AUTO-WEIGHTED TWO-DIMENSIONAL PRINCIPAL COMPONENT ANALYSIS WITH ROBUST OUTLIERS

Rui Zhang, Feiping Nie, and Xuelong Li*

School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL)
Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China.

ABSTRACT

Two-dimensional principal component analysis (2DPCA) serves as an efficient approach for both dimensionality reduction and high-quality reconstruction. However, conventional 2DPCA method is sensitive to the outliers such that associated results could be compromised. To strengthen the robustness of conventional 2DPCA method, we try to propose a novel robust two-dimensional principal component analysis with optimal mean (R2DPCA-OM) method to automatically achieve the optimal mean. Besides, the experimental results illustrate that the proposed R2DPCA-OM method could obtain the optimal subspaces and mean, such that dimensionality is reduced with less reconstruction error. Consequently, superiority and effectiveness of the proposed R2DPCA-OM method could be verified analytically and empirically.

Index Terms— principal component analysis, robustness, optimal mean.

1. INTRODUCTION

As for the high-quality reconstruction, principal component analysis (PCA)[1, 2, 3] surely serves as one out of many effective approaches [4, 5, 6, 7, 8], via which the dimensionality reduction could be achieved with minimizing the mean square error. To avoid coping with data matrix of large dimensionality, which is generated by re-shaping each tensor data into the vector form, two-dimensional principal component analysis known as 2DPCA [11, 12, 13, 14] is proposed for the efficient computation. As a result, not only the dimensionality reduction can be achieved but also the significant statistical properties for the input data could be maintained under 2DPCA. Although 2DPCA is much more efficient than PCA in diverse aspects, classical 2DPCA is still sensitive to the outliers, such that related experimental results of 2DPCA could be erroneous.

To address the defect previously mentioned, a novel robust 2DPCA method is proposed to achieve the optimal mean automatically in this paper. Different from traditional 2DPCA method, the proposed method utilizes the robust 2DPCA with optimal mean (R2DPCA-OM) as the objective function, which is robust to the outliers. Besides, associated algorithm seeks the optimal mean in each iteration instead of traditional data preprocessing, such that input data is centralized. Moreover, the proposed R2DPCA-OM method has a self-adaptive weight, which assigns the smaller weight to the term with larger outliers automatically to promote the robustness. Furthermore, the proposed R2DPCA-OM method could be extended into the capped R2DPCA-OM method to deal with even more ill-defined situation.

*Corresponding author. Email: feipingnie@gmail.com.

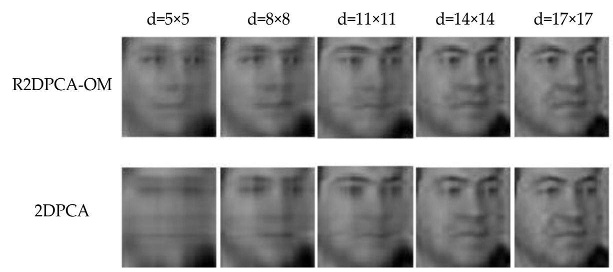


Fig. 1. The comparisons of data reconstruction are performed for R2DPCA-OM and 2DPCA[11] over the dataset FERET, where d is the reduced dimensionality.

