

“© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

# Robust 2D Joint Sparse Principal Component Analysis with F-norm Minimization for Sparse Modelling: 2D-RJSPCA

Muhammad Imran Razzak  
Advanced Analytics Institute,  
University of Technology, Sydney  
Australia  
imran.razzak@ieee.org

Raghib Abu Saris  
CPHHI, King Saud bin Abdulaziz  
University for Health Sciences  
Saudi Arabia  
sarisr@ksau-hs.edu.sa

Michael Blumenstein, Guandong Xu\*  
Faculty of Engineering & IT  
University of Technology, Sydney  
Australia  
michael.blumenstein,guandong.xu@uts.edu.au

**Abstract**—Principal component analysis (PCA) is widely used methods for dimensionality reduction and Lots of variants have been proposed to improve the robustness of algorithm, however, these methods suffer from the fact that PCA is linear combination which makes it difficult to interpret complex nonlinear data, and sensitive to outliers or cannot extract features consistently, i.e., collectively; PCA may still require measuring all input features. 2DPCA based on  $\ell_1$ -norm has been recently used for robust dimensionality reduction in the image domain but still sensitive to noise. In this paper, we introduce robust formation of 2DPCA by centering the data using the optimized mean for two-dimensional joint sparse as well as effectively combining the robustness of 2DPCA and the sparsity-inducing lasso regularization. Optimal mean helps to improve the robustness of joint sparse PCA further. The distance in spatial dimension is measure in F-norm and sum of different datapoint uses 1-norm. 2DR-JSPCA imposes joint sparse constraints on its objective function whereas additional plenty term help to deal with outliers efficiently. Both theoretical and empirical results on six publicly available benchmark datasets shows that Optimal mean 2DR-JSPCA provides better performance for dimensionality reduction as compare to non-sparse (2DPCA and 2DPCA-L1) and sparse (SPCA, JSPCA).

**Keywords**—CA, 2DPCA, Sparse PCA, PCA Optimization, Principal Component Analysis, 2DPCA, Sparse PCA, PCA Optimization, Principal Component Analysis

## I. INTRODUCTION

Dimensionality reduction is a classical problem in pattern recognition and plays an important roles due to its contribution to alleviate the so-called "curse of dimensionality". Its aim is to project the high-dimensional feature space into low-dimensional space by preserving the importance of statistical properties of data, so that it could be interpreted efficiently. During the last decades, we have witnessed many algorithms have been proposed to robustly reduce the effect of complex noise or outliers in realistic data. Among these methods, unsupervised dimensionality reduction methods are more useful in piratical application since data labeling is usually expensive and oftenly absense of prior knowledge for most scientific problems.

PCA is one of the extensively used unsupervised method, which projects high-dimensional data into liner orthogonal

space. However, one of the major drawback is that PCA is linear combination and loading are non-zero. This makes PCA data interpretation difficult and it is still sensitive to outliers (as its co-variance matrix is derived from  $\ell_2$ -norm) that affect its performance. Thus, it fails to deal with outliers that often appears in real world data. Moreover, before applying PCA and LDA, there is need to convert image into one dimensional vector, thus it may not exploit image's spatial structural information very well which is very important for image representation [9]. To overcome these issues, several variants of PCA have been proposed to improve the effectiveness of dimensionality reduction and robustness against outliers. Recently,  $\ell_1$ -norm-based subspace learning methods have been widely used for dimensionality reduction to deal with outliers, i.e., imposing  $\ell_1$ -norm on loss term [1], [3], [5], [5], [8]. Although, these methods are able to reduce the effect of outliers to some extent, low-dimensional feature space is linear combination of all original features in high-dimensional features which makes it difficult to interpret complex nonlinear data. To deal with this issue, 2D-PCA [15], 2DLDA [16], multi-linear PCA [6] and JGSPCA [4] etc. have been developed. Two-dimensional PCA (2DPCA) is more efficient, due to its direct formulation based on raw two-dimensional images i.e. two dimensional subspace learning methods directly calculate the class scatter metrics from images, hence can reveal the spatial structural information of image that is quite important for image classification task. However, the limitation of 2DPCA is the dense basis which makes it difficult to explain the resulting features. As such, it is desirable to select the most relevant or salient elements from a large number of features. Sparse modeling has received increased attention to deal with this issue that could be achieved by regularizing objective variables with a lasso penalty term using the  $\ell_1$ -norm. In 2DPCA,  $\ell_1$ -norm works as a robust measure for sample dispersion, not regularizing basis vectors. Although the above discussed variations either 1D or 2D are able to reduce the effect of outliers to some extent however one major disadvantage is that each new feature in low dimensional subspace is linear combination of all the original features exist in high dimensional space that effect the accuracy due to features redundancy. Whereas it is quite difficult to interpret new features. Sparseness could be the solution to overcome this issue. Sparse principal component

analysis has no ability to jointly select features. 2D Joint sparse PCA can effectively integrates feature selection into subspace to exclude the redundant features [10]. To address this issue, in this paper, we have presented a novel approach called 2D Robust Joint Sparse PCA (2DR-JSPCA) that simultaneously optimizes mean and sparse projection matrix in the criterion function that helps to integrate feature selection into subspace learning to exclude the redundant features.

