Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 6, pp. 4967–4980, DOI 10.12694/scpe.v25i6.3406

LOCAL WEIGHTED REPRESENTATION BASED MATRIX REGRESSION CLASSIFIER AND FACE RECOGNITION

FANFENG SHI*, ZHENG FANG†, AND XINXIN WU‡

Abstract. Nuclear-norm-based matrix regression (NMR) approaches utilize the nuclear norm for error term characterization, which strengthens the robustness of algorithm. However, NMR ignores the differences between samples from different classes, which leads to a poor feature representation. Moreover, NMR does not consider variations within different class, which affects the classification performance when classes are not homogeneous. To solve above problems, a local weighted representation-based matrix regression method (LWMR) is proposed. LWMR method solves two issues of current NMR methods that are based on nuclear norm. First, LWMR utilizes the prior distance information between test and training samples as weights, which improves the inter-class separation. Second, LWMR creates a new dictionary by averaging samples within different class and choosing the best representative sample for each class, which reduces the dictionary size and complexity. Experimental results on four widely used datasets demonstrate that LWMR method has faster calculation speed and better image performance than other regression models.

Key words: Matrix regression, Nuclear norm, Data representation, Image classification.

1. Introduction. Face recognition involves utilizing facial feature information to ascertain a person's identity. This process typically encompasses three main steps: face detection, feature extraction, and face classification. Among these steps, the design of the face classifier serves as the final component of face recognition. The efficacy of the classifier directly impacts the ultimate outcome of the face recognition process[6, 1]. Linear regression analysis is a common technique for image classification [8]. Naseem *et al.* proposed a linear regression classifier (LRC) for classifying face images [15]. To avoid overfitting, different regularization terms are usually added to LR models. Two commonly used regularizers are L_1 norm regularizer and L_2 norm regularizer. Linear regression (LR) with L_2 norm regularizer is commonly called ridge regression [10], whereas LR with an L_1 norm regularizer is known as lasso regression [16]. These methods constitute widely utilized sparse representation models. J. Wright *et al* introduced a robust face recognition technique through sparse representation classification (SRC) [17]. SRC utilizes all training data as a dictionary to represent test samples, and assumes that representation coefficients are sparse. Non-zero coefficients should predominantly correspond to training samples sharing the same category label as the test sample. To improve the robustness, they additionally posited sparsity in noise and introduced an enhanced SRC model. L. Zhang *et al* analyzed the principle of SRC and considered that the collaborative representation strategy outweighs L_1 norm-based sparse constraints in importance. Consequently, they introduced ridge regression-based collaborative representation classification (CRC) [18]. Nonetheless, CRC lacks a noise removal mechanism, rendering it unsuitable as a robust image classification algorithm. In above methods, they all improve the linear regression model by adding regularization constraints, while ignoring the optimization of dictionary construction for representing test samples. J. Xu *et al.* introduced the mean representation classification (MRC) method [20], which employs the mean value of intra-class training samples as a dictionary for representing test samples. This method acquires reconstruction coefficients of each class, then assigns test samples to a class with the smallest residual. P. Huang *et al.* introduced the local mean representation-based classifier (LMRC) method [7]. This algorithm considers intra-class variations and improves the reconstruction effect of the linear regression algorithm. Above-mentioned methods are all vector-based regression methods. For two-dimensional images

*School of Information Engineering, Yangzhou Polytechnic Institute, Yangzhou Jiangsu 225127, China (Corresponding author, shiff@ypi.edu.cn)
†School of Information Engineering, Yangzhou Polytechnic Institute, Yangzhou Jiangsu 225127, China
‡School of Computer Science, Nanjing Audit University, Nanjing Jiangsu 211815, China.

presented in matrix form, they must first be converted into vectors before employing these regression models [5]. This transformation step may lose some structural information. J. Yang *et al.* introduced the nuclear norm-based matrix regression (NMR) model [21]. This model, termed matrix regression, is designed for image characterization and classification using two-dimensional image matrices. NMR employs the nuclear norm to define decision rules, rendering it more robust to illumination changes and occlusions. Based on NMR algorithm, Xie *et al.* introduced a robust NMR algorithm by incorporating a non-convex function to characterize the rank of the error image and extending it to mixed noise conditions [19]. Additionally, Deng *et al.* proposed an NN-MRPE method based on NMR, preserving embedding by constructing a graph using nuclear norm residual evaluation to project high-dimensional data into a low-dimensional space [3]. Li *et al.* proposed an improved NMR method that leverages the low-rank property of the reconstructed image, applies nuclear norm regularization to the image [11]. Chen *et al.* proposed an L_1-norm-based NMR approach. The method applies the L_1 norm to the reconstruction coefficient matrix, and enhances the robustness of the NMR method to occlusion and illumination variations [2]. Lu *et al.* introduced a novel locality-preserving projection method termed nuclear norm-based two-dimensional locality-preserving projection (NN-2DLPP), which maintains the local structure of data in a low-dimensional subspace [13]. NN-2DLPP method restores the noisy data matrix via low-rank learning, eliminates noise from data, and projects denoised data onto a new subspace. Du *et al.* introduced a method termed adaptive occlusion dictionary learning based on kernel norm for face recognition, specifically designed to handle illumination changes and occlusions [4]. To further optimize the algorithm, Luo *et al.* proposed an approximate NMR model with elastic network regularization [14]. These methods leverage the inherent structure and characteristics of face data to enhance classification performance.

Methods based on NMR boost the robustness of matrix regression algorithms by refining regularization terms and the residual assessment model. However, these methods neglect the optimization of the dictionary construction for representing test samples. NMR utilizes all training samples as the dictionary to represent test samples, without considering the intra-class or inter-class variations among training samples, which could potentially diminish the performance of the matrix regression method.

