Incomplete-Data Oriented Dimension Reduction via Instance Factoring PCA Framework

Ernest Domanaanmwi Ganaa[1,3], Timothy Apasiba Abeo[1,4], Sumet Mehta[1],
Heping Song[1], and Xiang-Jun Shen[1,2(✉)]

[1] School of Computer Science and Communication Engineering,
JiangSu University, Zhenjiang 212013, Jiangsu, China
xjshen@ujs.edu.cn
[2] Jingkou New-Generation Information Technology Industry Institute,
JiangSu University, Zhenjiang, China
[3] School of Applied Science and Technology, Wa Polytechnic, 553, Wa, Ghana
[4] School of Applied Science, Tamale Technical University, 3ER, Tamale, Ghana

Abstract. In this paper, we propose an instance factoring PCA (IFPCA) framework for dimension reduction in incomplete datasets. The advantage of IFPCA over the traditional PCA is that, a penalty is imposed on the instance space via a scaling-factor to suppress the effect of outliers in pursuing projections. We geometrically use two scaling-factor strategies, total distance and cosine similarity metrics. Both strategies can learn the relationship between each data point and the principal projection in the feature space. In this way, better low-rank projections are obtained through scaling the data iteratively to suppress the impact of noise in the training set. Extensive experiments on COIL-20, ORL and USPS datasets prove the superiority of the proposed framework over state-of-the-art dimensionality reduction methods such as LSDA, gLPCA, RPCA-OM, PCA, LPP and RCDA.

Keywords: Principal component analysis · Instance factoring · Incomplete-data · Dimensionality reduction · Manifold learning

1 Introduction

Massive data is generated daily such as through city-installations of high speed cameras for public safety. These data are mostly incomplete due to sensor failures or environmental obstructions in recordings. This poses a great challenge to algorithms of video surveillance in processing such missing, noisy and high dimensional datasets.

Many manifold learning methods have been proposed for dimensionality reduction (DR). These methods can broadly be divided into global structure and local structure learning. Local structure learning methods such as, Locality Sensitive Discriminant Analysis (LSDA) [1] finds a projection that maximizes the margin between data points from different classes at each local area. Neighborhood Preserving Embedding (NPE) [2] is an unsupervised linear dimensionality reduction technique which solves the out of sample problem in Locally Linear Embedding (LLE). Locality Preserving

Projections (LPP) [3] finds a good linear embedding that preserves local structure information. And global structure learning methods, such as Linear Discriminant Analysis (LDA) [4] captures the global geometric structure of data by maximizing the between class distance and minimizing the within class distance. Isomap [5] is another global learning method which estimates the geodesic distance between samples and uses multidimensional scaling to induce a low dimensional manifold.

Among the global DR methods, PCA is the most popular, simplest and efficient technique [6]. But, it is sensitive to outliers or noisy data points [7–10]. Thus, several adaptations of PCA have been developed in the past few years to improve its performance. Representatives such as graph-Laplacian PCA (gLPCA) [11] learns a low dimensional representation of data that incorporates graph structures. Optimal mean robust principal component analysis (RPCA-OM) [12] removes the mean in a given dataset automatically by integrating the mean into the dimensionality reduction objective function. Abeo et al. [13] also extended the minimizing least squares idea of PCA to consider both data distribution and penalty weights in dealing with outliers. Yang et al. [14] estimated corrupt instances by making full use of both intra-view and inter-view correlations between samples, considering samples in the same view to be linearly correlated. Li et al. [15] proposed ordinal preserving projection (OPP) for learning to rank by using two matrices which work in the row and column directions respectively to leverage the global structure of the dataset and ordinal information of the observations. Most existing adaptations of the standard PCA learn how to select suitable features instead of suitable samples; because of this, corrupt instances have not be efficiently handled in the past. Research in DR is still being vigorously pursued by researchers to either improve the performance of existing techniques or develop new ones.