## II. 2DR-JSPCA

Dimensionality reduction methods try to preserve a certain kind of linear representation after projection. As the existing methods fail to select useful features or not efficient for outliers. In this section, we propose robust 2D joint sparse principal component analysis in detail. As described earlier, projection procedure involves all the original features and it may have redundant or irrelevant features. Inspired by SPCA and JSPCA, We aim to learn transformation matrix with sparsity. The objective function called optimal mean 2D robust JSPCA has more freedom to jointly select the useful features from low-dimensional representation. As the objective function is based on F-norm thus it helps to reduce the effect of outliers. The objective function imposes simultaneously optimize mean and sparse projection due to joint sparsity constraints on it objective function as show in in equation 4.

### A. Objective Function

Considering the outliers appearance in the dataset and consistent selection of optimal features, we propose two-dimensional joint sparse principal component analysis (2DR-JSPCA) for reconstructing the data matrix. 2DR-JSPCA aims to seek the projection matrix that makes the value of objective function minimum. The objective function given below in equation 1 uses optimal mean that helps to improve the robustness against outliers. The objective of proposed function is to effectively combine the robustness of 2DPCA and the sparsity-inducing lasso regularization by imposing jointly sparse constraints on its objective function as well as introducing the additional penalty term. As a resultant, we hope to obtain robust low dimensional subspace that is not is not uniquely determined up to an orthogonal transformation.

Consider the observed data  $A'$  ( e.g., image patch, video frame or speech signal), corrupted by noise  $\eta$ , i.e.

$$A = A' + \eta$$

Where  $A'$  is clean data and  $\eta$  is the additive noise in the image.

$$A = [A_1, \dots, A_N]$$

where  $A_j \in R^{m \times n}$ .

To this end, consider the optimization problem

$$\min_{Q,P} J(Q,P) = \min_{Q,P} \sum_{j=1}^N \|K_j - QP^T K_j\|_F^2 + \lambda \|Q\|_F^2 \quad (1)$$

where  $K_j$  is  $A_j - M$ . Matrix  $Q \in \mathbb{R}^{n \times d}$  transforms each sub-image into lower-dimensional subspace and the matrix  $P \in \mathbb{R}^{n \times d}$  recovers the data matrix. Furthermore, while we require  $P$  to be orthogonal ( $P^T P = I_d$ ), we do not require the orthogonality of the matrix  $Q$ . This enables the 2DR-JSPCA has more freedom to learn low dimensional space that approximate to high dimensional data in flexible manner.

To prove the objective function mentioned in equation 1, firstly we consider the objective function of 2DPCA and add optimal mean and sparseness respectively.  $V = [v_1, v_2, \dots, v_d] \in \mathbb{R}^{n \times d}$  is the projection matrix, where  $v_1$  is the first basis vector of 2DPCA that maximizes the  $\ell_1$ -norm-based dispersion of projected samples.

$$V^* = \operatorname{argmin}_{V^T V = I_d} \sum_{n=1}^N \|A_i - A_i V V^T\|_F^2 \quad (2)$$

Based on the Optimal mean, the above equation can be written as

$$V^* = \operatorname{argmin}_{V^T V = I_d} \sum_{n=1}^N \|(A_j i - M) - (A_j - M) V V^T\|_F^2 \quad (3)$$

The above objective function is called optimal mean 2DPCA with F-norm minimization. In order to model the sparsity, the above equation can be rewritten as

$$\min_{Q,P} J(Q,P) = \min_{Q,P} \sum_{j=1}^N \|(A_j - M) - QP^T(A_j - M)\|_F^2 + \lambda \|Q\|_F^2 \quad (4)$$

where the matrix  $Q \in \mathbb{R}^{n \times d}$  transforms each sub-image into lower-dimensional subspace and the matrix  $P \in \mathbb{R}^{n \times d}$  recovers the data matrix. Directly solving the above equation is difficult as both loss term and regularization term are non-smooth. To find optimal solution for equation II-A, we applied some mathematical relations. The following theorems plays important role in solving the minimizers of the optimization problem as shown in equation II-A.