This paper introduces a novel matrix regression image classification approach termed as local weighted representation-based matrix regression (LWMR). This method is designed to improve upon the NMR technique by refining the process of dictionary construction for representing test samples. LWMR generates training samples by assigning weights based on the similarity between test and training samples, thereby enhancing inter-class distinctions. Additionally, this study addresses intra-class variabilities by constructing a dictionary through the selection of the mean value of locally optimal samples within each class as representative samples. Subsequently, the NMR method is employed to decompose test samples into a linear combination of representative samples from the dictionary. The classification outcome is then determined by identifying the minimum residual error. The LWMR method outlined in this paper offers several advantages:

1. LWMR method employs a matrix-based regression model to preserve the structural information and improve the performance of the linear regression model.
2. LWMR method enhances the discriminative ability of the matrix regression model by increasing the variability of samples between classes through weighting.
3. LWMR method exhibits strong robustness against abnormal samples within intra-class training sets.

2. Related work.

2.1. Mean represents classification. The vector of each type of training sample image is expressed as $a_i^m \in R^d, d = p \times q$, where $m = 1, 2, ..., p_i$, p_i represents the number of training samples of ith class, p and q correspond to the height and width of a image. Therefore, the entire training set is expressed as $a = [a_1, a_2, ..., a_c]$. where c represents categories of samples. Test image is expressed as $b \in R^{p \times q}$. The mean representation classification method uses the mean of each type of sample as a representative sample to describe characteristics of the test sample. Hence, a test sample is represented as the product of the class mean sample b. The coefficient vector is denoted as:

$$b = mx \tag{2.1}$$

where $x = [x_1, x_2, ..., x_c]^T$ is the reconstruction coefficient vector corresponding to the representative sample. $m = [m_1, m_2, ..., m_c]$ denotes a dictionary representing sample composition, where each entry corresponds to

the mean sample of a class. The mean sample of the ith class is expressed as:

$$m_i = \frac{1}{p_i} \sum_{j=1}^{p_i} a_i^j, i = 1, 2, \ldots, c \tag{2.2}$$

The mean representation classification method uses the least sqaure method to solve for reconstruction coefficients:

$$\hat{x} = \left(m^T m\right)^{-1} m^T b \tag{2.3}$$

Calculate the reconstruction residual corresponding to the test sample:

$$e_i(b) = \|b - m_i \hat{x_i}\|_2 \tag{2.4}$$

where $\hat{x_i}$ represents the reconstruction coefficient vector of ith class. Finally, according to the principle of minimum reconstruction residual distance, the category of a test sample can be determined:

$$\text{identity } (b) = \arg\min_i \{e_i\} \tag{2.5}$$

2.2. Matrix regression classification. The matrix regression classification method utilizes all training samples to characterize test samples, and uses $A_1, A_2, ..., A_n$ as dictionaries to linearly represent test samples. Therefore, the test sample B is expressed as:

$$\begin{gathered} B = A(x) + E \\ A(x) = x_1 A_1 + x_2 A_2 + \ldots, + x_n A_n \end{gathered} \tag{2.6}$$

where $x = [x_1, x_2, ..., x_n]$ is the coefficient vector corresponding to training samples. E is the error matrix, which is usually low-rank in a optimal solution. Therefore, NMR calculates x by solving the subsequent optimization problem:

$$\min_x \operatorname{rank}(A(x) - B) \tag{2.7}$$

To enhance the generalization ability, NMR introduced ridge regression to the above model, added a regularization term to represent the coefficient x, and obtained a new matrix regression model:

$$\min_x \|A(x) - B\|_* + \frac{1}{2}\lambda \|x\|_2^2 \tag{2.8}$$

In the equation, λ is a positive constant. To optimize the model, NMR utilizes the nuclear norm $\|\cdot\|_*$ substitution $\operatorname{rank}(\cdot)$ function and avoids overfitting of the model by introducing regular terms.

3. Matrix regression classification with locally weighted representation.

3.1. motivation. NMR utilizes all training samples to construct a dictionary. The dictionary is utilized to represent test samples. Training samples are directly employed to represent test samples without distinction. The dictionary constructed directly from all training samples ignores the problem of insufficient differences between samples of different classes. When samples with large differences are in samples within a class, the classification performance of the model will be poor.

To enhance the performance, reconstructing the dictionary used for representing test samples has become a widely adopted and effective approach. MRC simplifies this process by employing the mean value of training samples from each class as the representative sample to construct a new dictionary. This method of reconstructing the dictionary by means of representative samples ignores the possibility that intra-class samples may appear as outliers, thus greatly reducing the accuracy and recognition rate of the algorithm. Therefore, building an effective dictionary is an important way to improve model performance.

Increasing the difference between classes can better improve the competitiveness between samples, which enhances the feature expression capability. In this paper, a weighting function based on NMR is designed,

which weights the similarity between training and test samples, and increases the difference between samples by weighting all training samples. To enhance the robustness in intra-class samples, the local mean processing method can not only reduce the impact of outliers on model performance, but also reduce parameters.

Inspired by aforementioned ideas, a matrix regression classification model based on local weighted representation is proposed. This model uses the similarity between training and test samples as weights. After weighting all training samples, the model calculates the local mean of training samples within each class to form a representative sample of each class. Finally, representative samples of each class are combined into a new dictionary for representing test samples. The model in this paper simultaneously considers the problem of insufficient differentiation of samples between classes and excessive variability of samples within classes, which improves the classification performance and robustness of the model.