2. ROBUST 2DPCA WITH OPTIMAL MEAN

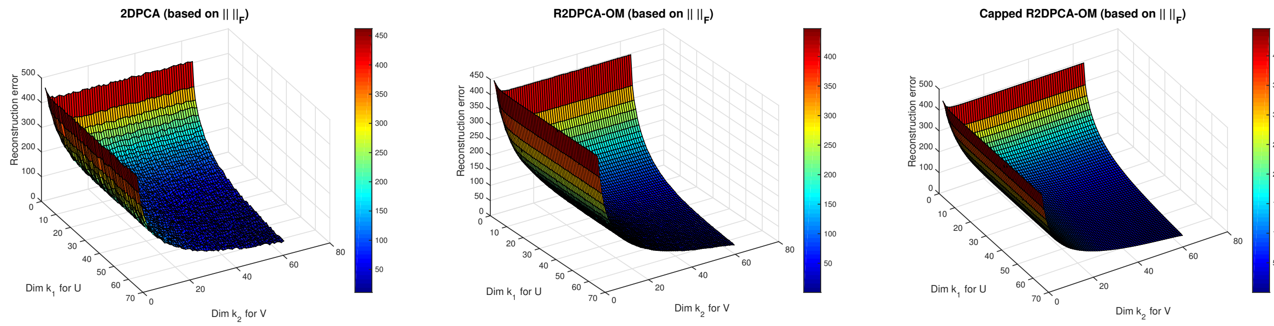
With centered tensor sample $\chi_{\mathbf{i}} \in \mathbb{R}^{\mathbf{I_1} \times \mathbf{I_2} \times \ldots \times \mathbf{I_N}}, (\mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{m})$, multilinear PCA (MPCA)[15, 16] could be illustrated as

$$\min_{\mathbf{S_i}, \mathbf{U^{(j)}}} \sum_{\mathbf{i=1}}^{\mathbf{m}} \|\chi_{\mathbf{i}} - \mathbf{S_i} \times_{\mathbf{1}} \mathbf{U^{(1)}} \times_{\mathbf{2}} \mathbf{U^{(2)}} \times_{\mathbf{3}} \ldots \times_{\mathbf{N}} \mathbf{U^{(N)}}\|^{\mathbf{2}} \tag{1}$$

where $\mathbf{U^{(j)}}, (\mathbf{j} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{N})$ is the orthonormal projection and $\mathbf{S_i}, (\mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{m})$ is the core tensor.

Apparently, two-dimensional PCA (2DPCA)[11] serves as a special case ($\mathbf{N} = \mathbf{2}$) of MPCA in (1). Accordingly, the

978-1-5090-4117-6/17/$31.00 ©2017 IEEE



(a) 2DPCA (b) R2DPCA-OM(our) (c) Capped 2DPCA-OM(our)

Fig. 2. The comparisons of reconstruction error are performed for 2DPCA [11], R2DPCA-OM and capped R2DPCA-OM under the noised data of dataset FEI.

Table 1. The reconstruction error under 2 different measures.

YALE($\times 10^3$)	The measure $\mathbf{m} = \|\cdot\|_*$		
$\mathbf{k_1 \times k_2}$	40×25	8×60	80×40
capped R2DPCA-OM	**5.0499**	**6.3413**	2.8597
R2DPCA-OM	**5.0499**	6.3638	**2.8561**
2DPCA[11]	5.3381	6.3653	3.3154

UMIST($\times 10^3$)	The measure $\mathbf{m} = \|\cdot\|_\mathbf{F}$		
$\mathbf{k_1 \times k_2}$	30×30	35×55	65×60
capped R2DPCA-OM	2.6707	**2.0671**	**0.5798**
R2DPCA-OM	**2.6686**	2.0939	0.5811
2DPCA[11]	2.6736	2.1050	0.6654

2DPCA problem can be represented as

$$\min_{\mathbf{M},\mathbf{B_i},\mathbf{U},\mathbf{V}} \sum_{\mathbf{i=1}}^{\mathbf{l}} \|\mathbf{A_i} - \mathbf{M} - \mathbf{U}\mathbf{B_i}\mathbf{V^T}\|_\mathbf{F}^\mathbf{2}$$

$$\mathbf{s.t.}\quad \mathbf{U^T U} = \mathbf{I_{k_1}} \ \ \mathbf{and} \ \ \mathbf{V^T V} = \mathbf{I_{k_2}} \tag{2}$$

where $\mathbf{M} \in \mathbb{R}^{\mathbf{m}\times\mathbf{n}}, \mathbf{B_i} \in \mathbb{R}^{\mathbf{k_1}\times\mathbf{k_2}}, \mathbf{U} \in \mathbb{R}^{\mathbf{m}\times\mathbf{k_1}}$ and $\mathbf{V} \in \mathbb{R}^{\mathbf{n}\times\mathbf{k_2}}$ are the associated variables with the data point $\mathbf{A_i} \in \mathbb{R}^{\mathbf{m}\times\mathbf{n}}, (\mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathbf{l})$. Besides, $\mathbf{M}$ serves as mean matrix, while $\mathbf{U}$ and $\mathbf{V}$ serve as the projection matrices in Eq. (2). Since $\mathbf{B_i}$ is free from any constraint, we could achieve the extreme value condition w.r.t. $\mathbf{B_i}$ in (2) as