The minimizers of the optimization problem 4 satisfy the equation

$$\left( \sum_{j=1}^N (A_j^T - M)(A_j - M) \right) P = \left( \sum_{j=1}^N (A_j^T - M)(A_j - M) \right) Q \quad (5)$$

*Proof:* First, utilizing definition of the Frobenius norm, the cyclic and linearity properties of the trace function, and the orthogonality of  $P$ , we rewrite the objective function  $J$  in a more computationally tractable way.

$$J(P, Q) = \sum_{j=1}^N \|(A_j - M) - (A_j - M)QP^T\|_F^2 + \lambda \|Q\|_F^2$$

$$J(P, Q) = \sum_{j=1}^N \text{tr} \left[ ((A_j - M)^T - PQ^T(A_j - M)^T) \right. \\ \left. (A_j - M) - (A_j - M)QP^T \right] + \lambda \text{tr}(Q^T Q)$$

$$J(P, Q) = \sum_{j=1}^N \text{tr} \left[ (ZZ^T - ZZ^T QP^T - PQ^T ZZ^T) \right. \\ \left. + PQ^T ZZ^T QP^T + \lambda \text{tr}(Q^T Q) \right]$$

$$J(P, Q) = \sum_{j=1}^N \text{tr} \left( ZZ^T - 2ZZ^T QP^T + \right. \\ \left. ZZ^T QQ^T + \lambda \text{tr}(Q^T Q) \right)$$

Where  $Z = A_j - M$ .

Now, differentiating above equation with respect to  $Q$ ,

$$\frac{\partial J}{\partial Q} = \sum_{j=1}^N (-2A_j^T A_j P + 2A_j^T A_j Q) + 2\lambda Q.$$

Therefore,

$$\frac{\partial J}{\partial Q} = 0 \Rightarrow \sum_{j=1}^N (-2ZZ^T P - M + (2ZZ^T)Q) + 2\lambda Q = 0 \quad (6)$$

$$\sum_{j=1}^N ((ZZ^T)P) = \sum_{j=1}^N ((ZZ^T)Q + \lambda Q) \quad (7)$$

$$\left( \sum_{j=1}^N ZZ^T \right) P = \left( \sum_{j=1}^N (\lambda I_n + ZZ^T) \right) Q \quad (8)$$

Hence, the result follows.  $\blacksquare$

In order to reduce the feature redundancy, we imposed orthogonal constraints  $I_n$  where  $Q$  is the project the weighted data matrix an  $P$  is used to recover it.

To optimize the mean matrix, we fix  $P$  and  $Q$  and differentiating with respect to  $M$ , we get.

$$M = \underset{M}{\text{argmin}} \sum_{n=1}^N \text{tr} \left[ (Md_n M^T + 2M^T d_n A_n P Q^T) \right]$$

$$-2M^T d_n A_n - 2M^T d_n M Q P^T + E_i + 2A_n^T d_i P Q^T + \lambda \text{tr}(Q^T Q)$$

$$\text{where } d_n = [ \|E\|_F^2 ]^{-1}$$

By differentiating with respect to  $M$  we get

$$M = \sum_{n=1}^N d_n A_n / \sum_{n=1}^N d_n \quad (9)$$

Observe that, if  $\lambda = 0$ , then  $P = Q$  as a results objective function degenerate to OMF-2DPCA [14].

$$\left( \sum_{j=1}^N ZZ^T \right) P = \left( \sum_{j=1}^N (ZZ^T) \right) Q$$

$$Q = \left[ \sum_{j=1}^N ZZ^T \right]^{-1} \left[ \sum_{j=1}^N ZZ^T \right] P = P.$$

Furthermore, the objective function simplifies to

$$J(Q, P) = \sum_{j=1}^N \|Z - ZPP^T\|_F^2$$

Hence, the proposed 2DR-JSPCA degenerates to OMF-2DPCA. As such, in some sense 2DR-JSPCA generalizes the 2DPCA.

Observe that, if  $\lambda > 0$ , then  $Q$  is sparse .

$$\min_{Q, P} \sum_{j=1}^N \|Z - ZQP^T\|_F^2 + \lambda \|Q\|_F^2 = \sum_{j=1}^N (-2ZZ^T P \\ + 2ZZ^T Q + 2\lambda Q)$$

$$V = \underset{V^T V = I_d}{\text{argmin}} \sum_{n=1}^N \|(X_i - M) - (X_i - M)VV^T\|_F^2$$

The objective function in equation 4 is to find sparse matrix  $Q$  and orthogonal matrix  $P$

$$\left( \sum_{j=1}^N A_j^T A_j \right) P = \sum_{j=1}^N (\lambda I_n + A_j^T A_j) Q$$

Hence, the proposed 2DR-JSPCA is sparse when  $\lambda > 0$ .

If  $UDV^T$  is the singular value decomposition (SVD) of  $\sum_{j=1}^N A_j^T A_j Q$ , then

$$P = UI_{n \times d} V^T \quad (10)$$

is orthogonal and minimizes Equation (II-A) for a given  $Q$ .

*Proof:*

Recall that  $U$  and  $V$  are orthogonal matrices of sizes  $n \times n$  and  $d \times d$ , respectively. As such,

$$P^T P = V I_{n \times d}^T U^T U I_{n \times d} V^T = I_d$$

■

Before showing the convergence of 2DR-JSPCA, we need to give the following lemma.

For any nonzero vector

$$P, Q \in R^c$$

, the following results hold:

$$\|P\|_F - \frac{\|P\|_F^2}{2\|Q\|_F} \leq \|Q\|_F - \frac{\|Q\|_F^2}{2\|Q\|_F}$$

Below in table I we describe an iterative algorithm of 2DR-JSPCA for training samples  $A_1, \dots, A_n$  of size  $m \times n$ , and regularization parameter  $\lambda$ .