3.2. Local Weighted Representation Dictionary. The motivation behind local weighted representation stems from the recognition of the significance of considering the local structure and relationships within data samples, particularly in image classification tasks. Traditional representation methods often treat all samples equally when constructing dictionaries or representations, disregarding potential variations and similarities among samples. Local weighted representation seeks to address this limitation by assigning weights based on the similarity between test and training samples. The construction of the local weighted representation dictionary is based on the distance Dis(a, b) between all training samples a and a test sample b as the similarity. The similarity matrix between training samples and a test sample is defined as:

$$S = D\left(a_i^j - b\right) \tag{3.1}$$

where the similarity matrix S is expressed as :

$$S = \begin{bmatrix} D\left(a_1^1 - b\right) & D\left(a_1^2 - b\right) & \cdots & D\left(a_1^{p_1} - b\right) \\ D\left(a_2^1 - b\right) & D\left(a_2^2 - b\right) & \cdots & D\left(a_2^{p_2} - b\right) \\ \vdots & \vdots & \ddots & \vdots \\ D\left(a_c^1 - b\right) & D\left(a_c^2 - b\right) & \cdots & D\left(a_c^{p_c} - b\right) \end{bmatrix} \tag{3.2}$$

where the distance between the test sample and all training samples is measured by Euclidean distance:

$$D\left(a_i^j - b\right) = \left\|a_i^j - b\right\|_2, i = 1, 2, \ldots, c; j = 1, 2, \ldots, p_i \tag{3.3}$$

Weights are assigned based on the similarity between all training samples and a test sample, and a weighted representation dictionary is constructed by adding weight to training samples:

$$W = \begin{bmatrix} \frac{1}{S_1^1} * a_1^1 & \frac{1}{S_1^2} * a_1^2 & \cdots & \frac{1}{s_1^{p_1}} * a_1^{p_1} \\ \frac{1}{S_2^1} * a_2^1 & \frac{1}{S_2^2} * a_2^2 & \cdots & \frac{1}{s_2^{p_2}} * a_2^{p_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{S_c^1} * a_c^1 & \frac{1}{S_c^2} * a_c^2 & \cdots & \frac{1}{s_c^{p_c}} * a_c^{p_c} \end{bmatrix} \tag{3.4}$$

where S_i^j represents the similarity between the jth sample a_i^j to be tested and training samples in the ith category. A weighted representation sample for each training sample is expressed as $\mathrm{w}\frac{1}{S_i^j} * a_i^j$.

Constructing a weighted representation dictionary through similarity weighting can increase the difference between samples of different class, enhance the weight of samples with smaller similarities in the reconstructed representation, and improve the classification effect of the regression model. However, when outliers appear in intra-class samples, it will affect the regression reconstruction performance of the model. Based on the weighted representation dictionary, the robustness of the model for the case of excessive intra-class sample differences is improved by constructing a locally weighted representation dictionary. After ranking the similarity corresponding to each class of training samples, the corresponding nearest neighbor parameter K is selected and the K-nearest neighbor matrix is constructed. Specific steps are shown below:

Step 1. Sort each row of the similarity matrix by similarity and take the smallest similarity value to construct a nearest neighbor similarity matrix, as follows.

$$S_K = \begin{bmatrix} S_1(1) & S_1(2) & \cdots & S_1(K) \\ S_2(1) & S_2(2) & \cdots & S_2(K) \\ \vdots & \vdots & \ddots & \vdots \\ S_c(1) & S_c(2) & \ldots & S_c(K) \end{bmatrix} \tag{3.5}$$

Among them, the value $S_i(j)$ representing the jth similarity of the ith category contains the K smallest similarity value in each category.

Step 2. Find the position of each similarity in the K-nearest neighbor matrix in the corresponding training sample category, denoted by the function $f(S_K)$, to obtain the position matrix P of the similarity matrix S_K, which is defined as:

$$P = \begin{bmatrix} f(S_1(1)) & f(S_1(2)) & \cdots & f(S_1(K)) \\ f(S_2(1)) & f(S_2(2)) & \cdots & f(S_2(K)) \\ \vdots & \vdots & \ddots & \vdots \\ f(S_c(1)) & f(S_c(2)) & \ldots & f(S_c(K)) \end{bmatrix} \tag{3.6}$$

In the equation, $P_i{}^j = f(S_i(j))$ represents the jth position of the sorted similarity value in the ith category in the weighted representation dictionary.

Step 3. Find the weighted training samples of the K nearest neighbours of each category in the weighted representation dictionary W as a locally weighted dictionary l. The local weighted dictionary is expressed as:

$$P = \begin{bmatrix} W_1^{P_1^1} & W_1^{P_1^2} & \cdots & W_1^{P_1^K} \\ W_2^{P_2^1} & W_2^{P_2^2} & \cdots & W_2^{P_2^K} \\ \vdots & \vdots & \ddots & \vdots \\ W_c^{P_c^1} & W_c^{P_c^2} & \cdots & W_c^{P_c^K} \end{bmatrix} \tag{3.7}$$

where $W_i{}^j$ represents the weighted representation sample with the smallest corresponding similarity among the weighted representation samples.

Step 4. In order to reduce the problem that excessive intra-class sample differences will lead to poor model robustness. By taking the mean value within the class, the class representative samples are constructed by averaging the weighted representation samples of each class in the local weighted dictionary, and the local weighted representative dictionary is constructed. The local weighted representative dictionary is expressed as:

$$r = \begin{bmatrix} \text{mean}\left(W_1^{P_1^1}, W_1^{P_1^2}, \ldots, W_1^{P_1^K}\right) \\ \text{mean}\left(W_2^{P_2^1}, W_2^{P_2^2}, \ldots, W_2^{P_2^K}\right) \\ \vdots \\ \text{mean}\left(W_c^{P_c^1}, W_c^{P_c^2}, \ldots, W_c^{P_c^K}\right) \end{bmatrix} \tag{3.8}$$

where the function mean$(\cdot)$ means taking the average of vectors in the brackets to construct a new representative sample.