$$\frac{\partial \sum_{\mathbf{i=1}}^{\mathbf{l}} \|\mathbf{A_i} - \mathbf{M} - \mathbf{U}\mathbf{B_i}\mathbf{V^T}\|_\mathbf{F}^\mathbf{2}}{\partial \mathbf{B_i}} = \mathbf{0}$$

$$\Rightarrow \mathbf{B_i} = \mathbf{U^T}(\mathbf{A_i} - \mathbf{M})\mathbf{V}. \tag{3}$$

By virtue of Eq. (3), problem (2) can be simplified into

$$\min_{\mathbf{M},\mathbf{U},\mathbf{V}} \sum_{\mathbf{i=1}}^{\mathbf{l}} \|\mathbf{A_i} - \mathbf{M} - \mathbf{U}\mathbf{U^T}(\mathbf{A_i} - \mathbf{M})\mathbf{V}\mathbf{V^T}\|_\mathbf{F}^\mathbf{2}$$

$$\mathbf{s.t.}\quad \mathbf{U^T U} = \mathbf{I_{k_1}} \ \ \mathbf{and} \ \ \mathbf{V^T V} = \mathbf{I_{k_2}}. \tag{4}$$

Given the possible situation that the outliers might be large for certain $\mathbf{i}$-th term in (4), current 2DPCA not only could not mitigate related situation but might worsen it as well. Therefore, we will try to enhance the robustness of 2DPCA in (4). Accordingly, robust 2DPCA problem can be represented as

$$\min_{\mathbf{M},\mathbf{U},\mathbf{V}} \sum_{\mathbf{i=1}}^{\mathbf{l}} \|\mathbf{A_i} - \mathbf{M} - \mathbf{U}\mathbf{U^T}(\mathbf{A_i} - \mathbf{M})\mathbf{V}\mathbf{V^T}\|_\mathbf{F}$$

$$\mathbf{s.t.}\quad \mathbf{U^T U} = \mathbf{I_{k_1}} \ \ \mathbf{and} \ \ \mathbf{V^T V} = \mathbf{I_{k_2}}. \tag{5}$$

Motivated by [3] and [17], robust 2DPCA with optimal mean (R2DPCA-OM) problem can be proposed as the following re-weighted form to solve the robust 2DPCA problem in (5)

$$\min_{\mathbf{M},\mathbf{U},\mathbf{V}} \sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}\|\mathbf{A_i} - \mathbf{M} - \mathbf{U}\mathbf{U^T}(\mathbf{A_i} - \mathbf{M})\mathbf{V}\mathbf{V^T}\|_\mathbf{F}^\mathbf{2}$$

$$\mathbf{s.t.}\quad \mathbf{U^T U} = \mathbf{I_{k_1}} \ \ \mathbf{and} \ \ \mathbf{V^T V} = \mathbf{I_{k_2}} \tag{6}$$

where the weight $\mathbf{w_i} \leftarrow \frac{\mathbf{1}}{\mathbf{2}\|\mathbf{A_i} - \mathbf{M} - \mathbf{U}\mathbf{U^T}(\mathbf{A_i} - \mathbf{M})\mathbf{V}\mathbf{V^T}\|_\mathbf{F}}$ is to be updated iteratively in the algorithm 1. The major superiority of the proposed re-weighted form as R2DPCA-OM in (6) is connected with the self-adaptive weight. In other words, the smaller weight would be assigned to the term with larger outliers automatically.