TABLE I. ALGORITHMIC PROCEDURE OF 2D JOINT SPARSE PCA

<p><b>Input:</b> <math>A_j \in \mathbb{R}^{m \times n}</math> for <math>j = 1, \dots, N</math> where <math>A</math> is centralized, and parameter <math>\lambda</math>.  <b>Output:</b> Matrices <math>P</math> and <math>Q</math></p>
<p><b>Step-I:</b> Initialize the matrix <math>P</math></p> <p>While not converge do</p> <p><b>Step-II:</b> Update the optimal mean <math>M = \arg\min_M \sum_{n=1}^N \text{tr}(E_n^T d_n E_n)</math> by using Equation (9)</p> <p><b>Step-III:</b> Minimize the objective function with respect to <math>Q</math> by computing the matrix <math>Q</math> using Equation (5)</p> <p><b>Step-IV:</b> Compute the SVD of <math>\sum_{j=1}^N Z Z^T Q</math></p> <p><b>Step-V:</b> Minimize the objective function with respect to <math>P</math> by updating the matrix <math>P</math> using Equation (10).</p> <p>end while</p>

### B. Convergence Analysis

Lemma II-A plays important role in determining the proof of convergence of proposed objective function.

Given all the variables in objective function equation 4, iterative scheme of 2DR-JSPCA described in table 1 monotonically decrease the objective function value in each iteration and converge to local optima.

*Proof:* Given the initial value of  $P$ , say  $P_0$ , we calculate  $Q_0$  by minimizing  $J(P_0, Q)$ .

Hence,

$$J(P_0, Q_0) \leq J(P_0, Q)$$

We calculate  $P_1$  by minimizing  $J(P, Q_0)$ . Therefore,

$$J(P_1, Q_0) \leq J(P_0, Q_0)$$

Since  $Q_1$  minimizes  $J(P_1, Q)$ , we have

$$J(P_1, Q_1) \leq J(P_1, Q_0) \leq J(P_0, Q_0).$$

Iteratively, we obtain

$$J(P_{t+1}, Q_{t+1}) \leq J(P_t, Q_t) \text{ for } t = 0, 1, 2, \dots$$

Since the SVD as shown in step-III of table I gives the optimal  $P^t$  that further reduces the objective value. Once we obtained optimal value of  $P$  and  $Q$ , the next iteration further converge  $P$  to local optima.

Hence, the sequence  $J(P_t, Q_t)$  is monotonically decreasing. Thus, by the Monotonic Convergence Theorem,  $J(P_t, Q_t)$  converges to a local optimal value.

$$\sum_{j=1}^N \text{tr} [(Z^T - P_\infty Q_\infty^T Z^T) (Z - Z Q_\infty P_\infty^T)] + \lambda \text{tr} (Q_\infty^T Q_\infty)$$

■

## III. DISCUSSION AND ANALYSIS

In order to evaluate the performance of proposed robust joint sparse PCA, in this section we have discussed and compared its performance on six popular image dataset including AR, ORL FERET, Yale, COIL20 and CMU PIE and compared its performance with unsupervised methods including PCA, PCAL1, JSPCA, 2DPCA, R2DPCA, PCA2DL1 and PCA2DL1-S on corrupted and non-corrupted benchmark datasets. In order to compare dimensionality reduction performance objectively and persuasively we evaluated using nearest neighbour classifier. Experiment is divided into two groups: Experiment-I is on non-contaminated(original) dataset and experiment-II is on contaminated datasets to validate the robustness against outliers. We have corrupted small portion of all dataset by adding random noise (salt and pepper or block of different sizes) as shown in figure 1. It shows some original images in first two rows and third row is corrupted by random blocks whereas fourth row is contaminated by 10% and 15% salt and pepper noise.

### A. Datasets

In order to evaluate the performance of proposed robust joint sparse PCA, in this section we have discussed and compared its performance on six popular image dataset including AR, ORL FERET, Yale, COIL20 and CMU PIE. Furthermore, we have contaminated the dataset to test the robustness against outliers, we have randomly selected 20% images to add noise in the datasets. We corrupted the dataset i.e. random noise well as block occlusions. Random noise is salt and pepper noise spread randomly at 10%, 15% on random selection of images from dataset as shown in figure 1. Similarly, block occlusion is added block of different sizes at random locations with variable size 5x5, 10x10, 10x15 as shown in figure 1. For evaluation on contaminated datasets, we have selected 60% and 70% and 80% samples per individual for each dataset as training dataset (3,5 and 7 ; 8, 13 and 18 ; 22, 27 and 32 ; for ORL, AR and Yale datasets respectively). We have conducted various number of experiments on each dataset and average classification accuracy is computed as shown in figure 3 , 4 and table II and table III .

