UNCORRELATED DISCRIMINATIVE LOW-RANK PRESERVING PROJECTION FOR DIMENSIONALITY REDUCTION

Ummadi Sathish Kumar*
Research Scholar, Department of Computer Science and Engineering, Acharya Nagarjuna University Guntur, Andhra Pradesh, India 522510,

Edara Srinivasa Reddy
Professor, Department of Computer Science and Engineering, Acharya Nagarjuna University Guntur, Andhra Pradesh, India
*Corresponding Author

ABSTRACT

Low-rank representation (LRR) is a tool which is used to ensure the inherent representation of the noticed samples. Even LRR make an effective representation, but which cannot make better classification and which is weak while handling new samples and which cannot acquire projection matrix in the training period. A novel uncorrelated low rank preserving projection (UDLRPP) is introduced as dimension reduction algorithm by assimilating the un-correlation constraint and the internal region relationship of the original samples into the low rank representation (LRR). The role of UDLRPP is, the LRR can ensnare the overall structure data and the internal geometrical structure data is parallelly preserved by manifold regularization term. The low-rank representation coefficients are used to acquire the constrained term is instigated by the adaptive graph. And by adding the un-correlation analysis constraint expression, UDLRPP can attain the foremost projection matrix to enhance the classification accuracy. Several investigations completed on different public image datasets and declared that the proposed UDLRPP can acquire improved recognition rate compared with the state-of-the-art feature extraction methods.

Key words: Low-rank representation, uncorrelated feature, discriminant analysis adaptive graph, feature extraction.

http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=10
**1. INTRODUCTION**

The high dimensionality in big data needs a heavy computation when the analysis needed. The number of input variables or features for a dataset is referred to as its dimensionality. Dimensionality reduction refers to techniques that reduce the number of input variables in a dataset. The fields including data mining, machine learning and computer vision to name just a few. It is often a necessary pre-processing step in many systems, usually employed for simplification of the data and noise reduction, in terms of reducing the number of random variables under consideration and it obtains a set of principal variables. It can be divided into feature selection and feature extraction. In feature selection, we try to find a subset of the original features set, to get a smaller subset that can be used to model the problem. It usually involves filter, wrap and embedding methods. In the extraction of feature, it reduces the data in a space of high dimensional space to a lower dimension space, i.e. a space with less number of dimensions. The goal of dimensionality reduction is to map the high-dimensional samples to a lower dimensional space such that certain properties are preserved. Usually, the property that is preserved is quantified by an objective function and the dimensionality reduction problem is formulated as an optimization problem. To solve this problem, many methods are proposed for dimensionality reduction. Principal component analysis (PCA) aims at preserving the global variance with the minimum reconstruction error by projecting data on a linear subspace spanned by principal component vectors. Linear discriminant analysis (LDA) searches the projection axes on which data from the same class are as close as to each other and requires data points from different classes are as far as possible. LDA has the over-fitting problem. However, such technique only finds the overall structure information of the noticed data for dimensionality reduction. A novel uncorrelated discriminative low-rank preserving projection (UDLRPP) algorithm is proposed for dimensionality reduction. Overview of this paper is summarized as given below.

1. Low-rank representation, uncorrelated discriminative projection and manifold learning are merged into a unified model.
2. The lowest rank representation of the original data is computed in low-dimensional projection space, which saves a lot of time.
3. An iterative algorithm is designed to update the projection matrix $P$ and low-representation coefficient matrix $K$.
4. By adding the discriminative constraint item, the features extracted are more discriminative. And using the uncorrelated constraint item, non-redundant features are obtained.
5. The low rankness of the acquired samples in low dimensional space is used to construct an informative graph for getting local manifold structure.

The rest of the paper is organized as follows. In Section 2, briefly review the background and related work. In Section 3, LRR and its variants are introduced. In Section 4, Discriminative LRPP explained. In Section 5, the proposed UDLRPP method explained. In Section 6, Extensive experiments are performed to evaluate UDLRPP algorithm. Section 7 gives the conclusion and future work.

**2. RELATED WORK**

Dimensionality reduction techniques have been widely used to represent the raw data in a compact way without losing too much useful information [1]. These techniques learn a lower dimensional subspace to represent the face such that the image analysis can be performed more efficiently. Principal Component Analysis (PCA) [2] and Linear Discriminant Analysis (LDA) [3] are two famous linear algorithms for unsupervised and supervised dimensionality reduction.
respectively, which have been widely studied and extensively used in many fields such as computer vision, pattern recognition and other biometrics [4]–[8].

Wang et al. [9] proposed a robust Compact Fisher Vector (CFV) descriptor for robust face recognition. They preserve the global structure of the data by using fewer dimensions. However, the world is not always flat, linear dimensionality reduction techniques cannot adequately reflect the nonlinear structure of world [10]. A series of nonlinear dimensionality reduction algorithms have been developed to solve this problem, with two attracting a wide range of attention particularly manifold based techniques and kernel based techniques. Isometric feature mapping (ISOMAP) [11], Laplacian Eigen maps (LE) [12], Local Linear Embedding (LLE) [13] and Local Tangent Space Alignment (LTSA) [14] are the most well-known manifold based algorithms. Locality Preserving Projection (LPP) [15], [16] is proposed to preserve the local structure of data in order to capture the non-linearity of the manifold. As this algorithm considers the manifold structure of samples, it achieves very good recognition performance in pattern recognition and draws wide attention in machine learning. However, the objective function of LPP only emphasizes the local structure and ignores the global structure. In order to settle the shortcomings of LPP, Cai et al. proposed an Orthogonal Locality Preserving Projection (OLPP) [17] algorithm and Zhao et al. proposed Uncorrelated Locality Preserving Projection (ULPP) [18] algorithm. Wang et al. proposed a Fast and Orthogonal Locality Preserving Projections (FOLPP) [19] for dimensionality reduction, which simultaneously minimizes the locality and maximizes the globality under the orthogonal constraint. However, all above improved methods do not consider the problem from the perspective of classification.