We propose a novel framework called incomplete-data oriented dimension reduction via instance factoring PCA framework (IFPCA) to address the sensitivity of PCA to corrupt instances. In this framework, a scaling-factor that imposes a penalty on the instance space is introduced to suppress the impact of outliers or corrupt instances in pursuing projections. Two strategies: total distance and cosine similarity metrics are used geometrically to iteratively learn the relationship between each instance and the principal projection in the feature space. Through this, the two strategies are able to distinguish between authentic and corrupt instances. Thus, low-rank projections are achieved through enhanced discrimination between relevant and noisy instances. The main contributions of this paper are summarized as follows:

1. We propose a novel framework by introducing a scaling-factor into the traditional PCA model to impose a penalty on the instance space in pursuing projections. The goal here is to significantly suppress the impact of outliers.
2. We further propose two scaling-factor strategies: total distance and cosine similarity metrics. These metrics iteratively evaluate the importance of each instance by learning the relationship between each instance and the principal projection in the feature space.
3. Finally, with the iterative discrimination ability, IFPCA can obtain better low-rank projections in incomplete datasets.

Extensive experiments on COIL-20, ORL and USPS datasets demonstrate the superiority of our method over state-of-the-art methods. The rest of the paper is organized

as follows: Sect. 2 presents formulation of the propose framework, experiments, results and complexity analysis are presented in Sect. 3, and conclusions and recommendations are made in Sect. 4.

2   The Proposed IFPCA Method

To illustrate our idea, we start by observing the objective function of PCA:

$$\min_{w^T w=1} \sum_{i=1}^{n} (x_i - ww^T x_i)^2 = min \left\| X - ww^T X \right\|_2^2 \tag{1}$$

where $\{\mathbf{w}\}_{j=1}^{d}$ is a subset of orthogonal projection vectors in $\Re^m$ and the set of data points $\{\mathbf{x}_i\}_{i=1}^{n}$ is zero-mean m-dimensional data points. It can be seen that, PCA uses a least square framework to minimize the sum distance between the original dataset X and the reconstructed dataset $ww^T X$. This geometrically will force the projection vector w to pass through the densest data points to minimize the sum distance. This can be seen in Fig. 1, where u_1 is the first principal projection vector. From this intuition, we evaluate the importance of instances by considering the relationship between each instance and the principal projection u_1. That is, the closer an instance to the projection vector u_1, the more important the instance in pursuing projections.

Therefore, we extend formula (1) to include a scaling-factor. This factor imposes a penalty on the instance space to suppress the impact of noise in incomplete datasets. The following is our propose function:

$$\min_{D,p} \left\| X - XDpp^T D \right\|_F^2 = \max_{D,p} p^T DX^T XDp \tag{2}$$
$$s.t. \, p^T D^2 p = 1$$

where p is a vector of sample space and $D = diag(d_1, d_2, \cdots, d_n)$ is a diagonal matrix that evaluates the importance of each instance in X. With this penalty, we are actually pursuing a projection $Z = Dp$ with $Z^T Z = I$ that considers the effect of instances. For example, if a lower scaling-factor d_i is assigned to the projection Z, the component of sample space Z_i is suppressed, which means the corresponding sample x_i contributes little to the projection Z.

To enforce the constraint in formula (2), we introduce the Lagrange multiplier (λ) and take partial derivatives w.r.t. p to obtain:

$$X^T XDp = \lambda Dp \tag{3}$$

It can be observed that Eq. (3) is a standard eigenvalue problem.

Mathematically, there is a direct relationship [16] between PCA and SVD when PCA components are calculated from the covariance matrix. The equation for singular value decomposition (SVD) of X is as follows:

$$X = u\Sigma v^T \quad s.t. \, u^T u = I_r, v^T v = I_r \tag{4}$$

In our proposed model, $v = DP$, where P is the set of r projections of p. Thus, the projection u in feature space can be obtained as follows:

$$u = XDP\Sigma^{-1} \tag{5}$$

where the low dimension feature space u is obtained with an injection of sample factors different from the traditional PCA. In this way, IFPCA can learn a low dimensional subspace from both sample and feature spaces of a dataset for improved performance.
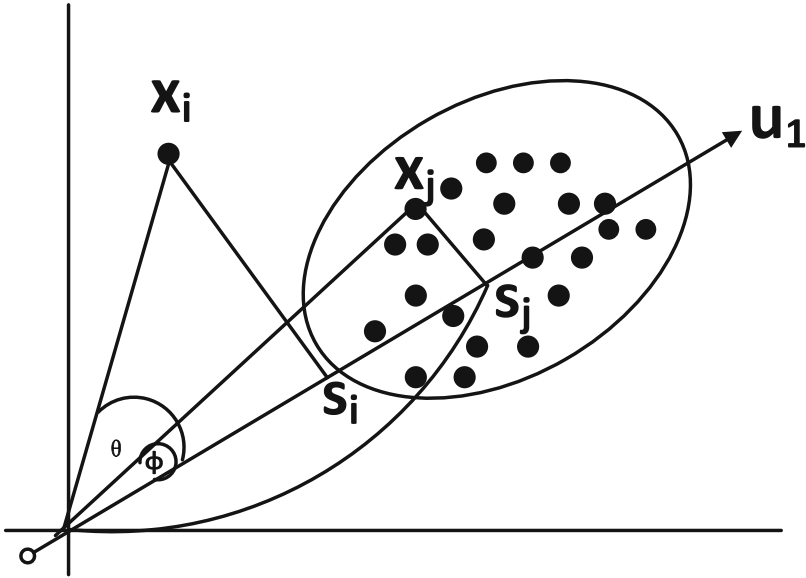

Fig. 1. Illustration of an instance relationship with the principal projection

2.1   Strategies of Building Matrix D

In this subsection, we model the relationship between scaling-factor D and the principal projection u_1 using two strategies: total distance and cosine similarity metrics. Both can be obtained geometrically as shown in Fig. 1.

The first strategy uses total distance metric to iteratively learn the relationship between each instance and the principal projection u_1. The total distance of an instance is defined as the square sum of the distances between the coordinate of each instance and the coordinates of every other instance in the training set to the projection u_1. The total distance of an instance is a natural way to evaluate its importance within the set. From Fig. 1, we can observe that the total distance of instance x_i which is outside the cluster will be relatively bigger than that of instance x_j within the cluster. Therefore, instance x_i is more likely to be an outlier or corrupt instance than x_j. From Fig. 1, the coordinate of instance x_i to the projection u_1 is computed through:

$$s_i = u_1^T x_i \tag{6}$$

Algorithm 1. The proposed IFPCA method

1: **Input:** Training set X
2: **Output:** The projection vector p
3: **Parameters:** ε
4: **Initialize:** Initialize D as an identity matrix
5: **while** not converged **do**
6:     obtain p based on eqn. (3).
7:     obtain u based on eqn. (5)
8:     Update D based on eqn. (7) or (9)
9:     Compute loss from formula (2).
10: **end while**

We then compute d_i through total distance metric as follows:

$$d_i = \sum_{i,j=1}^{n} (s_i - s_j)^2 \tag{7}$$

Thus, the bigger the d_i, the more likely x_i is a noisy or corrupt instance and hence its relevance will be scaled accordingly to suppress its effect on the projection.

The second strategy uses cosine similarity metric to build the scaling-factor D. This also iteratively learns the angle relationship between each instance in the training set and the principal projection u_1. Thus, by normalizing formula (6), the angle between each instance and the principal projection u_1 is defined as follows:

$$\cos \theta_i = \frac{u_1^T x_i}{\|u_1\| \cdot \|x_i\|} \tag{8}$$

From formula (8), a bigger $\cos \theta_i$ implies a smaller angle θ_i between instance x_i and the principal projection u_1 and vice versa. We illustrate the relationship between an instance and the principal projection in Fig. 1. From Fig. 1, it can be seen that, angle ϕ of instance x_j is relatively smaller than angle θ of instance x_i. Thus, x_j will be considered probably more important in finding best projections than x_i which might be noisy. Recall that d_i is a negative factor, we compute d_i through the similarity metric as follows:

$$d_i = \frac{1}{abs(\cos \theta_i) + \varepsilon} \tag{9}$$

where, $\varepsilon = 0.0001$ is a parameter to avoid d_i approaching infinity.

By iteratively scaling the data using these two strategies, the effect of corrupt instances in the training set will considerably minimize leading to better low-rank projections. The algorithm of the proposed IFPCA method is as shown in algorithm 1.

3   Experiments and Complexity Analysis

To demonstrate the effectiveness of the proposed IFPCA algorithm, we conduct experiments on COIL-20, ORL and USPS datasets using the proposed IFPCA and six state-of-the-art DR methods such as LSDA [1], gLPCA [11], RPCA-OM [12], PCA [17], LPP [3] and RCDA [18].