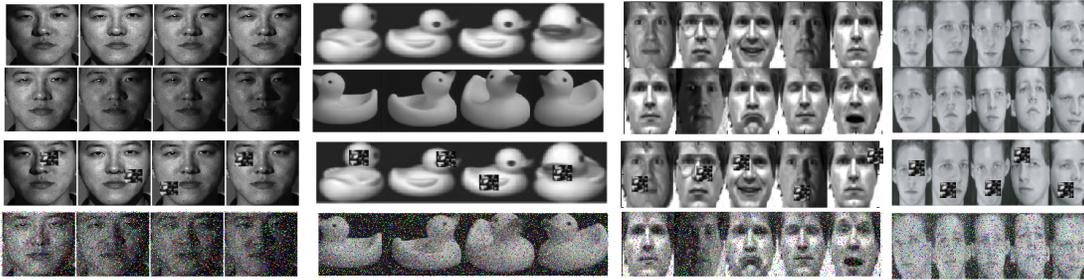


Fig. 1. Sample images of CMU PIE, COIL20, Yale [11] and AR [7] First two rows real dataset, Row 3 contaminated with block and Row 4 is contaminated with salt and paper noise 15%

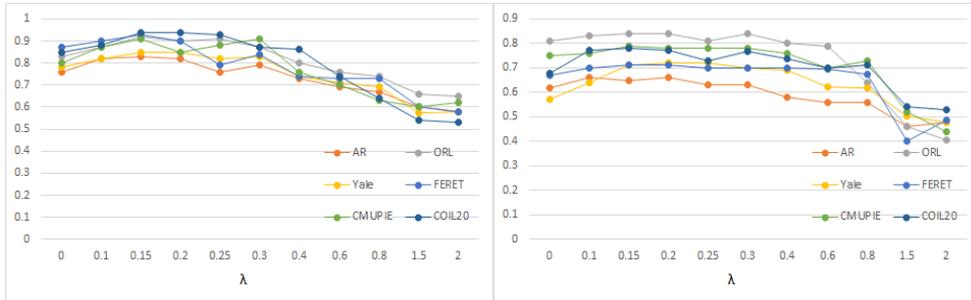


Fig. 2. Comparative evaluation at different value of  $\lambda$  for real and contaminated datasets

TABLE II. COMPARATIVE EVALUATION BASED ON AVERAGE CLASSIFICATION ACCURACY ON REAL DATASET AT OPTIMAL RESULT OF 2DJSPCA

Dataset	PCA	RPCA	SPCA	JSPCA	2DPCA	PCA2DL1	2DJSPCA
AR	0.6832 $\pm$ 0.005	0.6459 $\pm$ 0.008	0.7345 $\pm$ 0.0221	0.7896 $\pm$ 0.006	0.7589 $\pm$ 0.0071	0.8477 $\pm$ 0.0023	0.8482 $\pm$ 0.003
ORL	0.7891 $\pm$ 0.0028	0.8009 $\pm$ 0.0091	0.8322 $\pm$ 0.0011	0.8981 $\pm$ 0.0032	0.8843 $\pm$ 0.0411	0.8637 $\pm$ 0.0071	0.931 $\pm$ 0.0091
Yale	0.6886 $\pm$ 0.0031	0.5976 $\pm$ 0.0061	0.5723 $\pm$ 0.0009	0.7563 $\pm$ 0.0021	0.7911 $\pm$ 0.0091	0.7305 $\pm$ 0.0071	0.8721 $\pm$ 0.0531
FERET	0.8400 $\pm$ 0.0039	0.8322 $\pm$ 0.0039	0.8409 $\pm$ 0.0014	0.9222 $\pm$ 0.0022	0.9112 $\pm$ 0.0042	0.8900 $\pm$ 0.0022	0.9469 $\pm$ 0.0009
CMU PIE	0.7445 $\pm$ 0.0091	0.7666 $\pm$ 0.0027	0.8334 $\pm$ 0.0091	0.9011 $\pm$ 1.0011	0.8987 $\pm$ 0.0026	0.8607 $\pm$ 0.0015	0.939 $\pm$ 0.0041
COIL20	0.7923 $\pm$ 0.0023	0.7523 $\pm$ 0.0044	0.7744 $\pm$ 0.0012	0.8587 $\pm$ 0.0042	0.8639 $\pm$ 0.0036	0.8688 $\pm$ 0.0032	0.932 $\pm$ 0.0007

TABLE III. COMPARATIVE EVALUATION BASED ON AVERAGE CLASSIFICATION ACCURACY ON CONTAMINATED DATASET AT OPTIMAL RESULT OF 2DJSPCA