Yu et al.[20] proposed a Discriminant Locality Preserving Projections (DLPP) method to improve the classification performance of LPP. So with face image expression, posture change obviously, the recognition performance of these algorithms will be significantly reduced. In manifold learning methods, the local neighborhood structure of the original samples is well preserved, and the global structure information of the observed samples is ignored. It is well-known that the local structure information is beneficial for classification and recognition. As a subspace recovery technique, low-rank representation (LRR) [21] is successfully applied to the image processing. In LRR, the original image is divided into two parts: the clean part and the noise part. Therefore, LRR can effectively remove the noise, and it is very robust to noise. The adaptive distance penalty is introduced to low-rank representation, and Fei et al. [22] present adaptive distance penalty based low-rank representation algorithm, in which both the global structure information and the local neighborhood relationship of the original samples are well preserved. In order to construct a robust graph for image classification and clustering. Zheng et al. [23] propose a low-rank representation with local constraint (LRRLC) method. In LRRLC, the graph is constructed by combining a local regularization term into LRR. Although LRR can well capture the global structure information of the observed data via the low-rankness, and it achieves better clustering performance. IfLRR is directly performed on the original data, it will not only lead to the high computational cost, but also degrade performance of clustering and classification. LRR is an unsupervised algorithm, which does not make use of the label information of the observed data. The local structure information, which plays a key role in classification or clustering, is ignored by LRR. Thus, a novel uncorrelated discriminative low-rank preserving projection (UDLRPP) algorithm for feature extraction presented. In UDLRPP, the above defects about LRR can be well solved. Firstly, the low rank representation and the projection matrix are simultaneously obtained. The low-rank representation of the observed samples is computed in the low-dimensional space, which can reduce the amount of computation. The obtained projection matrix can deal with the new samples. Secondly, the label information of the training data is used in the training stage, which may get better classification performance. Thirdly, the local geometry structure information can
be preserved by manifold constrained regularization term. Finally, uncorrelated constraint can removes the redundant features and produce projection matrix.

3. LOW-RANK REPRESENTATION (LRR)

In this paper, it supposes that there are total \( N \) training samples from \( C \) classes. Let \( X_i = [X_{i,1}, X_{i,2}, \ldots, X_{i,N_i}] \in \mathbb{R}^{m \times N_i} \) be the data matrix formed by \( N_i \) training samples from \( i \)th object, in which \( X_{i,j} \) is a column vector corresponding to the sample from the \( j \)th sample of the \( i \)th class. All training data is denoted by \( X = [X_1, X_2, \ldots, X_n] \in \mathbb{R}^{m \times N} \), where \( N = \sum_{i=1}^{C} N_i \) is the total training sample size. In practice, the original dimension \( m \) of the samples is usually quite high. In order to reduce the dimensions, an optimal projection axis \( P = [P_1, P_2, \ldots, P_n] \in \mathbb{R}^{d \times m} \) is expected to obtain, which transforms the samples in \( m \)-dimensional subspace onto a \( d \)-dimensional subspace, where \( d \ll m \). In order to obtain global structure information of the observed samples, low-rank representation (LRR) method is presented for data representation. The objective of LRR is to capture the lowest rank representation matrix \( K \) of the observed matrix \( X \) over a given dictionary \( A \). The representation matrix \( K \) can be obtained by solving the optimization problem as follows.

\[
\min_{K} \text{rank}(K), \text{s.t.} \ X = AK \tag{1}
\]

However, the solution of the above optimization problem is NP-hard. As a common practice, the optimization problem (1) is usually relaxed to the following convex optimization problem by replacing rank function with nuclear norm.

\[
\min_{K} \| K \|_{*}, \text{s.t.} \ X = AK \tag{2}
\]

where \( \| K \|_{*} \) is the nuclear norm of matrix \( K \), i.e., the sum of its singular values. The original samples usually contain noise, or they are even corrupted. Thus Equation (2) can be rewritten as

\[
\min_{K,E} \| K \|_{*} + \lambda \| E \|_{2,1}, \text{s.t.} \ X = AK + E \tag{3}
\]

where \( \| E \|_{2,1} = \sum_{i=1}^{N} \sqrt{\sum_{i=1}^{m} e_{ij}^2} \) is the \( L_{2,1} \)-norm and \( \lambda \) is a scalar constant, which trades off the effect of the noise term and low-rank part. For simplicity, the data matrix \( X \) itself is usually chosen to act as the dictionary. So Equation (3) can be rewritten as

\[
\min_{K,E} \| K \|_{*} + \lambda \| E \|_{2,1}, \text{s.t.} \ X = XK + E \tag{4}
\]

The solution of the optimization problem (4) is usually obtained by the Augmented Lagrange Multiplier (ALM) method.

In LRR, the representation of the observed data is obtained under a global low-rank constraint, and so the global structure can be well captured. The intrinsic representation of the observed samples can be obtained by removing the noise contained in the data. Therefore, LRR based methods are applied to various learning tasks. Several LRR variants methods are given as follows.