3.1   Parameter Settings

For each dataset, we randomly sampled 60%, 70% of the instances for training and 40 and 30% respectively for testing in our experiments. The parameters of LSDA [1], gLPCA [11], RPCA-OM [12], PCA [17], LPP [3] and RCDA [18] were set according to their literature. We set the k-nearest-neighbors parameter K to 5 in IFPCA, and all other relevant comparative methods, in order to make a very fair comparison. We finally make use of the K-nearest neighbor (KNN) classifier for classifications. We record results for our framework as IFPCA-1 and IFPCA-2, where IFPCA-1 and IFPCA-2 represent cosine similarity and total distance metrics respectively. The experiments are repeated 15 times and we record the average classification accuracies, corresponding optimal dimensions and standard deviations for the various methods.

3.2   Results Discussions and Analysis

We discuss and analyze the results obtained for each method on the three datasets used for our experiments in this section.

Object Recognition. We validate the proposed methods on object recognition using COIL-20 dataset. This dataset [19] contains 1440 observations of 20 objects with 72 poses each. The objects were placed on a motorized turntable against a black background and rotated through $360°$ to vary the object pose with respect to a fix camera. Images of the 20 objects were taken at pose intervals of $5°$. The results for the various methods are shown in Table 1 with best results in bold. From Table 1, we can see that IFPCA-1 and IFPCA-2 both have superior performances than all the comparative methods. For optimal dimensions, IFPCA-1 and IFPCA-2 obtain phenomenal optimal dimensions in both samples as compare to the other six comparative methods. For the 60% sample, IFPCA-1 outperforms IFPCA-2 by a small margin of 0.16%, gLPCA by an impressive margin of 2.69%, RPCA-OM by 1.61%, PCA by 3.44%, LPP by 11.86%, LSDA by 11.25% and RCDA by 2.19%. Again, IFPCA-1 and IFPCA-2 obtain the lowest variances in both cases which prove their consistency in performance.

Thus, the proposed methods have superior performances than all other comparative methods in object recognition and optimal dimensions. This is because the two proposed methods can detect and sufficiently suppress the impact of noisy data points in the training set than the other comparative methods. Figure 2 shows the trend of classification accuracies of each method against the variation of dimensions. It is evident from Fig. 2 that IFPCA-1, IFPCA-2, RPCA-OM, PCA and RCDA attain stable performances in higher dimensions above 10. While the performances of gLPCA, LPP and LSDA decline considerably as the dimensions increase.

Face Recognition. We further validate the effectiveness of the proposed method on face recognition using the ORL face dataset. This dataset [20] has 40 subjects, each with 10 faces at different poses making a total of 400 images of size 112×92. However, the images were resized to 32×32 for our experiments. These images were taken at different times, lighting and facial expressions. The faces are in an upright position in

Table 1. Mean classification accuracies ± standard deviation (%) and (optimal dimensions) of the various methods on the COIL-20 dataset.

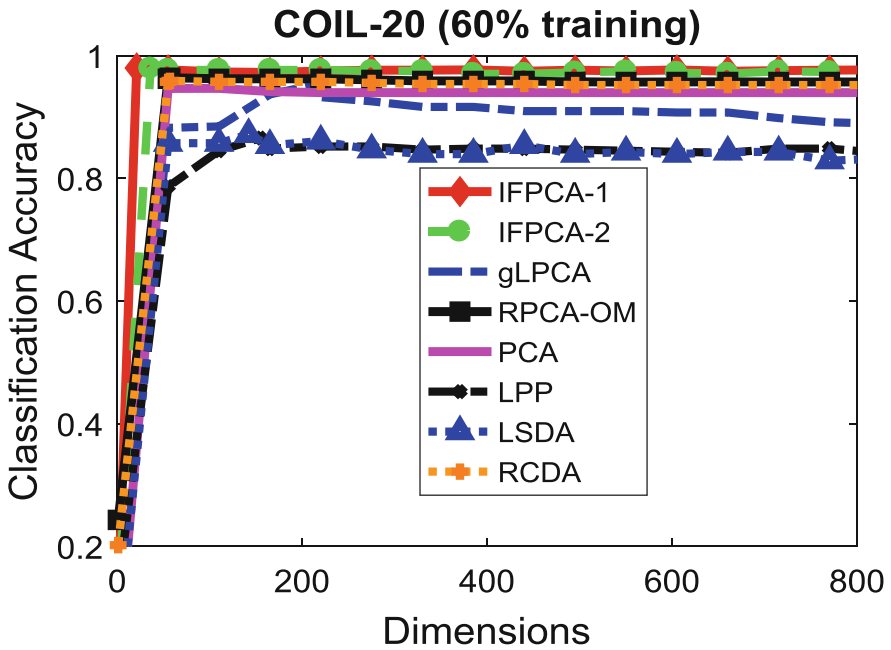| Sample	60%	70%
IFPCA-1	**97.84 ± 0.05(21)**	**98.89 ± 0.001(23)**
IFPCA-2	97.68 ± 0.009(36)	98.77 ± 0.003(37)
RCDA	95.65 ± 0.40(57)	97.21 ± 0.31(58)
PCA	94.40 ± 0.39(92)	95.50 ± 0.28(95)
RPCA-OM	96.23 ± 0.11(57)	98.46 ± 0.20(60)
gLPCA	95.15 ± 0.18(214)	96.31 ± 0.23(217)
LPP	85.98 ± 1.42(155)	87.49 ± 1.25(163)
LSDA	86.59 ± 1.48(142)	88.41 ± 1.39(144)