3.3. Matrix Regression Classification with Locally Weighted Representation. For each test sample, the local weighted representative sample of each class can be calculated through the above method, and representative samples are combined to build a local weighted representative dictionary $r = [r_1, r_2, \ldots, r_c] \in R^{d \times c}$. Use a local weighted dictionary r to linearly represent the test sample b, as follows:

$$\begin{gathered} b = r(x) + e \\ r(x) = x_1 r_7 + x_2 r_2 + \ldots + x_c r_c \end{gathered} \tag{3.9}$$

where, $x_1, x_2, \cdots, x_c$ represent the coefficient corresponding to the representative sample, and e represents the residual. Considering that the residual image $r(x) - b$ usually required in the optimization process is of low rank, the representation coefficient is learned through low-rank constraints:

$$\min_x \operatorname{rank}(r(x) - b) \tag{3.10}$$

The above optimization problem is an NP problem, which is usually converted into a nuclear norm problem to solve, as follows:

$$\min_x \|(r(x) - b)\|_* \tag{3.11}$$

In addition, the idea of ridge regression is introduced into the formula, and a regularized matrix regression model is obtained by adding regular terms:

$$\min_x \|r(x) - b\|_* + \frac{1}{2}\lambda\|x\|_2^2 \tag{3.12}$$

By applying the augmented Lagrange multiplier method, the model is rewritten as follows:

$$\min_x \|y\|_* + \frac{1}{2}\lambda\|x\|_2^2 \quad \text{s.t. } r(x) - b = y \tag{3.13}$$

The corresponding Lagrangian function is defined as:

$$L(y, x, z) = \|y\|_* + \tfrac{1}{2}\lambda\|x\|_2^2 + \operatorname{tr}\left(z^T(r(x) - y - b)\right) + \tfrac{\mu}{2}\|r(x) - y - b\|_F^2 \tag{3.14}$$

where $\mu > 0$ is a penalty parameter, z is the Lagrange multiplier, $\operatorname{tr}(\cdot)$ representing the trace operation function. Through iterative solution, the solution formula is as follows:

$$x^{k+1} = \left(r^T r + \frac{\lambda}{\mu} I\right)^{-1} r^T \left(b + y^k - \frac{1}{\mu} z^k\right) \tag{3.15}$$

$$y^{k+1} = D_{\frac{1}{\mu}}\left(r\left(x^{k+1}\right) - b + \frac{1}{\mu} z^k\right) \tag{3.16}$$

$$z^{k+1} = z^k + \mu\left(r\left(x^{k+1} - y^{k+1} - b\right)\right) \tag{3.17}$$

After calculating the representation coefficient x^*, calculate the reconstruction residual corresponding to the test sample:

$$e_i(b) = \left\|\hat{b} - \hat{b_i}\right\|_* = \left\|r\left(x^*\right) - r\left(\delta_i\left(x^*\right)\right)\right\|_* \tag{3.18}$$

where $\delta_i(x)$ retains the representation coefficients related to the current class and sets other coefficients to zero. Finally, according to the principle of minimum reconstruction residual distance, the category of the test sample is determined, that is:

$$\text{identity } (b) = \arg\min_i \{e_i\} \tag{3.19}$$

The flow of the algorithm is as follows:

4. Experimental results and analysis. Experiments to assess the performance of algorithms are presented in this chapter. Two benchmark datasets of face images and biometric fingerprints are used to test LWMR algorithm Some classical classification algorithms, such as SRC, LRC, and CRC are used to compared with LWMR method. The algorithm are also contrasted with the state-of-the-art NMR and SR-NMR matrix regression classification algorithms from the relevant literature. To make results more credible, we repeat each experiment 10 times. We conduct all experiments on MATLAB 2021b platform, using a machine with an Intel i7-10750 2.60GHz CPU, an NVIDIA GTX1650 GPU, and 16GB 2933MHz memory.

Algorithm 1 Matrix regression classification algorithm based on local weighted representation

Input: Training set, Test set, Neighbor parameter K, regularization parameter λ, penalty parameter μ, Lagrange multiplier z, maximum number of iterations.

Output: The category of a test sample

1: Preprocess samples to obtain training set $a = [a_1, a_2, \cdots, a_c]$ and test sample b.
2: Calculate the similarity matrix S between training samples and the test sample.
3: The similarity between training samples and the test sample serves as a weight, and a weighted representation dictionary W is constructed by adding weight to training samples.
4: Find weighted training samples of the K-nearest neighbors of each category in the weighted representation dictionary as a local weighted dictionary l.
5: By taking the mean value within the class, the class representative samples are constructed by averaging weighted representation samples of each class in the local weighted dictionary l, and the local weighted representative dictionary r is constructed.
6: Utilize a locally weighted dictionary r to linearly represent b, and solve for representation coefficients x through matrix regression.
7: Classify according to the minimum residual principle identity $(b) = \arg\min_i \{e_i\}$

Fig. 4.1: Images from Extended Yale B dataset.

4.1. Light change experiment. Image classification accuracy may be affected by noise data in images collected under varying illumination conditions. To assess the robustness of the algorithm to lighting changes, the Extended Yale B dataset is utilized [12]. This dataset contains 38 people, each acquiring 64 images.All images are converted to grayscale and are resized to 96×84 pixels. 10, 15, 20, or 25 images per subject are selected randomly as training samples and rest samples are used to test. Some sample images are shown in Fig. 4.1.