LRR-based discriminative projection (LRR-DP) [24]. LRR-DP utilizes the low-rank structure of the observed samples to construct the between-class and within-class scatter matrices, and its algorithm is given below.

**Step 1:** The sample data set \( X \) from \( c \) classes are given, and we calculate the optimal representation \( W \) based on LRR as follows.

\[
\min_{W} \| W \|_{*} + \lambda \| E \|_{2,1}, \text{s.t.} \ X = XW + E \tag{5}
\]
**Step 2:** The within-class $S_W^i$, between-class $S_B^i$, and noise scatter matrices $S_E^i$ are respectively constructed by

\[
S_W^i = \sum_{i,j} [X_{ij} - X_{ij}] [X_{ij} - X_{ij}]^T
\]

\[
S_B^i = \sum_{i,j} [X_{ij} - X_{ij}] [X_{ij} - X_{ij}]^T
\]

\[
S_E^i = \sum_{i,j} [X_{ij} - X_{ij}] [X_{ij} - X_{ij}]^T
\]

**Step 3:** Calculate the projection matrix $P$ by

\[
\max_P \frac{\text{tr}(P^T S_W^i P)}{\text{tr}(P^T S_E^i P) + \mu \text{tr}(P^T S_B^i P)}
\]

By removing the noise data from the observed samples, LRR-DP can extract robust features. Low-rank sparse preserving projections (LSPP) [25]. LSPP introduces the local geometric constraint to LRR, and its objective function is given as follows.

\[
\min_{P,K,E} \|K \|_* + \alpha \|K \|_1 + \beta \|E \|_{2,1} + \lambda \text{tr}(P^T (I - N) (I - N) P)^T) s.t. P^T X = P^T K X + E
\]

In LSPP, both the intrinsic local structure and the global structure can be captured. Low-rank preserving projections (LRPP) [26]. By projecting the observed samples on a low-dimensional subspace, LRPP can learn a low-rank weight matrix, and its objective function is described below

\[
\min_{P,W,E} \frac{1}{2} \sum_{i,j} (W_{ij} + W_{ji}) \|P^T X_i - P^T X_j\|_2^2 + \alpha \|W\|_* + \beta \|E\|_{2,1} s.t. P^T X = P^T K X + E
\]

LRPP can well capture the global structure information of the observed data, and the low-rank weight matrix is learned for preserving the local geometric information.

**4. DISCRIMINATIVE LOW-RANK PRESERVING PROJECTION**

LRR can obtain the lowest-rank representation of the observed samples the best data representation does not mean the best discriminative power. In LRR, the global structure information of the observed samples is captured, and the local neighborhood relationship of the observed samples is ignored. However, the local geometry information of samples often plays a key role in classification and clustering. In addition, the class information of the original samples is not utilized in LRR. LRR cannot directly deal with the new samples because of lacking the projection matrix. The local relationship among the observed data can be well preserved in the manifold learning methods. The neighborhood relationship, which corresponds to the nearest neighbor graph, plays a key role in the methods based on manifold learning. That is to say, the performance of manifold based methods heavily relies on the graph construction process. The conventional graph construction methods include $c$-graph or $k$ nearest neighbor (KNN-graph). LRR is a robust and effective subspace clustering method. Even if the original samples contain noise, it can achieve better performance. It is well-known that low-rank representation matrix $K$ in LRR can well reflect the nearest neighbor relationship among the observed data, and LRR graph is adopted in this algorithm. And also considered the samples label information. So the low-rank representation, manifold regularization constraint, and low-dimensional subspace learning are integrated together for dimensionality reduction. The local geometrical structure of the observed samples is important in clustering or classification. The manifold learning methods usually preserve the local structure information by utilizing the pair-wise Euclidean distances, which are very sensitive to errors and noises in the observed
data. It is well-known that LRR is robust to outliers and noises. The low rankness of the observed samples is used to construct an affinity graph in this paper. In order to use class label information, the global discriminative information is introduced based on LDA idea. Here, initially the global subspace structure of the original samples is firstly captured by LRR, which can obtain the lowest rank representation of the original data. Then, the manifold and discriminative constraint terms based on LRR-graph and LDA are respectively used to learn the local and global discriminative information. Then, a projection matrix $P$ is simultaneously learned for dimensionality reduction. The objective function is described as follows.

$$
\min_K \| K \|_* + \lambda \| E \|_{2,1} + \alpha (tr(P^T S_W P) - tr(P^T S_B P)) + \beta \sum_{i,j=1}^N \| P^T X_i - P^T X_j \|_2^2 K_{ij} s.t. P^T X = P^T X K + E, \ P^T P = 1, K \geq 0 \tag{10}
$$

where $p = [P_1, P_2, ..., P_m] \in \mathbb{R}^{m \times d}$, is a projection matrix which transforms the samples in $m$-dimensional subspace onto a $d$ dimensional subspace ($d \ll m$). The superscript $T$ denotes the transposition operation. The parameters $\alpha$ and $\beta$ are positive constants, and they respectively balance the effect of discriminative term and local manifold term in Equation (10).

$$
S_W = \frac{1}{N} \sum_{i=1}^C \sum_{j \in C_i} (X_j - \bar{X}_i)(X_j - \bar{X}_i)^T \quad \text{is within-class scatter and } \quad S_B = \frac{1}{N} \sum_{i=1}^C n_i (\bar{X}_i - \bar{X})^T (\bar{X}_i - \bar{X}) \quad \text{is between-class scatter.}
$$

