Comparative analysis of 2D-PCA based dimensionality reduction and feature extraction

Kai Zhang¹, Xiaoqing Hu*, Lei Kang¹, Qiaoyu Ma¹, Xin Zhang²

¹School of Science, China University of Geosciences (Beijing), Beijing 100083, China
²Beijing Normal university, School of Statistics, Beijing 100091, China

*Corresponding author’s e-mail: xiaoqinghu_321@163.com

Abstract. The complex spectral and spatial characteristics of hyperspectral remote sensing images (HSI) lead to higher time-consuming in classification task. To address this question, we introduced the 2D-PCA dimensionality reduction method of linear mapping in the two-dimensional spatial domain on the basis of linear dimensionality reduction in the spectral domain, thereby compressing the complex spatial structure information of HSI into a limited low-dimensional space, and realizing space-spectrum dimensionality reduction and information fusion. The experimental results on three classic data sets of Salinas, Tea Farm, and Indian Pines show that 2D-PCA has a strong ability to condense and compress spatial structure characteristics. Compared with popular deep learning frameworks such as CNN and Mixer-MLP, conventional machine learning models based on 2D-PCA have significant advantages in terms of computing time under the premise of controllable accuracy loss, which makes 2D-PCA a promising method for dimensionality reduction and feature expression in hyperspectral pixel-wise classification.

1. Introduction
Hyperspectral remote sensing technology is commonly used to detect ground targets, providing an effective database in classification by collecting the reflection spectra of light waves from ground materials in continuous wavelengths. Ground features contain two typical features in HSI. First, the different reflectance and absorption characteristics of feature in high-dimensional spectrum; And second, unique distribution and combination of features in the spatial domain could clearly show in remote sensing images. Spatial-spectral domain fusion is a direct, simple and effective method for feature extraction in classification of HSI. But the common HSI in a real application would hundreds and thousands of bands lead to low classification accuracy and higher time-consuming. Therefore, dimensionality reduction is an important research in the application of hyperspectral remote sensing.

Dimensionality reduction algorithms usually mapping the original high-dimensional spectral information into the low-dimensional subspace by the means of linear or nonlinear. At present, widely used algorithms are Principal Component Analysis (PCA)[1] and Linear Discriminant Analysis (LDA)[2], etc., both of which achieve dimensionality reduction by pulling the sample matrix into a one-dimensional vector. These methods only reduce the dimensionality of the spectral, and further spatial-spectral feature fusion is still required. For example, feature extraction algorithms based on
theoretical models (Gabor filter, wavelet transform, etc.), or extraction of spatial-spectral features and fusion by the means of convolution process or encoding of deep learning frameworks (Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Transformer, etc.). These algorithms are extremely time-consuming.

Each pixel takes a fixed window as an image sample in classification and analysis of HSI. There are obvious shortcomings in traditional dimensionality reduction algorithms: (1) The original spatial information of the image is destroyed by the flattening matrix, which could lead to decrease in discrimination of feature extraction. Meanwhile, the destruction of spatial information is irreversible, and not compensated with enhanced algorithms. (2) The information of each sample in the HSI are composed of continuous spectrums. Image sample with fixed window in the immediate area has high computational complexity.

The 2D-PCA(3) is widely used in signal detection(4), image dimensionality reduction(5), image classification and image reconstruction(6), and face image recognition(7). In this paper, 2D-PCA is used in HSI, which not only reduces the dimensionality and extracts features in two-dimensional spatial domain but also finds effective linear mapping. Specifically, the 2D-PCA is used to dimensionality reduction and feature extraction in Salinas, Tea Farm and Indian Pines, and then machine learning algorithm are used as classifiers. Finally, compared with the deep learning frameworks to verify the effectiveness of 2D-PCA in spatial feature extraction and dimensionality reduction.

2. Principle of the method

2.1. 2D-PCA

In the 2D-PCA algorithm, \( X \) denotes an \( n \)-dimensional unit column vector, \( A_{b} \) denotes the spatial image of a single band, and each image sample size is \( m \times n \), \( b \in \{1,2,\cdots,B\} \). The image sample in \( A_{b} \) is mapping into the vector \( X \) and undergoes linear transformation, and then the mapping image sample \( Y \) is obtained. In order to fit mapping vector \( X \), the total divergence of the mapping samples used to measure the mapping effect. The total divergence can expressed the trace of the covariance matrix.

\[
J(X) = tr(S_x)
\]

(1)

Where \( S_x \) is the covariance matrix of the mapping eigenvectors of the training samples, and \( tr(S_x) \) is the trace of the covariance matrix. \( S_x \) is defined as.

