm performs best among all methods and the recognition rates of BCRC-CNN are nearly 100% in most cases. Furthermore, it is worth stressing that the curve of BCRC-CNN is always flat as the increase in the number of training samples, implying effectiveness and robustness of BCRC-CNN.

**5 Conclusion**

In this study, we proposed a BCRC-CNN for face classification. The critical innovation is an efficient hybrid RBC model in the CNN feature space, which is used to accomplish face recognition. Meanwhile, the residuals from the forward representation and the reverse modelling are integrated with the minimum cost for robust classification. The strength of the novel scheme in this study lies in successfully utilizing the auxiliary residual information from the test sample, and enhancing the recognition rate, across a variety of appearance variations. The experimental results demonstrate the effectiveness of BCRC-CNN adequately. We believe that this kind of novel RBC algorithm will successfully improve representation capabilities in image recognition in the future.
[18] Bouvrie, J.: ‘Notes on convolutional neural networks’, Neural Nets, 2006, pp. 38–44

[19] Gu, J., Wang, Z., Kuen, J., et al.: ‘Recent advances in convolutional neural networks’, Comput. Sci., 2015, arXiv preprint arXiv: 1512.07108

[20] Simonyan, K., Zisserman, A.: ‘Very deep convolutional networks for large-scale image recognition’, Comput. Sci., 2014, arXiv preprint arXiv: 1409.1556

[21] Martinez, A.M.: ‘The AR face database’. CVC Technical Report 24, 1998

[22] ‘Feret data’. Available at http://www.itl.nist.gov/iad/humanid/feret/feret-master.html

[23] ‘The database of faces’. Available at http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[24] Naseem, I., Togneri, R., Bennamoun, M.: ‘Linear regression for face recognition’, IEEE Trans. Pattern Anal. Mach. Intell., 2010, 32, (11), pp. 2106–2112

[25] Kabeer, V., Thasleema, T.M., Narayanan, N.K.: ‘Face recognition using state space parameters and artificial neural network classifier’, IEEE Int. Conf. on Innovations in Information Technology, Dubai, UAE, November 2007, pp. 476–480

[26] Xu, Y., Zhang, D., Yang, J., et al.: ‘A two-phase test sample sparse representation method for use with face recognition’, IEEE Trans. Circuits Syst. Video Technol., 2011, 21, (9), pp. 1255–1262