Here $C_i$ is the set of column indices that belong to the $i$th class, i.e., for $j \in C_i$, $X_j$ belongs to the $i$th class, $x$ corresponds to the mean of the all training data, $X_i$ corresponds to the mean of the data from the $i$th class, and the other variables in Equation (10) have the same meaning with the corresponding variables in Section 2. In Equation (10), the first term $\| K \|_*$ represents the lowest rank representation of the data, which can capture the global structure information of the observed samples. The second term $\| E \|_{2,1}$ is used to handle sample-specific corruptions and outliers, and ensure that the noise matrix is sparse. The first two items ensure that the essential low-rank representation of the original samples can be gotten. The third term $tr(P^T S_W P) - tr(P^T S_B P)$ is the discriminative constraint term, which is used to capture the discriminative information by making the distances of between class as large as possible, and the distances among within-class samples as small as possible in the low-dimensional projection space. The fourth term $\sum_{i,j=1}^N \| P^T X_i - P^T X_j \|_2^2 K_{ij}$ is the manifold constraint term, by which the adaptive weigh is obtained, and it is guaranteed to make the projected samples in low-dimensional space to respect the similarity neighborhood relationship based on the learned weight matrix in dimensionality reduction process. In this process, noises and outliers in the original data are effectively weakened, which can ensure that the weight graph truly reflect the neighbor relationship among all the samples. If the weight $K_{ij}$ takes a negative value in practical applications, it is lacking of physical interpretation for data representation. So the values in weight matrix $K$ are constrained to be non-negative. It is totally different from the variant methods based on LRR. It is effectively encodes the local and global structure information of the observed samples, and the learned weight graph and discriminative constraint term enhances the separate ability.

4.1. Optimization

There are multiple variables in the optimization problem (10). The optimal solutions in Equation (10) cannot be simultaneously solved. So an alternatively iterative algorithm is designed to solve the above optimization problem. The variable $P$ is firstly fixed to update $K$, and then $K$ is fixed to update $P$.

4.1.1. Fix $P$ to update $K$ and $E$

When the variable $P$ is given, the problem (10) can be converted to the optimization problem as follows.
\[
\begin{align*}
\min_{K, P} & \|K\|_* + \lambda \|E\|_{2,1} + \beta \sum_{i,j=1}^N \|P^T X_i - P^T X_j\|_2^2 W_{ij} \quad \text{s.t.} \quad P^T X = P^T XK + E, K \geq 0
\end{align*}
\] (11)

In order to obtain the solution of the optimization problem (11), it is firstly converted to the optimization problem by introducing an auxiliary variable \(W\) as follows.

\[
\begin{align*}
\min & \|K\|_* + \lambda \|E\|_{2,1} + \beta \sum_{i,j=1}^N \|P^T X_i - P^T X_j\|_2^2 W_{ij} \quad \text{s.t.} \quad P^T X = P^T XK, K \geq 0
\end{align*}
\] (12)

1. Computation of \(K\)

When \(W\), \(E\) and \(P\) are given, the augmented Lagrangian function of problem (12) for \(K\) is given as follows.

\[
\begin{align*}
\min_K & \|K\|_* + \frac{\mu}{2} \|P^T X - P^T XK - E + \frac{T_1}{\mu} K - W + \frac{T_2}{\mu} \|_F^2
\end{align*}
\] (13)

The linearized alternating direction method with adaptive penalty (LADMAP) [27] is used to solve the problem (13).

2. Computation of \(W\)

When \(K\), \(E\) and \(P\) are given, the augmented Lagrangian function of problem (12) for \(W\) is given as follows.

\[
\begin{align*}
\min_W & \beta \sum_{i,j=1}^N \|P^T X_i - P^T X_j\|_2^2 W_{ij} + \frac{\mu}{2} \|Z - W\|_F^2
\end{align*}
\] (14)

Suppose that \(Y_{ij} = \|P^T X_i - P^T X_j\|_2^2\) and \(J = K + \frac{T_2}{\mu}\), Then the optimization problem (14) can be rewritten as

\[
\begin{align*}
\min_W & \beta \sum_{i,j=1}^N Y_{ij} W_{ij} + \frac{\mu}{2} \|J - W\|_F^2
\end{align*}
\] (15)

After optimization of Equation (15) is simplified as

\[
\sum_{i=1}^N \min(Y_{ii} W_i + \frac{\mu}{2} \|J_i - W_i\|_2^2)
\] (16)

The problem (16) is convex and smooth for any \(W_i\), and it has analytic solution. By solving the derivatives with respect to \(W_i\), the optimal solution of the problem (16) can be obtained. Based on the obtained \(W_i\), we can get the \(W\).