\[
S_x = E( (Y - E(Y))^T (Y - E(Y)) )
= E( [ (A - E(A)) X ] [ (A - E(A)) X ]^T )
\]

(2)

At the same time, the total divergence matrix is:

\[
G_x = E( (A - E(A))^T (A - E(A)) X )
\]

(3)

Then the formula of the trace of the covariance is:

\[
J(X) = X^T G_x X
\]

(4)

After the image matrix is mapping into \( X \), the overall scatter of the mapping sample reaches a maximum. But it is not enough for an image to have only one mapping axis. It usually needs to select a series of mapping vector \( X_1, X_2, \cdots, X_d \), which satisfy the standard orthogonal constraint and maximizing total divergence. The mapping vector is:
\[
\{X_1, X_2, \ldots, X_d\} = \arg \max J(X) \\
X_i^T X_j = 0
\]  

(5)

Where \(i \neq j\), and \(i, j \in (1, 2, 3, \ldots, d)\). In fact, the optimal vector \(X_1, X_2, \ldots, X_d\) are the eigenvector corresponding to the first \(d\) largest eigenvalues of \(G\). The optimal vector \(X_1, X_2, \ldots, X_d\) of 2D-PCA are used for feature extraction. For image sample \(A\) is:

\[
Y_k = A_{b_k} X_k
\]  

(6)

Where \(k \in (1, 2, 3, \ldots, d)\). A set of mapping vectors \(Y_1, Y_2, \ldots, Y_d\) was obtained, which represents the principal component of the image sample \(A_{b_k}\). Each principal component of 2D-PCA is a vector, and a feature matrix \(B = [Y_1, Y_2, \ldots, Y_d]\) consists of all principal components. Finally, the mapped image matrix is obtained based on the mapping of the feature map matrix.

### 2.2. Deep learning frameworks

In this paper, CNN and MLP-Mixer[8] are selected as the comparison models. The convolutional layer is applied to extract the deep and shallow features of the image, and the fully connected layer is used to mapping and classification in the CNN. The structure of CNN is show in Figure 1. In addition, the MLP mixer relying on repeat mapping to achieve information interaction in spectral and spatial domain. The specific structure is shown in Figure 2.

![Figure 1. CNN network structure](image1)

![Figure 2. MLP-Mixer Network Structure](image2)

MLP-Mixer is composed of multiple patch linear embedding layers, Mixer layers and a classifier head. The Mixer layer mainly consists of two different forms: Channel-mixing MLP allows communication between different bands and is applied to individual bands of image patches. Token-mixing MLP is used in images with different spatial positions on the columns of every patches.
Moreover, feature inputs are integrated by skip-connection and Layer Norm, and feature inputs are converted to probabilities by nonlinear activation unit (GELU).

Specifically, MLP-Mixer divides each image sample with a window of 15×15 into patches. Using non-overlapping patches as input, and then features in every patch are extracted in Token-mixing MLPs, and features on every channel are extracted in Channel-mixing MLPs.

2.3. Classification based on 2D-PCA
In this paper, feature extraction of image samples performed by 2D-PCA. The image samples are mapped into a low-dimensional space with a size range of 1×1, 2×2, 3×3, …, 15×15. Then, the models used as comparisons are Back Propagation (BP) and Random Forest (RF). BP has one hidden layer with 128 neurons. RF is composed of 100 sub-decision trees with classification function. Each sub-tree is split by Gini metrics and split until the leaf nodes contain only non-distinguishable samples.

3. Experiments and analysis
3.1. Dataset
In this paper, we conducted experiments on three widely used HSI: Indian Pines, Salinas, and Tea Farm. The detailed parameters are shown in Table 1.

| Dataset  | Indian Pines | Salinas | Tea Farm |
|----------|--------------|---------|----------|
| Data size| 145×145      | 512×217 | 348×512  |
| Bands    | 200          | 224     | 80       |
| Spatial resolution (m) | 20          | 3.7     | 2.25     |
| Spectral range (μm)   | 0.4~2.5      | 0.4~2.5 | 0.417~0.855 |
| Number of categories  | 16           | 16      | 10       |
| Total number of samples | 10249      | 54129   | 53734    |

Figure 3. False-color image and ground truth of three HSI. a-c: Salinas, d-f: Tea Farms, g-i: Indian Pines.

Indian Pines dataset was collected in Northwest Indiana, USA, with a spectral range of 0.4~2.5 μm and spatial resolution of 20 m. The size of the single band image is 145×145 and contains 10249 valid samples except for background point. There are 16 classes of crops, including corn, oats, woods, etc. Salinas dataset was acquired using the AVIRS imaging Spectrometer in Salinas Valley, California, USA. The Spectral range is 0.4~2.5 μm and spatial resolution is 3.7 m. After eliminating the bands
with severe water absorption, 204 effective bands are reserved. Tea Farms dataset was gathered at Tea Garden Provincial Base, Fanglu Village, Changzhou City, Jiangsu Province, China. The image data consists of 80 bands and 10 typical labels, the image size is 348×512, and the spatial resolution is 2.25 m. Specifically, the false-color maps and labels of the three HSI are shown in Figure 3.