Besides, the objective function $\mathbf{J}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ of the problem (6) could further be expanded into

$$\mathbf{J}(\mathbf{M}, \mathbf{U}, \mathbf{V}) = \sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}\mathbf{Tr}(\mathbf{A_i^T}\mathbf{A_i} + \mathbf{M^T}\mathbf{M} - \mathbf{2}\mathbf{M^T}\mathbf{A_i} - (\mathbf{A_i} - \mathbf{M})^\mathbf{T}\mathbf{U}\mathbf{U^T}(\mathbf{A_i} - \mathbf{M})\mathbf{V}\mathbf{V^T}). \tag{7}$$

Based on the result in (7) and the extreme value condition

Table 2. The reconstruction error under the dataset FERET.

The measure $m = \|\cdot\|_F$ $(\times 10^3)$				
Dimensionality	2×2	4×4	6×6	8×8
R2DPCA-OM	**1.0128**	**0.8605**	**0.7295**	**0.6364**
2DPCA[11]	1.0913	0.9254	0.7642	0.6509

w.r.t. $\mathbf{M}$, we have

$$\frac{\partial \mathbf{J}(\mathbf{M},\mathbf{U},\mathbf{V})}{\partial \mathbf{M}} = \mathbf{0}$$
$$\Rightarrow \sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}(\mathbf{M}-\mathbf{A_i}) = \mathbf{U}\mathbf{U^T}\mathbf{R}\mathbf{V}\mathbf{V^T} = \mathbf{U}\mathbf{N_1}\mathbf{V^T}.$$

On the other hand, the term as $\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}(\mathbf{M}-\mathbf{A_i})$ can be disintegrated into

$$\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}(\mathbf{M}-\mathbf{A_i}) = \mathbf{U}\mathbf{N_1}\mathbf{V^T} + \overline{\mathbf{U}}\mathbf{N_2}\mathbf{V^T} + \mathbf{U}\mathbf{N_3}\overline{\mathbf{V}}^\mathbf{T} + \overline{\mathbf{U}}\mathbf{N_4}\overline{\mathbf{V}}^\mathbf{T} \quad (8)$$

where $\overline{\mathbf{U}}$ and $\overline{\mathbf{V}}$ are the orthogonal complement spaces for $\mathbf{U}$ and $\mathbf{V}$ respectively with $\mathbf{N_i}, (\mathbf{i}=\mathbf{1},\mathbf{2},\mathbf{3},\mathbf{4})$ being the coefficient matrix. Based on Eq. (7) and (8), $\mathbf{M}$ can be derived as

$$\overline{\mathbf{U}}\mathbf{N_2}\mathbf{V^T} + \mathbf{U}\mathbf{N_3}\overline{\mathbf{V}}^\mathbf{T} + \overline{\mathbf{U}}\mathbf{N_4}\overline{\mathbf{V}}^\mathbf{T} = \mathbf{0}$$
$$\Rightarrow \mathbf{M} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}\mathbf{A_i}}{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}} + \mathbf{U}\mathbf{N}\mathbf{V^T}$$

where $\mathbf{N} \in \mathbb{R}^{\mathbf{k_1}\times\mathbf{k_2}}$ is an arbitrary constant matrix. By substituting $\mathbf{M} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}\mathbf{A_i}}{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}} + \mathbf{U}\mathbf{N}\mathbf{V^T}$, Eq. (6) can be reformulated into

$$\min_{\mathbf{M},\mathbf{U^T}\mathbf{U}=\mathbf{I_{k_1}},\mathbf{V^T}\mathbf{V}=\mathbf{I_{k_2}}} \sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i} \|\mathbf{A_i} - \frac{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}\mathbf{A_i}}{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}} - \mathbf{U}\mathbf{U^T}(\mathbf{A_i} - \frac{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}\mathbf{A_i}}{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}})\mathbf{V}\mathbf{V^T}\|_\mathbf{F}^\mathbf{2} \quad (9)$$

where both $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices. Based on above result, we notice that Eq. (9) does not depend on the coefficient matrix $\mathbf{N}$. Hence, we could choose $\mathbf{N}$ as null matrix for the convenience, such that the optimal mean $\mathbf{M}$[1] can be delivered as $\mathbf{M} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}\mathbf{A_i}}{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}}$. Based on Eq. (9) and the optimal mean above, Eq. (6) can be further rewritten as