3. Computation of \(E\)

When \(K\), \(W\) and \(P\) are given, the augmented Lagrangian function of problem (12) for \(E\) is given as follows.

\[
\begin{align*}
\min_E & \lambda \|E\|_{2,1} + \frac{\mu}{2} \|P^T X - P^T XK - E + \frac{T_1}{\mu} \|_F^2
\end{align*}
\] (17)

The other variables are fixed, and \(E\) can be updated by

\[
E = \arg \min \lambda \|E\|_{2,1} + \frac{\mu}{2} \|P^T X - P^T XK - E + \frac{T_1}{\mu} \|_F^2
\] (18)

4.1.2. Fix \(Z\) and \(E\) to update \(P\):

When \(K\) and \(E\) are fixed, the problem (10) can be converted to the following equivalent problem.

\[
\begin{align*}
\min_P & \alpha (tr(P^T S_W P) - tr(P^T S_B P)) + \beta tr(P^T XLX^TP) \quad \text{s.t.} \quad P^T X = P^T XK + E, P^T P = I
\end{align*}
\] (19)
where $L = D - W$ is graph Laplacian, $D$ is a diagonal matrix whose diagonal elements $D_{ii} = \sum_j W_{ij}$. Based on the constrained term $P^T X = P^T X K + E$ the problem in (19) can be transformed into an unconstrained problem as follows.

$$\min_{P} \left\| P^T X - P^T X K \right\|_F^2 + \alpha (tr(P^T S_w P) - tr(P^T S_B P)) + \beta Tr(P^T X L X^T P)s.t. \ P^T P = I$$  \hspace{1cm} (20)$$

The above Equation (20) can further equal to the following equation.

$$\min_{P} Tr((X - X K)(X - X K)^T + \alpha (S_w - S_B) + \beta X L X^T) s.t. \ P^T$$  \hspace{1cm} (21)$$

The solution of (21) can be gotten by solving the minimum eigenvalues problem.

$$((X - X K)(X - X K)^T + \alpha (S_w - S_B) + \beta X L X^T)P = \eta P$$  \hspace{1cm} (22)$$

5. UNCORRELATED DISCRIMINATIVE LOW-RANK PRESERVING PROJECTION (UDLRPP)

The reduced projection features are obtained onto the common space by solving the Equation of (22) is represented as $P = [P_1, P_2, ..., P_n]$. The column vectors $P_i$ in $P$ are the eigenvectors corresponding to the first $d$ smallest eigen values $\eta$ in Equation (22). Although, the obtained common space is quite discriminative, the different components of the reduced feature vectors are correlated and this space is called Correlated Common Space (CCS), denoted by $P_{ccs}$. In the proposed approach, instead of classifying directly from CCS, first transform features of the correlated common space $P_{ccs}$ to the Uncorrelated Common Space (UCS) $P_{ucs}$, and then classification is done. For this, UCS is obtained from the CCS with the help of the transformation matrix $U = [u_1, u_2, ..., u_d]$[28]. The columns of $U$ (eigenvectors) are essentially the solutions of the following generalized eigen value equation corresponding to the first $d$ lowest eigen values

$$P_{ccs} (L_{ccs} + B_{ccs}) P_{ccs}^T = \lambda_{ccs} P_{ccs} G_{ccs} P_{ccs}^T u$$  \hspace{1cm} (23)$$

Where $L_{ccs}$ and $B_{ccs}$ are Laplacian and between-class transformation matrices respectively. These matrices are obtained from the CCS. The matrix $G_{ccs} = I - (1/V_N) ee^T$, where $I$ is an identity matrix, and $e = (1,1,...,1)^T$. Therefore, the transformed UCS can be obtained by linear projection:

$$P_{ucs} = U^T P_{ccs}$$  \hspace{1cm} (24)$$

Finally, kNN classifier used for classification onto the uncorrelated common discriminative space.

**Algorithm 1: UDLRPP**

Input: Training samples matrix $X$, maximum iterations $maxiter$, and parameter

Initialize: PCA projection matrix is used to initialize $P$, $K=M=E=T_1=T_2=0$, $\mu_{max} = 10^{10}$, $\rho = 1.1$.

For $iter = 1: max iter$

1. The other variables are fixed, and $K$ is updated

By Equation (13):

$$\min_{\mu} \left\| K \right\| + \frac{\mu}{2} \left( \left\| P^T X - P^T X K - E + \frac{T_1}{\mu} \right\|_F^2 + \left\| K - W + \frac{T_2}{\mu} \right\|_F^2 \right)$$

http://www.iaeme.com/IJARET/index.asp 1263 editor@iaeme.com
2. The other variables are fixed, and $W$ is updated by Equation (16):
\[
\sum_{i=1}^{N} \min \left( Y_i, W_i + \frac{\mu'}{2} \| J_i - W_i \|^2 \right)
\]

3. The other variables are fixed, and $E$ is updated by Equation (18):
\[
E = \arg \min \lambda \| E \|_2 + \frac{\mu}{2} \| P^T X - P^T X K - E + \frac{T_1}{\mu} \|^2_F.
\]

4. The multipliers is updated by
\[
T_1 = T_1 + \mu (P^T X - P^T X K - E)
\]
\[
T_2 = T_2 + \mu (K - W)
\]

5. The parameter $\mu$ is updated by
\[
\mu = \min (\rho \mu, \mu_{\text{max}})
\]

6. The projection matrix $P$ is updated by Equation (22)
\[
((X - X K)(X - X K)^T + \alpha (S_w - S_b) + \beta XLX^T)P = \eta P
\]

7. Applied Un-correlation Constraint on Projection Matrix $P$ by Equation (23) and (24)
\[
P_{ccs}(L_{ccs} + B_{ccs})P_{ccs}^T u = \lambda_{ccs} P_{ccs} G_{ccs} P_{ccs}^T u
\]
\[
P_{uCS} = U^T P_{ccs}
\]

End

Output: Un-correlated Projection matrix $P$.