The recognition results of different methods on Extended Yale B dataset are shown in Table 4.1. As shown in Fig. 4.1, this dataset contains many images with severe lighting conditions, which will generate more noise and have higher requirements on the robustness of the recognition algorithm. From Table 4.1, it is clearly that the sparse representation method SRC has a poor recognition rate and is not robust to images with large illumination changes. Both LRC and CRC methods utilize all training samples as a dictionary. However, the inclusion of noisy images in the dictionary prevents the generation of better reconstructed samples, thus impeding the enhancement of recognition results. Different from the classic algorithm, the matrix regression algorithm NMR takes into account the structural information, introduces the nuclear norm to describe the error term, improves the robustness to structural noise, and obtains a higher recognition rate. The proposed method

Table 4.1: Results of different algorithms on Extended Yale B

Methods	Different numbers of training samples			
	10	15	20	25
SRC	52.97%	60.61%	67.01%	71.10%
LRC	71.65%	82.04%	87.57%	90.00%
CRC	85.64%	90.11%	92.38%	94.54%
NMR	90.35%	94.59%	96.68%	97.73%
LWMR	91.93%	95.62%	97.41%	98.35%

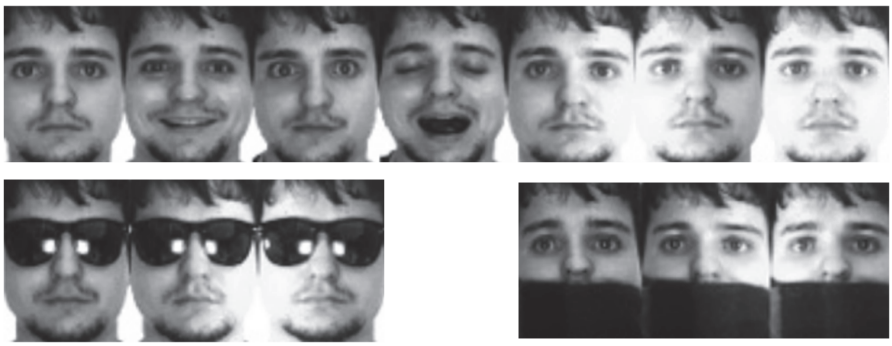


Fig. 4.2: Images from AR dataset.

reduces the impact of noisy samples by selecting dictionary samples and obtains the highest recognition rate.

4.2. Real occlusion experiment. To assess the recognition performance in occluded scenes, AR dataset is utilized. AR dataset [9] contains 120 volunteers, each with 26 images, including different expressions, angles, occlusions, etc., totaling 3120 images. All images are downsampled to 165×120 pixels. Some samples of the dataset are shown in Fig. 4.2.

Firstly, we selected 8 random frontal face images from the unoccluded part of the dataset as training samples. We divide the occlusion part of AR face dataset into two test sets: sunglasses occlusion and scarf occlusion. Each test set contains 6 occlusion images of each type of sample. The classification results of different classification algorithms under real occlusion and scarf occlusion by sunglasses in AR dataset are shown in Fig. 4.3 and Fig. 4.4.

We can see from the histogram in Fig. 4.3 that LWMR algorithm achieves the highest recognition effect in real scenes blocked by sunglasses. For scenes blocked by sunglasses, there are fewer parts of the face image that are continuously blocked, and only a small amount of eyes are blocked. Therefore, the traditional SRC algorithm, LRC algorithm and CRC algorithm have achieved good classification results. NMR and LWMR algorithms do not have obvious advantages in classification under continuous occlusion. However, compared with traditional regression algorithms, matrix regression fully utilizes the two-dimensional structural information and achieve better performance. From Fig. 4.2, the face image is blocked by a large scarf, which seriously affects the face image and reduces the effectiveness of the face image. Fig. 4.4 shows results of face images of different methods in scarf occlusion scenes. It can be seen from the figure that traditional regression algorithms: SRC algorithm, LRC algorithm and CRC algorithm have poor recognition results in scenes with large-area continuous occlusion. Recognition rate below 55%. The algorithm based on matrix regression achieves a higher recognition rate. By retaining the two-dimensional structural information, the performance is improved by imposing low-rank constraints on the error image. LWMR algorithm in this article further improves the recognition rate of the algorithm by constructing a more effective dictionary and enhances the effectiveness in different scenarios.

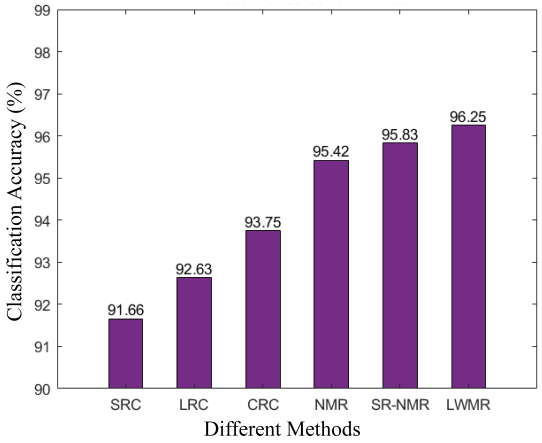


Fig. 4.3: Recognition rate under sunglasses occlusion in AR dataset.

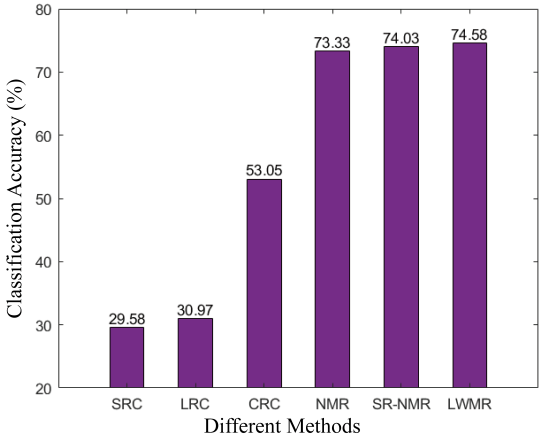


Fig. 4.4: Recognition rate under scarf occlusion in AR dataset.

4.3. Random occlusion experiment. To further assess the effectiveness of LWMR, we conducted experiments using three subsets of the Extended Yale B dataset. Subset 1 and subset 2 served as the training set, and subset 3 functioned as the test set. We added different degrees of occlusion to the test set, using baboon images to randomly occlude test images. Occlusion is divided into 5 levels, accounting for 10%, 20%, 30%, 40%, 50%, and 60% of overall images. The image after adding occlusion is shown in Fig. 4.5.