Fig. 2. Classification accuracies against the variation of dimensions on the COIL-20 dataset by the various methods

frontal view with a slight left-right rotation. The results for the various methods are shown in Table 2 with best results in bold.

The results as shown in Table 2, clearly indicate that IFPCA-1 and IFPCA-2 both have exceptional performances compared to all the comparative methods in face recognition in both the 60 and 70% training samples. For the 70% training sample, IFPCA-1 achieves a remarkable face recognition accuracy of 98.67% and the best optimal dimension of 25 with gLPCA and LPP achieving the worst dimensions of 279 each. Thus, for

Table 2. Mean classification accuracies ± standard deviation (%) and (optimal dimensions) of the various methods on the ORL dataset.

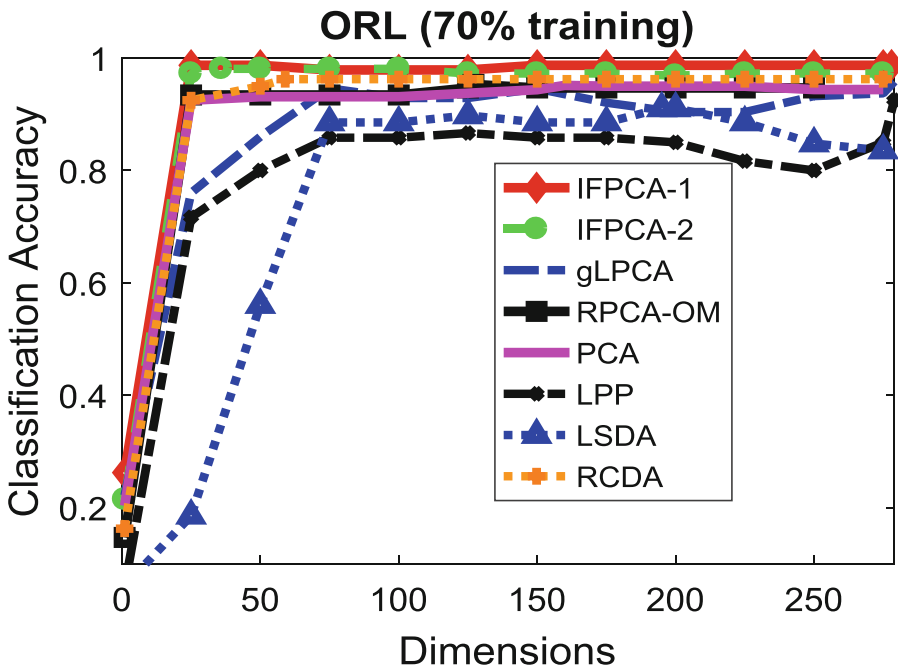| Sample	60%	70%
IFPCA-1	**97.41 ± 0.09 (24)**	**98.67 ± 0.21 (25)**
IFPCA-2	96.57 ± 0.14 (34)	98.29 ± 0.11 (36)
RCDA	95.51 ± 0.28 (58)	96.47 ± 0.23 (59)
PCA	92.90 ± 0.68 (158)	94.32 ± 0.45 (160)
RPCA-OM	94.32 ± 0.85 (126)	95.83 ± 0.37 (127)
gLPCA	94.02 ± 0.20 (279)	95.61 ± 0.33 (279)
LPP	91.01 ± 1.07 (279)	92.12 ± 0.83 (279)
LSDA	88.24 ± 0.90 (155)	90.54 ± 0.85 (195)