6. EXPERIMENTS AND ANALYSIS

Here, experiments are performed to exhibit the capability of the proposed UDLRPP algorithm. The performance of UDLRPP algorithm is measured on six public image databases and the experiment setting on these databases is given as follows. The parameters optimal values in all algorithms are chosen in the tests, and the values of parameters in the proposed algorithm are 1, 5, 0.1 for $\alpha$, $\beta$, $\lambda$ in Yale database, 9, 0.01, 6 in Yale B database, 3, 5, 1 in FERET database and 3, 5, 20 in CMU PIE database respectively. It is difficult to directly judge the effectiveness of the extracted features or not. Thus, after dimension reduction, the nearest neighbor classifier is used for classification and recognition after feature extraction by the above methods. According to the gained recognition results, it is possible to judge the effectiveness of the proposed feature extraction algorithm.
Uncorrelated Discriminative Low-Rank Preserving Projection for Dimensionality Reduction

6.1. Face recognition using YALE Database

Here UDLRPP is checked for different face images. As shown in Figure 1, Yale face database [29] pictures contain 165 GIF-scale pictures for 15 people. There are 10 pictures per topic, one per facial exposition or setup: center-light, glass, happy, left-light, glass-free, ordinary, right-light, sad, sleepy, amazed, and wink. Four pictures are elected of each individual (center light, left light, normal and happy), and the remaining of the database was the test collection. Thus, the sample size for training is 60 and the sample size for tests is 105. Each picture is cropped and re-dimensioned. Figure 2 shows the specimen pictures of a candidate from the Yale face database.

Figure 1 Block-diagram of the proposed UDLRPP method.

Figure 2 Yale Face image Database Samples
The picture samples are split simultaneously so that 1 (l = 2, 3, 4, 5, 6) pictures are gathered and noted for each individual person to create a training set, while others are used as a test set. Here, the database is randomly split into 60% training and 40% test specimens to check the capability of the suggested algorithm. The identification frequency outcomes for YALE face dataset are shown in the Figure 3. Table 1 shows the UDLRPP identification in YALE database is greater than LPP, OLPP, FOLPP, -DP, LRPP, LSPP and HEOULPP. LRR

![Figure 3](image-url)

**Figure 3** Face recognition Rates (%) vs Dimensions on Yale Database

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Training Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>LPP</td>
<td>74.2 (10D)</td>
</tr>
<tr>
<td>OLPP</td>
<td>74.6 (10D)</td>
</tr>
<tr>
<td>LRR-DP</td>
<td>90.2 (10D)</td>
</tr>
<tr>
<td>LSPP</td>
<td>91.1 (10D)</td>
</tr>
<tr>
<td>LRPP</td>
<td>92.6 (10D)</td>
</tr>
<tr>
<td>FOLPP</td>
<td>72.8 (10D)</td>
</tr>
<tr>
<td>HEOULPP</td>
<td>84.6 (10D)</td>
</tr>
<tr>
<td>UDLRPP</td>
<td>96.3 (10D)</td>
</tr>
</tbody>
</table>

D: Dimensions

**6.2. Face recognition using ORL Database**

The ORL face database [30] consists of a total of 400 face images, which including a total of 40 people, 10 samples per person, some of which are shown in Figure 4. Face images are cropped and resized to 64×64, which including small changes in facial expressions and gestures, scale changes within 20%, difference in with and without glasses.
Table 2 outlines the best outgrowths and the optimum dimension for each method. The UDLRPP method appeared the best. Figure 5 exhibit the recognition rate versus the dimensions. The UDLRPP algorithm achieved the highest recognition rates among all methods.

### Table 2: Face Recognition Rates (%) of Methods for ORL Database

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Training Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>LPP</td>
<td>73.4 (20D)</td>
</tr>
<tr>
<td>OLPP</td>
<td>73.8 (20D)</td>
</tr>
<tr>
<td>LRR-DP</td>
<td>84.1 (20D)</td>
</tr>
<tr>
<td>LSPP</td>
<td>86.0 (20D)</td>
</tr>
<tr>
<td>LRPP</td>
<td>86.2 (20D)</td>
</tr>
<tr>
<td>FOLPP</td>
<td>71.5 (20D)</td>
</tr>
<tr>
<td>HEOULPP</td>
<td>83.5 (20D)</td>
</tr>
<tr>
<td>UDLRPP</td>
<td>89.6 (20D)</td>
</tr>
</tbody>
</table>

D: Dimensions

### 6.3. Face recognition using AR Database

AR Database comprises of more than 4000 pictures of 126 color face pictures (70 males and 56 females). In two sessions (separated by two weeks), the image of the most individuals (65 men and 55 women) is taken. The 13 gray-scale pictures were easily interpretable for each session, each session to the size of 50 x 40 for each event.

The pictures from each event consist of three illuminated pictures, three wearing scarf pictures, three sun glass pictures and four expressive variants. Some of which are shown in the Figure 6.
Here, Table 3 reports the maximum recognition rates and the corresponding dimensions of the different methods with different numbers of training samples. The proposed method UDLRPP achieved the best recognition rate compared with LPP, OLPP, FOLPP, LRR-DP, LRPP, LSPP and HEOULPP. Figure 7 illustrates the recognition rate versus the dimensionality which shows that the proposed method consistently outperformed the other methods in most experimental cases.