Fig. 4.6 shows results of five comparative methods and LWMR method under different degrees of occlusion. As the occlusion ratio increases, the recognition rate of each algorithm gradually decreases. As is seen from Fig. 4.6, LWMR method obtains the best performance and has good robustness to continuous occlusion scenes. SRC algorithm, LRC algorithm and CRC algorithm are classic regressions. As the degree of occlusion increases, the recognition rate of the algorithm decreases significantly, because as the degree of occlusion increases, more and more key information of the image will be lost, which affects the recognition algorithm. The latest SR_NMR algorithm achieves a good recognition rate because the algorithm increases the sparsity constraint of matrix regression and improves the robustness to noise.

4.4. Experiment analysis. From experimental results, the performance, advantages and disadvantages of different algorithms can be analyzed. This article selects three classic datasets from different scenes to

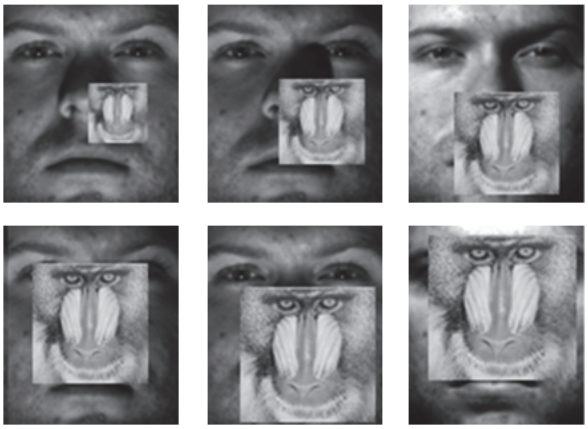


Fig. 4.5: Face images with different occlusion ratios.

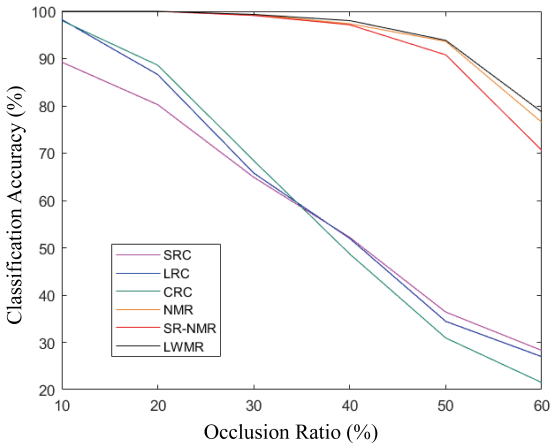
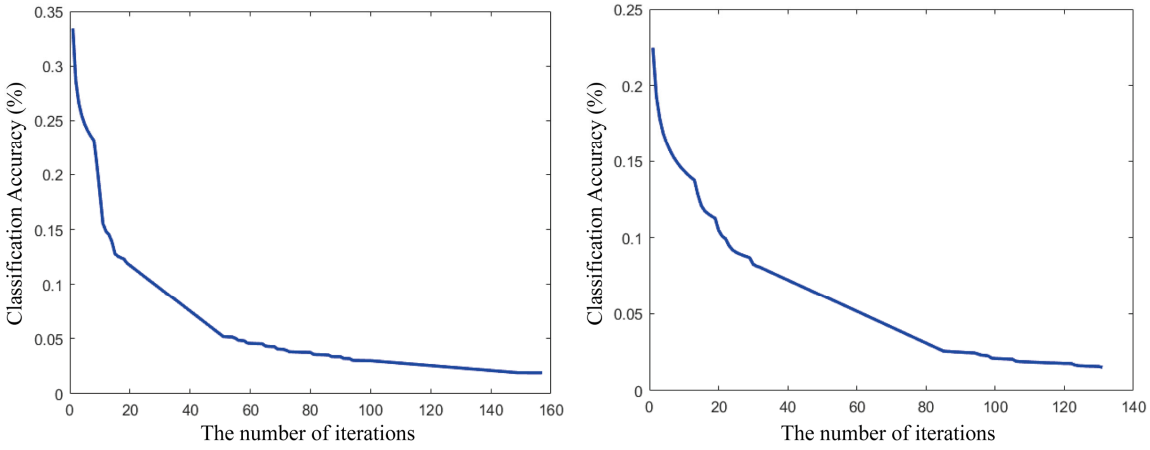


Fig. 4.6: Classification accuracies under different occlusion ratios.

evaluate the performance of the different method, including Yale B dataset with severe illumination changes, AR dataset with different degrees of occlusion in real scenes, and the dataset with random block occlusion. LWMR has achieved good recognition performance on three datasets, indicating that LWMR has good robustness to different types of noise.

Moreover, Yale B dataset was employed to assess variations in recognition performance among different algorithms under various lighting conditions. As shown in Table 4.1, different illumination will have a great impact on the recognition rate, especially in scenes with severe illumination changes. Traditional SRC algorithms, LRC algorithms and CRC algorithms lack robustness to noise and do not have good identification and discrimination capabilities for large changes in illumination. NMR algorithm and SR-NMR algorithm fully take into account the structural information. Moreover, they enhance the recognition performance effectively through the application of low-rank constraints on the error image. This paper further processes the noise data based on matrix regression, and constructs a robust dictionary by weighting training samples and selecting them locally, thereby reducing the impact of noise samples on the recognition rate. The experimental data shows that LWMR algorithm can perform well in scenes with severe illumination changes. It has better recognition effect.



(a) Extended Yale B (b) AR

Fig. 4.7: Convergence on different datasets.