Fig. 3. Classification accuracies against the variation of dimensions on the ORL dataset by the various methods

face recognition accuracy, IFPCA-1 outperforms IFPCA-2 by a little margin of 0.38%, gLPCA by 3.06%, RPCA-OM by 2.84%, PCA by 4.35%, LPP by 6.55%, LSDA by 8.13% and RCDA by 2.20%. Figure 3 shows the trend of classification accuracies of each method against the variation of dimensions. It is clear from Fig. 3 that the performances of IFPCA-1, IFPCA-2, RPCA-OM, PCA and RCDA have once again been stable irrespective of increases in dimensions. While that of gLPCA, LPP and LSDA

are considerably unstable. The results further show that IFPCA-1 and IFPCA-2 have the most consistent performances in the 60 and 70% samples respectively.

Handwritten Digit Recognition. In our quest to further demonstrate the effectiveness of our framework, we run experiments on the USPS dataset. This dataset [21] consists of handwritten digits from 0 to 9. The training and testing sets consist of 7291 examples and 2007 examples respectively. Each example has 256 attributes or pixels that describe each digit. The results for the various methods are shown in Table 3 with best results in bold.

From Table 3, IFPCA-1 and IFPCA-2 once again out perform all the comparative methods in handwritten digit recognition for both training samples of the USPS dataset. For the 60% sample, IFPCA-1 has a digit recognition accuracy of 0.55% more than IFPCA-2, 1.87% more than gLPCA, 1.69% more than RPCA-OM, 3.06% more than PCA, 3.09% more than LPP, 3.73% more than LSDA and 1.75% more than RCDA. IFPCA-1 further obtains the best optimal dimensions of 27 and 29 in the 60 and 70% training samples respectively.
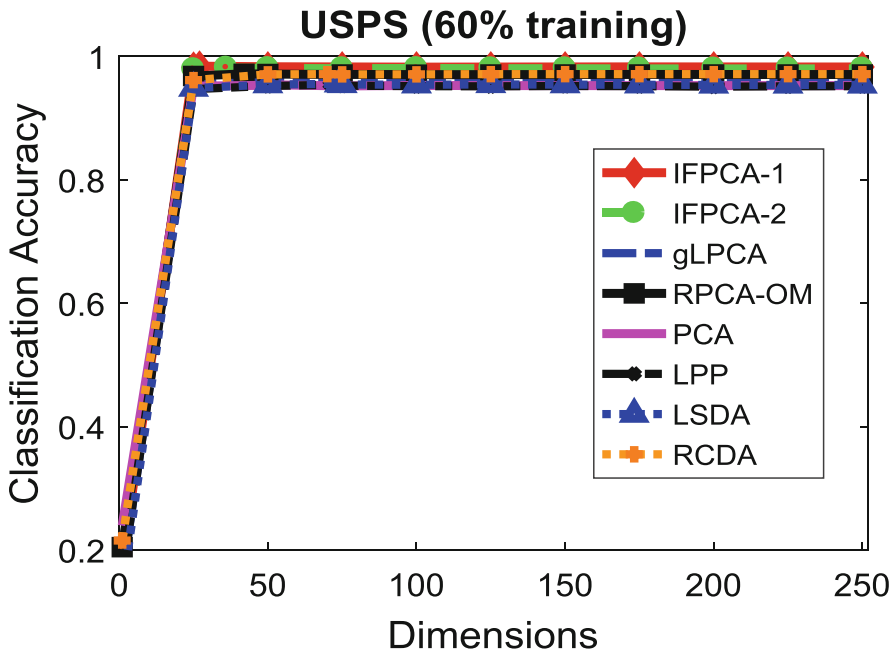

Fig. 4. Classification accuracies against the variation of dimensions on the USPS dataset by the various methods

Table 3. Mean classification accuracies $\pm$ standard deviation (%) and (optimal dimensions) of the various methods on USPS dataset.