Dataset	PCA	RPCA	SPCA	JSPCA	2DPCA	PCA2DL1	2DJSPCA
AR	0.5741 $\pm$ 0.0023	0.5387 $\pm$ 0.0022	0.6178 $\pm$ 0.0091	0.6481 $\pm$ 0.0091	0.6576 $\pm$ 0.0049	0.6277 $\pm$ 0.0053	0.6711 $\pm$ 0.0203
ORL	0.6385 $\pm$ 0.0012	0.7411 $\pm$ 0.00321	0.8201 $\pm$ 0.0081	0.7181 $\pm$ 0.0087	0.8161 $\pm$ 0.0094	0.838 $\pm$ 0.0021	0.8511 $\pm$ 0.00221
Yale	0.5153 $\pm$ 0.0034	0.4865 $\pm$ 0.0083	0.5177 $\pm$ 0.0073	0.5978 $\pm$ 0.0065	0.5983 $\pm$ 0.0043	0.621 $\pm$ 0.0091	0.7300 $\pm$ 0.0091
FERET	0.6831 $\pm$ 0.0029	0.6948 $\pm$ 0.0042	0.59851 $\pm$ 0.0065	0.7186 $\pm$ 0.0043	0.6771 $\pm$ 0.0054	0.6184 $\pm$ 0.0087	0.7413 $\pm$ 0.0065
CMU PIE	0.577 $\pm$ 0.0032	0.5981 $\pm$ 0.0007	0.6771 $\pm$ 0.0054	0.6987 $\pm$ 1.00054	0.7181 $\pm$ 0.0091	0.6886 $\pm$ 0.0083	0.754 $\pm$ 0.0088
COIL20	0.5743 $\pm$ 0.0024	0.6081 $\pm$ 0.0032	0.7179 $\pm$ 0.0049	0.7474 $\pm$ 0.0076	0.7144 $\pm$ 0.0077	0.7786 $\pm$ 0.0053	0.7911 $\pm$ 0.0141

## B. Evaluation

In order to compare the dimensional reduction of different methods objectively and persuasively, we have used nearest neighbour to obtain classification accuracy. We have repeated experiment on each dataset ten times and average evaluation results shows that as classifier, 2DR-JSPCA achieved better accuracy as compared to JSPCA [17], SPCA [2], 2DRDPCA [13] and 2DPCAL1-S [12] as shown in table II, III and figure 3, 4. In first experiment, we have selected datasets with original but re-sized to 32x32. Table II shows the variation of classification accuracy with different subspace dimensionality at optimal  $\lambda = 0.18$ . For evaluation purpose, we have selected 60% and 70% and 80% samples per individual for each dataset as training dataset and rest of the datasets for validation. Due to the complexity of datasets such as illumination, variations and occlusions etc, it is quite challenging to obtain high classi-

fication accuracy, however, the experimental results shows that 2DR-JSPCA obtained better classification accuracy as compare to the PCA, 2DPCA, SPCA and JSPCA. It is due to sparsity in two dimension, selection of robust features as well as discarding the redundant patterns. Furthermore, it enables the 2DR-JSPCA has more freedom to learn low dimensional space that approximate to high dimensional data in flexible manner. In addition, regularization term  $\|Q\|_F^2$  is convex and can easily be optimized as it can gradually trending to smaller value iteratively. Moreover it reduces the constraints and enables our method to jointly select features.

To investigate the performance of 2D joint sparse PCA, 25% of dataset are contaminated with .e. random noise and block occlusions. Rectangular noise located at different position of different size (10x10 and 20 x 20) is added as shown in figure 1 whereas random noise is salt and pepper noise