Occluded images are used to test the effectiveness and robustness of face recognition algorithms. AR occlusion face images from realistic scenarios are utilized used to compare the capability of different methods. Fig. 4.3 shows the histogram of recognition rates for different algorithms. However, as shown in Fig. 4.4, the recognition performance of three classic algorithms declined sharply due to different scarf occlusions. The continuous large-area occlusion obscured some essential facial features, leading to low recognition accuracy. By incorporating the structural information of the image, NMR and SR-NMR algorithms perform low-rank decomposition on the error image to achieve higher recognition rates. By constructing a more effective dictionary that preserves the structural information, the algorithm presented in this paper enhances the noise robustness of matrix regression.

To assess the performance of continuous occlusion, we conducted random occlusion experiments with varying proportions. As shown in Fig. 4.6, the figure shows that results of different algorithms declines as the face image occlusion proportion increases. The recognition rate of the classic SRC, LRC, and CRC algorithms declines rapidly with the proportion of continuous occlusions, which greatly affects their performance. The matrix regression algorithm, on the other hand, has a slower decrease and better robustness as the occlusion ratio increases.As shown in Fig. 4.6, the LWMR algorithm presented in this study consistently demonstrates superior recognition rates compared to alternative methods. This achievement is attributed to the algorithm's integration of local weights into training samples based on their structural information. This strategy enables the LWMR algorithm to construct a more resilient dictionary, thereby enhancing its overall robustness.

4.5. Convergence analysis. The convergence of the model is important to verify the quality of a model. ADMM algorithm [22] can help the model iterate to the optimal value faster and more effectively. When the loss of the model decreases, the number of iterations increases. It indicates that the model has better performance. This paper uses the ADMM optimization algorithm to solve the model iteratively, and examined the model convergence in the experiment to validate the stability of the face recognition model. If the decline is slow at the beginning, it means that the learning rate of the model is set relatively low and the learning gradient is relatively slow. If the learning rate of a model is set too high, the global optimal value will not be learned, and the loss will fluctuate and not decrease.

We used the loss as a criterion for model convergence and the number of iterations as an indicator of model effectiveness. Fig. 4.7 illustrates the convergence of different datasets. Fig. 4.7 (a) and (b) demonstrate that the algorithm converged well on both two datasets, learning effective parameters that improved the face image classification.

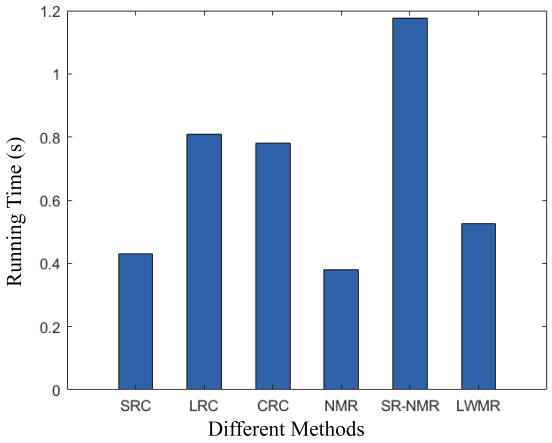


Fig. 4.8: Running times of different algorithms.

4.6. Time complexity analysis. To validate the performance of the LWMR model presented in this paper, we compared and analyzed its running time, and further examined its time complexity. We used AR face dataset as the benchmark to measure the running time of different algorithms.

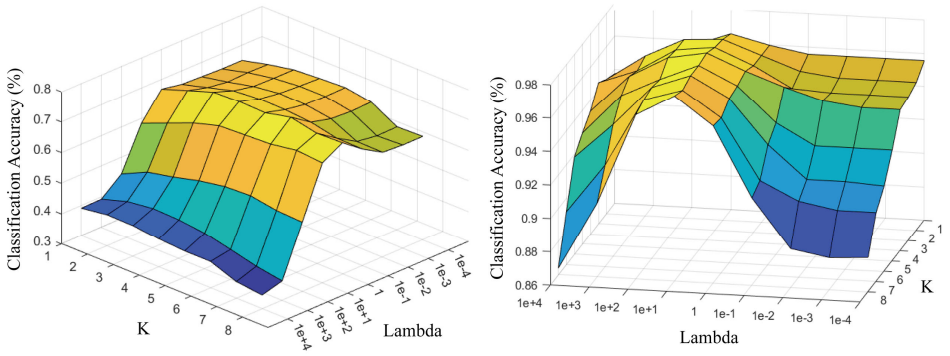
Fig. 4.8 shows that LRC algorithm and CRC algorithm have the lowest time complexity and the fastest running time, as they only involve linear regression models with closed-form solutions. However, these two algorithms have low recognition rates and poor noise robustness in complex scenarios such as extensive continuous occlusions.

LWMR algorithm presented in this paper outperforms other algorithms in terms of speed and recognition rate for scenarios with severe illumination variations and extensive continuous occlusions. SRC algorithm is somewhat robust, but it has high time complexity and lower recognition rate for such scenarios than LWMR algorithm. The complexity of LWMR is $O\left(k\left(m^2\right)+mn\right)$. The complexity of SRC is $O\left(n^2\left(m+n\right)^2\right)$. LWMR algorithm proposed in this paper exhibits low time complexity and high noise robustness across various scenarios.

4.7. Parameter sensitivity analysis. In the experiment, the regularization parameter and the local selection parameter K were used in the optimization formula of the model. Fig. 4.8 shows the impact of parameters λ and K. This paper sets parameters in the range of $\{1e-4, 1e-3, 1e-2, 1e-1, 1, 1e+1, 1e+2, 1e+3, 1e+4\}$ to conduct parameter sensitivity analysis. Local selection parameters are set within the data interval of $\{1, 2...train-size\}$, and the model recognition accuracy is used as the measurement unit for analysis.