Sample	60%	70%
IFPCA-1	**98.40 $\pm$ 0.03(27)**	**98.87 $\pm$ 0.002(29)**
IFPCA-2	97.85 $\pm$ 0.011(36)	98.56 $\pm$ 0.014(37)
RCDA	96.65 $\pm$ 0.33(70)	97.51 $\pm$ 0.25(71)
PCA	95.34 $\pm$ 0.18(55)	96.40 $\pm$ 0.15(60)
RPCA-OM	96.71 $\pm$ 0.17(42)	97.52 $\pm$ 0.22(44)
gLPCA	96.53 $\pm$ 0.27(53)	97.03 $\pm$ 0.36(55)
LPP	95.31 $\pm$ 0.18(71)	96.29 $\pm$ 0.31(73)
LSDA	95.67 $\pm$ 0.21(74)	96.50 $\pm$ 0.19(80)

Table 4. Computation time in seconds for training and testing for each method

Dataset	COIL20		ORL		USPS	
	Training	Testing	Training	Testing	Training	Testing
IFPCA-1	3.26	4.62×10^{-3}	1.93×10^{-1}	1.18×10^{-3}	1.69	2.20×10^{-3}
IFPCA-2	2.12	2.71×10^{-3}	1.68×10^{-1}	1.13×10^{-3}	3.82	5.40×10^{-3}
RCDA	3.51	$2.24 \times 10{-2}$	2.05	1.95×10^{-2}	4.02	5.13×10^{-2}
PCA	5.29×10^{-1}	1.95×10^{-2}	8.44×10^{-2}	1.95×10^{-2}	4.42×10^{-1}	2.80×10^{-2}
RPCA-OM	1.12	1.48×10^{-2}	1.33×10^{-1}	4.24×10^{-3}	1.78	1.01×10^{-1}
gLPCA	6.10×10^{-1}	5.93×10^{-3}	5.81×10^{-2}	2.46×10^{-3}	5.33×10^{-1}	6.50×10^{-3}
LPP	6.08×10^{-1}	$1.43 x 10^{-2}$	8.00×10^{-2}	5.32×10^{-3}	5.44×10^{-1}	7.48×10^{-3}
LSDA	5.62×10^{-1}	3.18×10^{-2}	1.16×10^{-1}	2.05×10^{-2}	4.24×10^{-1}	1.28×10^{-2}

IFPCA-1 and IFPCA-2 obtain the lowest variances than the comparative methods. The consistency in the performances of IFPCA-1 and IFPCA-2 proved their ability to discover better intrinsic structure of the dataset. Figure 4 shows the trend of classification accuracies of each method against the variation of dimensions. From Fig. 4 all the methods show stable performances in dimensions above 20, but with the proposed methods in the lead.

Complexity Analysis. We compare the computational times of the proposed methods to the other six comparative methods. All algorithms were implemented in MATLAB R2016b version 9.1.0.441655 64-bit using a personal computer with Intel (R) Core(TM) i5-7500 CPU @ 3.40 GHz with 8.00 GB memory and Windows 7 operating system environment. The convergence of the proposed framework depends on the importance evaluation diagonal matrix D. The computation time of an eigenvalue problem on a training set X of size $m \times n$ is $O(m^3)$. This means that a complexity of $O(m^3)$ is required by the proposed framework to compute the projection vector p since our framework is an eigenvalue problem. For the inner loop, if it takes k number of iterations in pursuing D for convergence to be attained, the complexity is $O(kmn)$. Hence,

the total complexity of the framework becomes $O(t(m^3 + kmn))$, where t is the number of iterations of the outer loop. Table 4 shows the computation time for each method on all three datasets.

4   Conclusions and Recommendation

We proposed in this paper a novel incomplete-data oriented dimension reduction via instance factoring PCA framework. Different from other variants of PCA, a scaling-factor that imposes a penalty on the instance space is introduced to suppress the impact of noise in pursuing projections. Two strategies, cosine similarity and total distance metrics are used geometrically to iteratively learn the relationship between each instance and the principal projection.

Comprehensive experiments on COIL-20, ORL and USPS datasets demonstrate the effectiveness of the proposed framework in both dimension reduction and classification tasks. This is because it is able to obtain low-rank projections in incomplete datasets by suppressing the effect of noisy or corrupt instances. This shows that our framework is more noise tolerant than the other comparative methods. We will extend this framework to low rank representation in the near future.