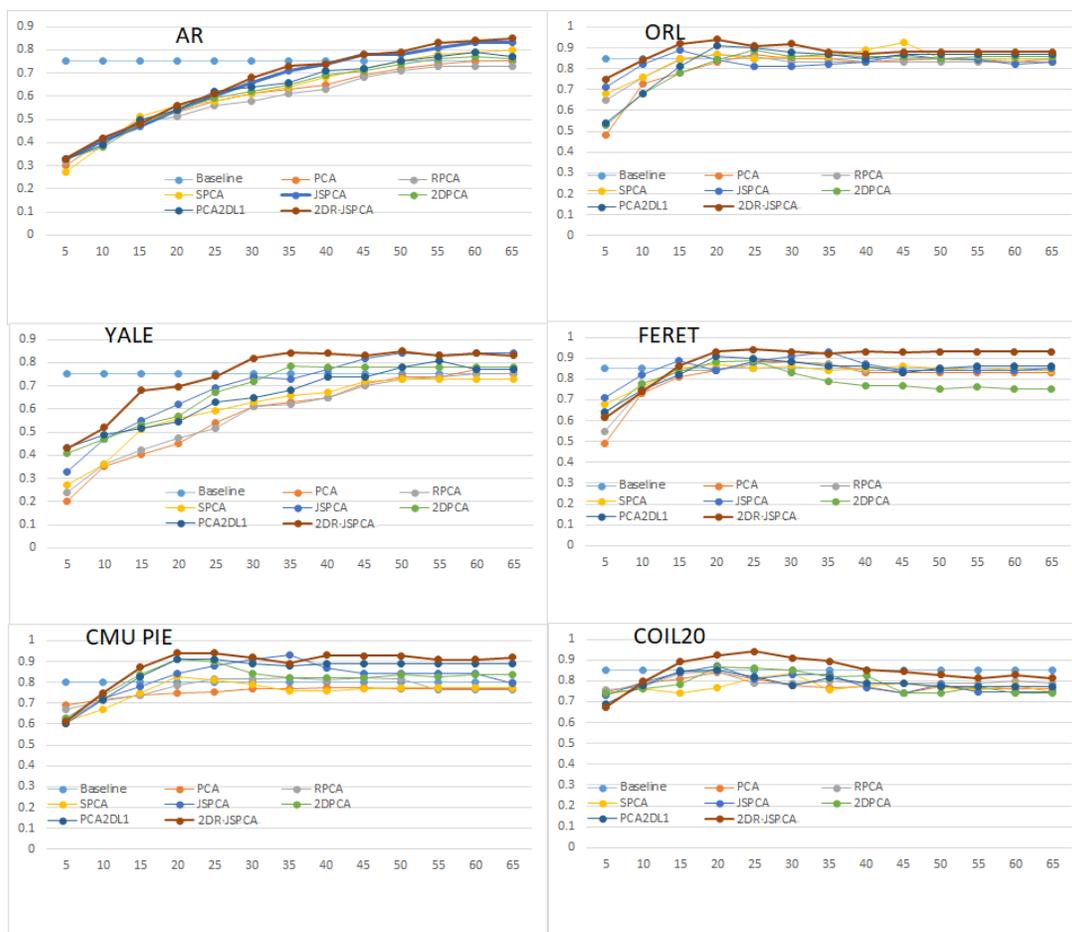


Fig. 3. Comparison: Classification accuracy on real dataset

spread randomly at 10%, 15% on random selection of images from datasets. After dataset corruption, we have selected 70% of corrupted images for training and rest of the images are part of validation datasets. Results shows that 2DR-JSPCA performed well for corrupted data as compare to other PCA based methods as shown in figure 4 and table III, however, it suffers from random corruption due to its feature selection ability. In conclusion, we can say that 2DR-JSPCA is robust to slight variations rather than random variations in the datasets.

2DR-JSPCA provided poor in term of reconstruction as compared to non-sparse methods due to the loss of extensive information, however, in comparison to sparse methods, 2DR-JSPCA reconstruction is better and able to select those features for that are effective for reconstruction. In conclusion, we can say that 2DR-JSPCA finds the representative features from high dimensional space that are used for classification. Results showed that 2DJSPCA out perform the other PCA based methods especially SPCA and Joint SPCA in term of classification. Traditional methods are not able to to interpret new feature whereas it is quite important to interpret new features especially when they have spatial meaning. 2DR-JSPCA outperform state of the art sparse methods Joint sparse PCA as it selected features by maintaining the images spatial structural information.

#### IV. CONCLUSION

In this paper, we presented new subspace learning methods robust joint sparse two dimensional principal component analysis by removing the mean automatically and relaxing the orthogonal constraints of transformation matrix and imposing a penalty function on regularization term. Both mathematical analysis and empirical results proved that 2DR-JSPCA with optimal mean has more freedom to jointly select the useful features from low-dimensional representation and robust against outliers as well as reduces the data dimensionality efficiently, moreover addition of penalty function into objective function results in robustness against outliers. Evaluation results show the improvement in effectiveness of 2DR-JSPCA for image reconstruction and classification. In conclusion, the numerical results suggest that our method is superior to previous approaches.

#### REFERENCES

- [1] J Paul Brooks, JH Dulá, and Edward L Boone. A pure  $l_1$ -norm principal component analysis. *Computational statistics & data analysis*, 61:83–98, 2013.
- [2] Alexandre d’Aspremont, Laurent E Ghaoui, Michael I Jordan, and Gert R Lanckriet. A direct formulation for sparse pca using semidefinite programming. In *Advances in neural information processing systems*, pages 41–48, 2005.