The experiment shows that the performance of the model has a non-monotonic relationship with the parameters. The model achieves the highest accuracy when the value is 1. This indicates that weights of two models before and after regularization are equal, and only minor adjustments of regularization parameters are needed. Optimal experimental results were obtained when the regularization parameter was set to 0.9. As shown in Fig. 4.9, when the K increases, the recognition rate of the model gradually increases, but when it reaches the maximum value, the recognition rate decreases. A common practice is to select two numbers before the maximum value as optimal identification parameters. Some of training samples are negative samples. By assigning weights to samples and eliminating a few negative samples, we can boost the performance and the robustness.

5. Conclusion. This paper presents LWMR, a matrix regression classification method designed to enhance discriminative capability. LWMR takes into account the variability among samples, avoids the masked part by the mean operation. Additionally, LWMR treats each sample in a weighted way, and takes into account the variability of the samples within the class, so it achieves a better recognition rate. Evaluation on three datasets validates the superior performance of LWMR over existing matrix regression methods, particularly evident in scenarios featuring severe illumination variations and continuous occlusions. In the further work, we



(a) AR-scarf (b) AR-sunglass

Fig. 4.9: Accuracies with different lamda parameters.

will explore difference weighting scheme and apply to more datasets.

REFERENCES

[1] F. Boutros, M. Fang, M. Klemt, B. Fu, and N. Damer, *Cr-fiqa: face image quality assessment by learning sample relative classifiability*, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 5836–5845.

[2] Z. Chen, X.-J. Wu, and J. Kittler, *A sparse regularized nuclear norm based matrix regression for face recognition with contiguous occlusion*, Pattern recognition letters, 125 (2019), pp. 494–499.

[3] Y.-J. Deng, H.-C. Li, Q. Wang, and Q. Du, *Nuclear norm-based matrix regression preserving embedding for face recognition*, Neurocomputing, 311 (2018), pp. 279–290.

[4] L. Du and H. Hu, *Nuclear norm based adapted occlusion dictionary learning for face recognition with occlusion and illumination changes*, Neurocomputing, 340 (2019), pp. 133–144.

[5] P. Filzmoser and K. Nordhausen, *Robust linear regression for high-dimensional data: An overview*, Wiley Interdisciplinary Reviews: Computational Statistics, 13 (2021), p. e1524.

[6] H.-H. Huang, F. Yu, X. Fan, and T. Zhang, *A framework of regularized low-rank matrix models for regression and classification*, Statistics and Computing, 34 (2024), p. 10.

[7] P. Huang, C. Qian, G. Yang, and J. Hua, *Local mean representation based classifier and its applications for data classification*, International Journal of Machine Learning and Cybernetics, 9 (2018), pp. 969–978.

[8] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, *Linear regression*, in An introduction to statistical learning: With applications in python, Springer, 2023, pp. 69–134.

[9] A. Kumar, M. Kumar, and A. Kaur, *Face detection in still images under occlusion and non-uniform illumination*, Multimedia Tools and Applications, 80 (2021), pp. 14565–14590.

[10] T. D. la Tour, M. Eickenberg, A. O. Nunez-Elizalde, and J. L. Gallant, *Feature-space selection with banded ridge regression*, NeuroImage, 264 (2022), p. 119728.

[11] Q. Li, H. He, H. Lai, T. Cai, Q. Wang, and Q. Gao, *Enhanced nuclear norm based matrix regression for occluded face recognition*, Pattern Recognition, 126 (2022), p. 108585.

[12] Y. Lin, Z. Lai, J. Zhou, J. Wen, and H. Kong, *Multiview jointly sparse discriminant common subspace learning*, Pattern Recognition, 138 (2023), p. 109342.

[13] Y. Lu, C. Yuan, Z. Lai, X. Li, W. K. Wong, and D. Zhang, *Nuclear norm-based 2dlpp for image classification*, IEEE Transactions on Multimedia, 19 (2017), pp. 2391–2403.

[14] L. Luo, Q. Tu, J. Yang, and J. Yang, *An adaptive line search scheme for approximated nuclear norm based matrix regression*, Neurocomputing, 289 (2018), pp. 23–31.

[15] I. Naseem, R. Togneri, and M. Bennamoun, *Linear regression for face recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32 (2010), pp. 2106–2112.

[16] L. Nguyen, D. K. Nguyen, T. Nguyen, B. Nguyen, and T. X. Nghiem, *Analysis of microalgal density estimation by using lasso and image texture features*, Sensors, 23 (2023), p. 2543.

[17] J. Peng, W. Sun, H.-C. Li, W. Li, X. Meng, C. Ge, and Q. Du, *Low-rank and sparse representation for hyperspectral image processing: A review*, IEEE Geoscience and Remote Sensing Magazine, 10 (2021), pp. 10–43.

[18] H. Su, Y. Yu, Q. Du, and P. Du, *Ensemble learning for hyperspectral image classification using tangent collaborative representation*, IEEE Transactions on Geoscience and Remote Sensing, 58 (2020), pp. 3778–3790.

[19] J. Xie, J. Yang, J. J. Qian, Y. Tai, and H. M. Zhang, *Robust nuclear norm-based matrix regression with applications to*

robust face recognition, IEEE Transactions on Image Processing, 26 (2017), pp. 2286–2295.
[20] J. Xu and J. Yang, *Mean representation based classifier with its applications*, Electronics letters, 47 (2011), pp. 1024–1026.
[21] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, *Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (2016), pp. 156–171.
[22] H. Zhang, J. Gao, J. Qian, J. Yang, C. Xu, and B. Zhang, *Linear regression problem relaxations solved by nonconvex admm with convergence analysis*, IEEE Transactions on Circuits and Systems for Video Technology, (2023).

Edited by: Jingsha He
Special issue on: Efficient Scalable Computing based on IoT and Cloud Computing
Received: Feb 28, 2024
Accepted: Apr 7, 2024