$$\max_{\mathbf{U},\mathbf{V}} \sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}\mathbf{Tr}(\mathbf{U^T}(\mathbf{A_i}-\mathbf{M})\mathbf{V}\mathbf{V^T}(\mathbf{A_i}-\mathbf{M})^\mathbf{T}\mathbf{U})$$
$$\mathbf{s.t.} \quad \mathbf{U^T}\mathbf{U} = \mathbf{I_{k_1}} \ \mathbf{and} \ \mathbf{V^T}\mathbf{V} = \mathbf{I_{k_2}} \quad (10)$$

[1]Moreover, the optimal mean for the problem (2) can be easily gained by substituting $\mathbf{w_i} = \mathbf{1}$ as $\mathbf{M} = \frac{1}{\mathbf{l}}\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{A_i}$.

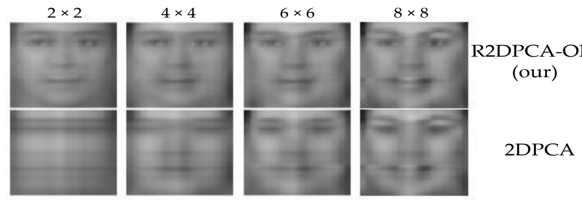


Fig. 3. Samples of the reconstructed images under the table 2.

Input: $\mathbf{A_i} \in \mathbb{R}^{\mathbf{m}\times\mathbf{n}}, (\mathbf{i}=\mathbf{1},\mathbf{2},\ldots,\mathbf{l})$.
Output: $\mathbf{Y} = \mathbf{U^T}\mathbf{C}\mathbf{V}$ represents the projection, which reduces the dimensionality of $\mathbf{C}$ to $\mathbb{R}^{\mathbf{k_1}\times\mathbf{k_2}}$ for any given data $\mathbf{C} \in \mathbb{R}^{\mathbf{m}\times\mathbf{n}}$.

```
Initialize w_i = 1, (i = 1, 2, ..., l) and VV^T = I_n;
while not converge do
    Update M ← Σ_{i=1}^l w_i A_i / Σ_{i=1}^l w_i;
    Update P_1 ← Σ_{i=1}^l w_i (A_i − M) V V^T (A_i − M)^T;
    Update U ← arg max_{U^T U = I_{k1}} Tr(U^T P_1 U);
    Update P_2 ← Σ_{i=1}^l w_i (A_i − M)^T U U^T (A_i − M);
    Update V ← arg max_{V^T V = I_{k2}} Tr(V^T P_2 V);
    for i = 1 : n do
        Update w_i ← 1 / (2||A_i − M − U U^T (A_i − M) V V^T||_F);
    end
end
return U ∈ R^{m×k1} and V ∈ R^{n×k2};
```

Algorithm 1: Robust 2DPCA with optimal mean (R2DPCA-OM) method under the problem (6)

where $\mathbf{M} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}\mathbf{A_i}}{\sum_{\mathbf{i=1}}^{\mathbf{l}} \mathbf{w_i}}$.

According to Eq. (10), the R2DPCA-OM method can be proposed in the algorithm 1 correspondingly.

• **Extension of the robust 2DPCA problem in (5).** Given the possible situation that the outliers might be extraordinarily huge for certain $\mathbf{i}$-th term in (5), the superiority of the proposed robust problem in (5) and (6) might be largely compromised. Actually, we could address this situation by introducing the capped form of the problem (5) as

$$\min_{\mathbf{M},\mathbf{U},\mathbf{V}} (\sum_{\mathbf{i=1}}^{\mathbf{l}} \min(\|\mathbf{A_i}-\mathbf{M}-\mathbf{U}\mathbf{U^T}(\mathbf{A_i}-\mathbf{M})\mathbf{V}\mathbf{V^T}\|_\mathbf{F}, \varepsilon))$$
$$\mathbf{s.t.} \quad \mathbf{U^T}\mathbf{U} = \mathbf{I_{k_1}} \ \mathbf{and} \ \mathbf{V^T}\mathbf{V} = \mathbf{I_{k_2}} \quad (11)$$

where ε is the threshold parameter. We could observe that if the outliers of certain i-th term in (11) is very large, Eq. (11) would automatically replace the related term by the threshold

Table 3. The average recognition rate comparisons.