**Table 3** Face Recognition Rates (%) Of Methods For AR Database

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Training Samples</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>15D</td>
<td>30D</td>
<td>45D</td>
<td>60D</td>
<td>80D</td>
</tr>
<tr>
<td>LPP</td>
<td>71.7 (15D)</td>
<td>74.2 (30D)</td>
<td>77.2 (45D)</td>
<td>78.4 (60D)</td>
<td>79.3 (80D)</td>
<td></td>
</tr>
<tr>
<td>OLPP</td>
<td>72.2 (15D)</td>
<td>74.8 (30D)</td>
<td>77.7 (45D)</td>
<td>79.4 (60D)</td>
<td>81.9 (80D)</td>
<td></td>
</tr>
<tr>
<td>LRR-DP</td>
<td>84.1 (15D)</td>
<td>85.3 (30D)</td>
<td>86.1 (45D)</td>
<td>87.9 (60D)</td>
<td>88.6 (80D)</td>
<td></td>
</tr>
<tr>
<td>LSPP</td>
<td>87.2 (15D)</td>
<td>89.4 (30D)</td>
<td>90.1 (45D)</td>
<td>90.9 (60D)</td>
<td>91.4 (80D)</td>
<td></td>
</tr>
<tr>
<td>LRPP</td>
<td>86.4 (15D)</td>
<td>86.2 (30D)</td>
<td>88.1 (45D)</td>
<td>89.3 (60D)</td>
<td>90.8 (80D)</td>
<td></td>
</tr>
<tr>
<td>FOLPP</td>
<td>74.5 (15D)</td>
<td>76.1 (30D)</td>
<td>80.8 (45D)</td>
<td>83.2 (60D)</td>
<td>87.3 (80D)</td>
<td></td>
</tr>
<tr>
<td>HEOULPP</td>
<td>80.1 (15D)</td>
<td>87.4 (30D)</td>
<td>90.1 (45D)</td>
<td>92.3 (60D)</td>
<td>96.9 (80D)</td>
<td></td>
</tr>
<tr>
<td>UDLRPP</td>
<td>90.8 (15D)</td>
<td>92.3 (30D)</td>
<td>96.4 (45D)</td>
<td>97.7 (60D)</td>
<td>98.8 (80D)</td>
<td></td>
</tr>
</tbody>
</table>

D: Dimensions

**Figure 6** AR Face image Database Samples
6.4. Face recognition using YALEB Database

The Extended Yale B face database contains lot of face images of 38 subjects under different pose and illumination conditions. For this experiment, a subset of the database selected that includes only those under illumination conditions. Each person had 64 different front images. All the images were resized to a resolution of 32×32 pixels. Some images from one person are illustrated in Figure 8. Randomly selected r images for each person as training samples, and the rest of the samples for testing.

Table 4 shows the recognition rates for the corresponding dimensions and the number of training samples. The proposed method UDLRPP obtained the best recognition rate when the number of training samples for each class varied from 4 to 6. Figure 9 illustrate the recognition rate versus the dimensionality, and it shows that the UDLRPP algorithm achieved a higher recognition rate than other methods.
Table 4 Face Recognition Rates (%) Of Methods For Yale b Database

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Training Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 (10D)</td>
</tr>
<tr>
<td>LPP</td>
<td>28.3</td>
</tr>
<tr>
<td>OLPP</td>
<td>29.2</td>
</tr>
<tr>
<td>LRR-DP</td>
<td>92.2</td>
</tr>
<tr>
<td>LSPP</td>
<td>68.3</td>
</tr>
<tr>
<td>LRPP</td>
<td>72.6</td>
</tr>
<tr>
<td>FOLPP</td>
<td>52.3</td>
</tr>
<tr>
<td>HEOULPP</td>
<td>94.1</td>
</tr>
<tr>
<td>UDLRPP</td>
<td>96.4</td>
</tr>
</tbody>
</table>

D: Dimensions

6.5. Face recognition using FERET Database

The FERET face image database contains 200 individuals and each person has 7 images which include different illumination conditions and facial expressions. All the original gray images were resized to 32 × 32 pixels. Figure 10 illustrates several images of one person. Here, randomly grouped the images of each person into two parts. One part was used as training samples with r images being chosen for each individual, and the other as testing samples.

![Figure 10 FERET Face image Database Samples](image)

Table 5 Face Recognition Rates (%) Of Methods For Feret Database

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Training Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 (10D)</td>
</tr>
<tr>
<td>LPP</td>
<td>20.2</td>
</tr>
<tr>
<td>OLPP</td>
<td>21.2</td>
</tr>
<tr>
<td>LRR-DP</td>
<td>20.2</td>
</tr>
</tbody>
</table>
Uncorrelated Discriminative Low-Rank Preserving Projection for Dimensionality Reduction

<table>
<thead>
<tr>
<th>Method</th>
<th>36.2</th>
<th>50.6</th>
<th>47.4</th>
<th>51.1</th>
<th>52.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSPP</td>
<td>(10D)</td>
<td>(50D)</td>
<td>(80D)</td>
<td>(140D)</td>
<td>(180D)</td>
</tr>
<tr>
<td>LRPP</td>
<td>40.6</td>
<td>52.4</td>
<td>52.8</td>
<td>53.2</td>
<td>54.6</td>
</tr>
<tr>
<td></td>
<td>(10D)</td>
<td>(40D)</td>
<td>(100D)</td>
<td>(130D)</td>
<td>(180D)</td>
</tr>
<tr>
<td>FOLPP</td>
<td>32.3</td>
<td>44.2</td>
<td>50.3</td>
<td>53.1</td>
<td>53.6</td>
</tr>
<tr>
<td></td>
<td>(10D)</td>
<td>(40D)</td>
<td>(80D)</td>
<td>(140D)</td>
<td>(180D)</td>
</tr>
<tr>
<td>HEOULPP</td>
<td>46.1</td>
<td>55.2</td>
<td>56.1</td>
<td>57.3</td>
<td>58.4</td>
</tr>
<tr>
<td></td>
<td>(10D)</td>
<td>(20D)</td>
<td>(40D)</td>
<td>(70D)</td>
<td>(180D)</td>
</tr>
<tr>
<td>UDLRPP</td>
<td>48.4</td>
<td>58.2</td>
<td>59.3</td>
<td>61.0</td>
<td>61.6</td>
</tr>
<tr>
<td></td>
<td>(10D)</td>
<td>(20D)</td>
<td>(40D)</td>
<td>(70D)</td>
<td>(180D)</td>
</tr>
</tbody>
</table>