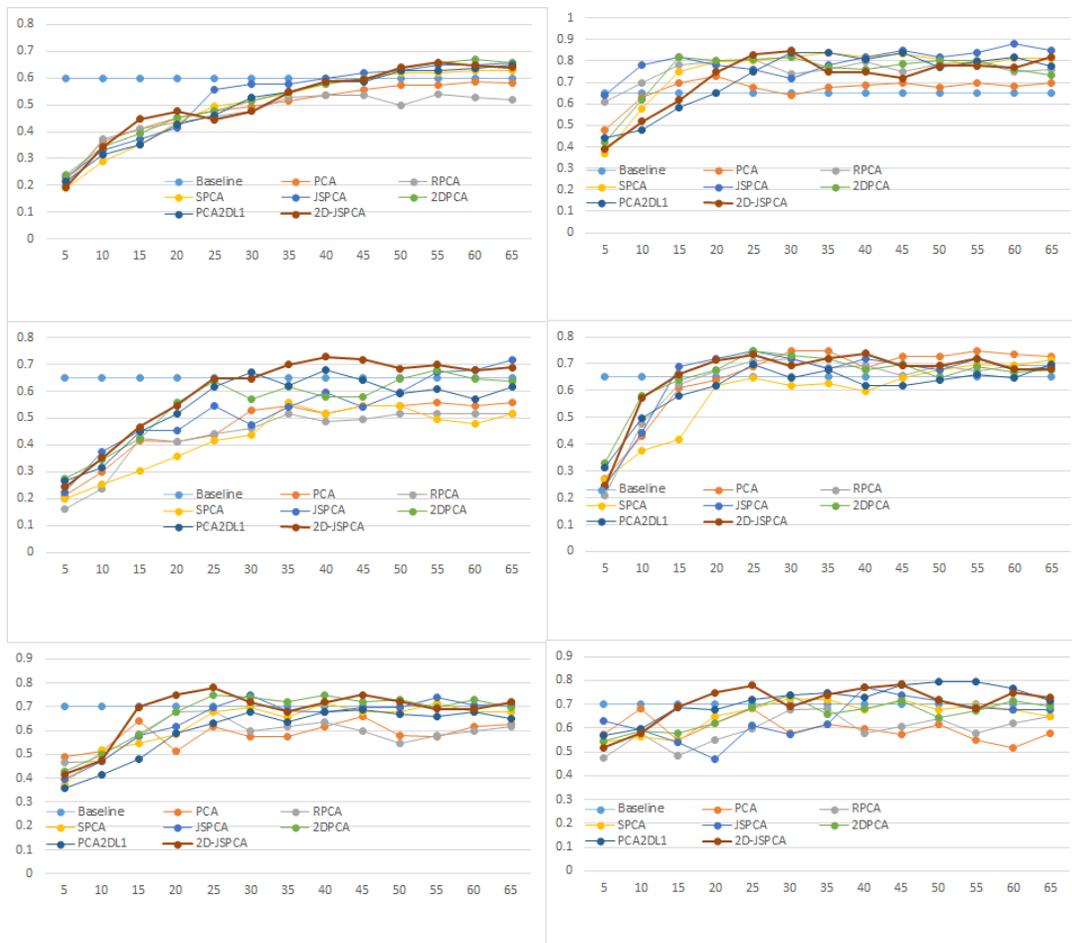


Fig. 4. Comparison: Classification accuracy on contaminated datasete

- [3] Chris Ding, Ding Zhou, Xiaofeng He, and Hongyuan Zha. R 1-pca: rotational invariant 1 1-norm principal component analysis for robust subspace factorization. In *Proceedings of the 23rd international conference on Machine learning*, pages 281–288. ACM, 2006.
- [4] Zohaib Khan, Faisal Shafait, and Ajmal Mian. Joint group sparse pca for compressed hyperspectral imaging. *IEEE Transactions on Image Processing*, 24(12):4934–4942, 2015.
- [5] Nojun Kwak. Principal component analysis based on l1-norm maximization. *IEEE transactions on pattern analysis and machine intelligence*, 30(9):1672–1680, 2008.
- [6] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetianopoulos. Mpc: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks*, 19(1):18–39, 2008.
- [7] Aleix M Martínez and Avinash C Kak. Pca versus lda. *IEEE transactions on pattern analysis and machine intelligence*, 23(2):228–233, 2001.
- [8] Feiping Nie, Heng Huang, Chris Ding, Dijun Luo, and Hua Wang. Robust principal component analysis with non-greedy l1-norm maximization. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 1433. Citeseer, 2011.
- [9] Muhammad Imran Razzak, Muhammad Khurram Khan, Khaled Alghathbar, and Rubiyah Yousaf. Face recognition using layered linear discriminant analysis and small subspace. In *Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on*, pages 1407–1412. IEEE, 2010.
- [10] Muhammad Imran Razzak, Raghieb Abu Saris, and Guandong Xu. Two dimensional joint sparse principal component analysis with f-norm minimization for sparse modelling: 2d-jspca. In *arXiv preprint arXiv:1708.05431*, 2017.
- [11] Terence Sim, Simon Baker, and Maan Bsat. The cmu pose, illumination, and expression (pie) database. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 53–58. IEEE, 2002.
- [12] Haixian Wang and Jing Wang. 2dpc: 2d-pca with l1-norm for simultaneously robust and sparse modelling. *Neural Networks*, 46:190–198, 2013.
- [13] Qianqian Wang and Quanxue Gao. Robust 2dpc and its application. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 79–85, 2016.
- [14] Qianqian Wang, Quanxue Gao, Xinbo Gao, and Feiping Nie. Optimal mean two-dimensional principal component analysis with f-norm minimization. *Pattern Recognition*, 68:286–294, 2017.
- [15] Jian Yang, David Zhang, Alejandro F Frangi, and Jing-yu Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence*, 26(1):131–137, 2004.
- [16] Jian Yang, David Zhang, Xu Yong, and Jing-yu Yang. Two-dimensional discriminant transform for face recognition. *Pattern recognition*, 38(7):1125–1129, 2005.
- [17] Shuangyan Yi, Zhihui Lai, Zhenyu He, Yiu-ming Cheung, and Yang Liu. Joint sparse principal component analysis. *Pattern Recognition*, 61:524–536, 2017.