Dataset	Method	Recognition rate on the original data $\mathbf{X_i^o}$	Recognition rate on the noised data $\mathbf{X_i^o} + \mathbf{X_i^n}$
AT&T	PCA[1]	88.00±0.56%	80.55±0.81%
	RPCA-OM[3]	91.05±0.52%	87.61±0.58%
	2DPCA[11]	90.64±0.94%	82.93±0.70%
	R2DPCA-OM(our)	**92.13**±0.48%	88.32±0.34%
	Capped R2DPCA-OM(our)	91.78±0.44%	**88.45**±0.32%
USPS	PCA[1]	90.63±0.58%	84.01±0.86%
	RPCA-OM[3]	92.28±0.41%	89.25±0.48%
	2DPCA[11]	91.86±0.60%	85.74±0.96%
	R2DPCA-OM(our)	93.30±0.83%	**89.56**±0.54%
	Capped R2DPCA-OM(our)	**93.44**±0.75%	89.30±0.33%

ε. In other words, the capped R2DPCA-OM in (11) could avoid the ill-defined situation mentioned above. Besides, solving the problem (11) is basically the same as solving the problem (5) as previously mentioned. The only difference reflects on a novel threshold-sensitive weight, which is introduced to the proposed R2DPCA-OM in (6) as

$$\mathbf{w_i} = \frac{\mathbf{Ind}}{\mathbf{2\|A_i - M - UU^T(A_i - M)VV^T\|_F}} \tag{12}$$

where the indicative function Ind is defined as

$$\mathbf{Ind} = \begin{cases} \mathbf{1}, \ \mathbf{\|A_i - M - UU^T(A_i - M)VV^T\|_F} \leq \varepsilon \\ \mathbf{0}, \ Otherwise \end{cases}.$$

Equipped with the weight defined in (12), the algorithm 1 could be extended to unraveling the capped robust 2DPCA in (11) correspondingly.

3. EXPERIMENT

In this section, numerical experiments are performed to verify the effectiveness of the proposed approaches.

3.1. Reconstruction Error Comparison

Firstly, we compare the proposed robust 2DPCA with optimal mean (R2DPCA-OM) and capped R2DPCA-OM with the 2DPCA approach[11] on three datasets as FEI, UMIST and YALE via the reconstruction error. We randomly select 25% of each dataset and set 20% size of the selected images with Gaussian noise to compare the reconstruction error represented by $\sum_\mathbf{i} \mathbf{m}(\mathbf{X_i^o} - \mathbf{X_i^r})$, where $\mathbf{X_i^o}$ is the original image and $\mathbf{X_i^r}$ is the reconstructed data. Moreover, the measure $\mathbf{m}$ is chosen as both $\|\cdot\|_\mathbf{F}$ and $\|\cdot\|_*$ to ensure a just comparison.

1. From the figure 1 and 3, we can observe that the R2DPCA-OM approach gain the reconstructed data with uniformly better quality compared to the 2DPCA[11] method under the same reduced dimensionality.

2. From the figure 2 and the table 1, reconstruction errors of the proposed R2DPCA-OM and capped R2DPCA-OM methods are consistently less than that of the 2DPCA method[11] under both measures.

3. From the figure 2, we could observe that surfaces of the proposed R2DPCA-OM and capped R2DPCA-OM methods are more smooth than that of the 2DPCA method, which represent stronger robustness to the outliers for the proposed approaches.

3.2. Recognition Rate Comparison

Finally, we compare the average recognition rates of PCA[1], RPCA-OM[3], 2DPCA[11], R2DPCA-OM and capped R2DP CA-OM in the table 3 based on both the original data $\mathbf{X_i^o}$ and the noised data $\mathbf{X_i^o} + \mathbf{X_i^n}$, where the noise term $\mathbf{X_i^n}$ is randomly generated. Besides, we utilize the datasets as AT&T and USPS, in which the training sets and the test sets are randomly chosen. Classification accuracy is further compared via shortest distance classifier.