D: Dimensions

Figure 11 Face recognition Rates (%) vs Dimension on FERET Database

The recognition rates and the corresponding dimensions of the different methods are given in Table 5. The proposed UDLRPP algorithm achieved the best recognition rate compared with other methods. Figure 11 shows the recognition rate versus the dimensionality. The recognition rate of UDLRPP was not good when the dimension was above 60. However, when the dimension was above 60, the performances of these methods dropped significantly. So, that the advantage of UDLRPP is obvious when the dimension is lower.

6.6. Face recognition using CMU PIE Database

The CMU PIE face database contains 68 individuals with 41368 face images. These images were captured under varying illumination, pose and expression. Here, a subset of images selected that contains 1632 images of 68 individuals, each of which has 24 images. The subset includes facial expressions, pose and illumination changes. All of these face images are resized to 64×64, some of which are shown in Figure 12.

Figure 12 CMU PIE Face image Database Samples

Table 6 shows the best outcomes and the optimum dimension for each method. The UDLRPP method achieved the highest recognition rate compared with other methods such as LPP, OLPP, LRR-DP, LSPP, LRPP, FOLPP, and HEOULPP when dimensions were below 60. Figure 13 exhibit the recognition rate versus the dimensions.
6.7. Computational Complexity

The time complexity of UDLRPP algorithm is analyzed here. For getting better clarity, some existed variables are re-explained as follows. The lowest rank for X is denoted by \( r_X \), Tis the iteration number in the proposed algorithm. The dimension in the original space is denoted by \( m \), the dimension after projection is \( d \) and the number of training samples is \( N \). Through analysis, find that the computational cost of UDLRPP is mainly fixated by step 1 and step 6 in Algorithm. It necessarily to compute the singular value decomposition in step 1, and the time complexity is at most \( O(dN^2) \) each repetition.

### Table 6 Face Recognition Rates (%) Of Methods For Cmu Pie Database

<table>
<thead>
<tr>
<th>Method</th>
<th>Number of Training Samples</th>
<th>6</th>
<th>10</th>
<th>14</th>
<th>18</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(10D)</td>
<td>(15D)</td>
<td>(30D)</td>
<td>(100D)</td>
<td>(160D)</td>
</tr>
<tr>
<td>LPP</td>
<td></td>
<td>28.3</td>
<td>40.2</td>
<td>61.2</td>
<td>90.5</td>
<td>91.0</td>
</tr>
<tr>
<td>OLPP</td>
<td></td>
<td>27.3</td>
<td>40.2</td>
<td>75.2</td>
<td>92.1</td>
<td>92.2</td>
</tr>
<tr>
<td>LRR-DP</td>
<td></td>
<td>45.2</td>
<td>72.1</td>
<td>90.2</td>
<td>91.8</td>
<td>91.2</td>
</tr>
<tr>
<td>LSPP</td>
<td></td>
<td>42.2</td>
<td>70.6</td>
<td>85.3</td>
<td>88.9</td>
<td>91.6</td>
</tr>
<tr>
<td>LRPP</td>
<td></td>
<td>72.6</td>
<td>87.8</td>
<td>94.2</td>
<td>94.3</td>
<td>94.6</td>
</tr>
<tr>
<td>FOLPP</td>
<td></td>
<td>60.3</td>
<td>85.3</td>
<td>94.6</td>
<td>94.7</td>
<td>94.8</td>
</tr>
<tr>
<td>HEOULPP</td>
<td></td>
<td>82.1</td>
<td>91.2</td>
<td>96.4</td>
<td>98.4</td>
<td>98.5</td>
</tr>
<tr>
<td>UDLRPP</td>
<td></td>
<td>82.4</td>
<td>91.3</td>
<td>96.4</td>
<td>99.8</td>
<td>99.8</td>
</tr>
</tbody>
</table>

D: Dimensions

It necessarily to compute a generalized eigen value problem in step 6, and the time complexity is \( O(m^3) \) each repetition. The number of repetitions are considered, and the and the computational complexity of UDLRPP algorithm is \( O(TdN^2 + Tm^3) \)

![Figure 13 Face recognition Rates (%) vs Dimensions on CMU PIE Database](image.png)

7. CONCLUSIONS AND FUTURE WORK

In this paper, a new supervised feature extraction method was developed which is called uncorrelated discriminative low-rank preserving projection (UDLRPP). In this method, low rankness, the sparsity and preserving projection of the original data are enciphered to construct
an adaptive graph. Here, both the local manifold and the global structure information of the original samples can be finely ensnared. Simultaneously, the uncorrelated and discriminative constraint terms are imported into the objective function to removes the redundancy between features to enhance the discriminant ability. Thereby, the classification and recognition ability can be enhanced. Many tests were performed on different public image databases, which conclude that UDLRPP can achieve the excellent performance against other popular feature extraction algorithms. The service ability of UDLRPP required to be verified in large-scale databases.

REFERENCES