References

1. Deng, C., He, X., Zhou, K., Han, J., Bao, H.: Locality sensitive discriminant analysis (2007)
2. Teoh, A.B.J., Han, P.Y.: Face recognition based on neighbourhood discriminant preserving embedding. In: International Conference on Control, Automation, Robotics and Vision, pp. 428–433 (2009)
3. Feng, G., Hu, D., Zhou, Z.: A direct locality preserving projections (DLPP) algorithm for image recognition. Neural Process. Lett. **27**(3), 247–255 (2008)
4. Chen, J., Ye, J., Li, Q.: Integrating global and local structures: a least squares framework for dimensionality reduction. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
5. Yang, M.H.: Face recognition using extended isomap. In: International Conference on Image Processing (2002)
6. Wiriyathammabhum, P., Kijsirikul, B.: Robust principal component analysis using statistical estimators. arXiv preprint arXiv:1207.0403 (2012)
7. Jolliffe, I.T., Cadima, J.: Principal component analysis: a review and recent developments. Philos. Trans. Roy. Soc. A Math. Phys. Eng. Sci. **374**(2065), 20150202 (2016)
8. Roweis, S.T.: EM algorithms for PCA and SPCA. In: Advances in Neural Information Processing Systems, pp. 626–632 (1998)
9. Kriegel, H.-P., Kröger, P., Schubert, E., Zimek, A.: A general framework for increasing the robustness of PCA-based correlation clustering algorithms. In: Ludäscher, B., Mamoulis, N. (eds.) SSDBM 2008. LNCS, vol. 5069, pp. 418–435. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69497-7_27
10. Vaswani, N., Bouwmans, T., Javed, S., Narayanamurthy, P.: Robust PCA, subspace learning, and tracking. IEEE Signal Process. Mag. **35**, 32–55 (2018)

11. Jiang, B., Ding, C., Luo, B., Tang, J.: Graph-Laplacian PCA: closed-form solution and robustness. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3492–3498 (2013)
12. Nie, F., Yuan, J., Huang, H.: Optimal mean robust principal component analysis. In: International Conference on Machine Learning, pp. 1062–1070 (2014)
13. Abeo, T.A., Shen, X.J., Gou, J.P., Mao, Q.R., Bao, B.K., Li, S.: Dictionary-induced least squares framework for multi-view dimensionality reduction with multi-manifold embeddings. IET Comput. Vision **13**(2), 97–108 (2019)
14. Yang, W., Shi, Y., Gao, Y., Wang, L., Yang, M.: Incomplete-data oriented multiview dimension reduction via sparse low-rank representation. IEEE Trans. Neural Netw. Learn. Syst. **29**, 6276–6291 (2018)
15. Li, C., Liu, J., Liu, Y., Xu, C., Liu, Q., Lu, H.: Ordinal preserving projection: a novel dimensionality reduction method for image ranking. In: Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, p. 17. ACM (2012)
16. Wall, M.E., Rechtsteiner, A., Rocha, L.M.: Singular value decomposition and principal component analysis. In: Berrar, D.P., Dubitzky, W., Granzow, M. (eds.) A Practical Approach to Microarray Data Analysis, vol. 5, pp. 91–109. Springer, Boston (2003). https://doi.org/10.1007/0-306-47815-3_5
17. Vidal, R., Ma, Y., Sastry, S.S.: Generalized Principal Component Analysis. IAM, vol. 40. Springer, New York (2016). https://doi.org/10.1007/978-0-387-87811-9
18. Huang, K.K., Dai, D.Q., Ren, C.X.: Regularized coplanar discriminant analysis for dimensionality reduction. Pattern Recogn. **62**(Complete), 87–98 (2017)
19. Alom, M.Z., Josue, T., Rahman, M.N., Mitchell, W., Yakopcic, C., Taha, T.M.: Deep versus wide convolutional neural networks for object recognition on neuromorphic system. arXiv preprint arXiv:1802.02608 (2018)
20. Jiang, R., Al-Maadeed, S., Bouridane, A., Crookes, D., Celebi, M.E.: Face recognition in the scrambled domain via salience-aware ensembles of many kernels. IEEE Trans. Inf. Forensics Secur. **11**(8), 1807–1817 (2016)
21. Proedrou, K., Nouretdinov, I., Vovk, V., Gammerman, A.: Transductive confidence machines for pattern recognition. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) ECML 2002. LNCS (LNAI), vol. 2430, pp. 381–390. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-36755-1_32