1. From the table 3, R2DPCA-OM and capped R2DPCA-OM perform consistently better than other three approaches on the recognition rate especially under the noised data due to their robustness.

4. CONCLUSION

In this paper, we propose and investigate the robust two-dimensional principal component analysis with optimal mean. With seeking the optimal mean in each iteration, both the R2DPCA-OM and the capped R2DPCA-OM methods achieve the optimal subspaces with less reconstruction errors. Via the application of both proposed approaches, the dimensionality can be reduced with high-quality reconstruction. Eventually, the effectiveness and the superiority of the proposed R2DPCA-OM and capped R2DPCA-OM methods are verified empirically and analytically.

5. REFERENCES

[1] I.T. Jolliffe, *Principal component analysis*, 2nd ed. Springer-Verlag, New York, NY, USA, 2002.

[2] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang, "Robust principal component analysis with non-greedy l_1-norm maximization," in *IJCAI*, pp. 1433-1438, 2011.

[3] F. Nie, J. Yuan, and H. Huang, "Optimal mean robust principal component analysis," in *Proc. ICML*, pp. 2755-2763, 2014.

[4] T.F. Cox and M.A. Cox, *Multidimensional scaling*, Chapman & Hall, London, England, 1994.

[5] J. Lee and M. Verleysen, *Nonlinear dimensionality reduction*, Springer, Berlin, Germany, 2007.

[6] F. Nie, S. Xiang, Y. Jia, and C. Zhang, "Semi-supervised orthogonal discriminant analysis via label propagation," *Pattern Recognition*, vol. 42, no. 11, pp. 2615-2627, 2009.

[7] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint $l_{2,1}$-norms minimization," in *NIPS*, 2010.

[8] F. Nie, H. Huang, and C. Ding, "Low rank matrix recovery via efficient schatten p-norm minimization," in *AAAI*, 2012.

[9] R. Devore, I. Daubechies, M. Fornasier, and C. Güntürk, "Iterative re-weighted least squares," *Comm. Pure. Appl. Math.*, vol. 44, no. 6, pp. 1-9, 2009.

[10] I. Daubechies, R. Devore, M. Fornasier, and C. Güntürk, "Iteratively re-weighted least squares minimization for sparse recovery," *Comm. Pure. Appl. Math.*, vol. 63, no. 1, pp. 1-38, 2008.

[11] J. Ye, "Generalized low rank approximations of matrices," in *Proc. ICML*, pp. 887-894, 2004.

[12] J. Yang, D. Zhang, A. Frangi, and J. Yang, "Two-dimensional PCA: a new approach to appearance-based face representation and recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131-137, 2004.

[13] X. Li, Y. Pang and Y. Yuan, "l_1-norm based 2DPCA," *IEEE Trans. on System, Man and Cybernetics, Part B*, vol. 40, no. 4, pp. 1170-1175, 2010.

[14] Y. Pang, D. Tao, Y. Yuan, and X. Li, "Binary two-dimensional PCA," *IEEE Trans. on System, Man and Cybernetics, Part B*, vol. 38, no. 4, pp. 1176-1180, 2008.

[15] H. Lu, K. Plataniotis, and A. Venetsanopoulos, "MPCA: multilinear principal component analysis of tensor objects," *IEEE Trans. on Neural Networks*, vol. 19, no. 1, pp. 18-39, 2008.

[16] H. Lu, K. Plataniotis, and A. Venetsanopoulos, "Multilinear principal component analysis of tensor objects for recognition," *International Conference on Pattern Recognition*, vol. 2, no. 1, pp. 776-779, 2006.

[17] G. Lan, C. Hou, and D. Yi. "Robust Feature Selection via Simultaneous Capped l_2-Norm and $l_{2,1}$-Norm Minimization," *IEEE International Conference on Big Data Analysis*, 2016.