3.2. Experiment
To meet the optimal size of the 2D-PCA mapping, The image sample is mapped to a different range of feature space and then verified by BP and RF. The results are shown in Figure 4.

![Figure 4](image_url)

Figure 4. Accuracy of different 2D-PCA mapping size (a) Indian Pines, (b) Salinas, (c)Tea Farms.

As we can observe from Figure 4, the classification results after mapping with 2D-PCA showed an increase followed by a decrease for different data. The inflection point occurs when the mapping size of 2D-PCA is 2×2. Consequently, the mapping size 2×2 of each image sample is used as a subsequent comparison. The results are shown in the Table 2, Table 3.

| Dataset      | CNN   | MLP-Mixer | 2D-PCA |
|--------------|-------|-----------|--------|
| Indian Pines | 83.19 | 86.55     | 89.41  |
| Salinas      | 99.43 | 98.92     | 99.59  |
| Tea Farms    | 99.69 | 98.72     | 98.75  |

Table 3. Time-consuming (s) of different model for HSI.

| Dataset     | CNN   | MLP-Mixer | 2D-PCA |
|-------------|-------|-----------|--------|
| Indian Pines| 509   | 305       | 0.64   |
| Salinas     | 2583  | 1551      | 0.33   |
| Tea Farms   | 199   | 964       | 1.49   |

In remote sensing applications, the classification results showed slight differences between the method and the deep learning frameworks, but deep learning frameworks will take longer. From Table 2, we can see that classification accuracy of BP and RF reach 86.68% and 87.83% based on 2D-PCA, which is slightly better than results of deep learning frameworks. Therefore, the method is less time-consuming than deep learning frameworks. Feature extraction and dimensionality reduction by 2D-PCA is more applicable and efficient in classification of HSI. The results are shown in Figure 5, Figure 6, and Figure 7.

4. Conclusion
The 2D-PCA is used for dimensionality reduction and feature extraction. Comparing the feature extraction algorithms of convolution and coding in CNN and MLP-Mixer, Image samples are mapped by 2D-PCA to obtain simpler and more representative features. In this paper, the HSI are dominated by agriculture, forestry, and plants, with a small portion of buildings. The 2D-PCA shows good applicability, and still has good effect on features of natural, urban and traffic with complex spatial
and abstract features. Therefore, researching more targeted HSI classification algorithms has become an important research direction.

Figure 5. Classification of Indian Pines in (a) BP, (b) RF, (c) CNN, (d) MLP-Mixer, (e) Original.

Figure 6. Classification of Salinas in (a) BP, (b) RF, (c) CNN, (d) MLP-Mixer, (e) Original.

Figure 7. Classification of Tea Farm in (a) BP, (b) RF, (c) CNN, (d) MLP-Mixer, (e) Original.

References

[1] Chen, G. Y. (2021). Multiscale filter-based hyperspectral image classification with PCA and SVM. Journal of Electrical Engineering, 72(1), 40–45.

[2] Lu, J., Platanitis, K. N., & Venetsanopoulos, A. N. (2003). Face recognition using LDA-based algorithms. IEEE Transactions on Neural Networks, 14(1), 195–200.

[3] Yang, J., Zhang, D., Frangi, A. F., & Yang, J. Y. (2004). 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.

[4] Xie, H.B., & Liu, H. (2017). Myoelectrical signal classification based on S transform and two-directional 2DPCA. In ESANN 2017-Proceedings, 25th European Symposium on Artificial Neural Networks, Bruges. pp. 477–482.

[5] Li, J., & Yan, D. (2017). Face recognition using composite classifier with 2DPCA. In Seventh International Conference on Electronics and Information Engineering. Shandong University (Weihai), Weihai, Shandong, China. Vol. 10322, p. 103221P

[6] Wang, D.G., Li, Y., & Jin, F.L. (2013). SAR images recognition combined bidirectional 2DPCA with KPCA. In Advanced Materials Research. Xiamen, Fujian, China. pp. 4045-4049.

[7] Zhou, S., & Zhang, D. (2018). Bilateral angle 2DPCA for face recognition. IEEE Signal Processing Letters, 26(2), 317-321.

[8] Tolstikhin, I., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., & Dosovitskiy, A. (2021). Mlp-mixer: An all-mlp architecture for vision. arXiv preprint arXiv:2105.01601.