

Journal of
Applied Remote Sensing

RemoteSensing.SPIEDigitalLibrary.org

Land cover classification of PolSAR image using tensor representation and learning

Mingliang Tao
Jia Su
Ling Wang

SPIE.

Mingliang Tao, Jia Su, Ling Wang, "Land cover classification of PolSAR image using tensor representation and learning," *J. Appl. Remote Sens.* **13**(1), 016516 (2019), doi: 10.1117/1.JRS.13.016516.

Land cover classification of PolSAR image using tensor representation and learning

Mingliang Tao,^{a,b,*} Jia Su,^{a,b,*} and Ling Wang^{a,b}

^aNorthwestern Polytechnical University, Research and Development Institute, Shenzhen, China

^bNorthwestern Polytechnical University, School of Electronics and Information, Xi'an, China

Abstract. We propose a tensor representation for polarimetric synthetic aperture radar data and extend the usage of tensor learning technique for feature dimension reduction (DR) in image classification. Under the tensor algebra framework, each pixel is modeled as a third-order tensor object by combining multiple polarimetric features and incorporating neighborhood spatial information together. A set of training tensors are determined according to the prior knowledge of the ground truth. Then a tensor learning technique, i.e., multilinear principal component analysis, is applied on the training tensors set to find a tensor subspace that captures most of the variation in the original tensor objects. This process serves as a feature DR step, which is critical for improving the subsequent classification accuracy. Further, the projected tensor samples after DR are fed to the k -nearest neighbor classifier for supervised classification. The performance is verified in both simulated and real datasets. The extracted features are more discriminative in the feature space, and the classification accuracy is significantly improved by at least 10% compared with other existing matrix-based methods. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.13.016516](https://doi.org/10.1117/1.JRS.13.016516)]

Keywords: polarimetric synthetic aperture radar; land cover classification; dimension reduction; tensor representation and learning; multilinear principal component analysis.

Paper 180605 received Jul. 17, 2018; accepted for publication Dec. 28, 2018; published online Feb. 9, 2019.

1 Introduction

Polarimetric synthetic aperture radar (PolSAR) has been an important instrument for active remote sensing since it can provide scattering information under different combinations of wave polarizations.¹ Inclusion of polarization diversity captures abundant structural and textural information of the medium and allows for the discrimination of different types of scattering mechanisms. Therefore, PolSAR data are an important source for land cover classification.

Generally, the classification scheme could be categorized into three classes. The first class is based on appropriate statistical modeling of PolSAR data. The most well-known method is the Wishart classifier proposed by Lee et al.,² which derives an optimal Bayesian classifier based on the assumption that scattering vectors from a homogeneous region follow a complex joint Gaussian distribution. However, the performance deteriorates in heterogeneous regions because of inaccurate statistical model description, and thus more refined statistical models are needed.^{3,4} Some advanced non-Gaussian models are investigated for characterizing heterogeneity of the scattering medium by incorporating the texture parameter,^{5,6} in which the classification accuracy is improved using more representative statistical model of the data.

The second class aims to utilize the polarimetric parameters or features that deriving from the decomposition theorems, such as H/A/Alpha, Freeman–Durden, and Yamaguchi. Each of these features provides a physical interpretation of the scattering mechanism of the illuminated area.² Recent literatures^{7,8} have shown that a careful combination of multiple features could improve the classification accuracy. Moreover, the incorporation of spatial information is also beneficial for performance improvement.^{9,10}

*Address all correspondence to Mingliang Tao, E-mail: mltao@foxmail.com

The third class mainly concentrates on the design of advanced classifier. De et al.¹¹ proposed that the use of deep learning technique. More advanced classification scheme using decision hierarchical classifier¹² and multilayer autoencoders¹³ are also proposed. The idea of combining the result of multiple classifiers could be found in Ref. 14. The performances of these two kinds of methods rely on a proper selection of used polarimetric features. Without well-selected discriminative features, it would be difficult to obtain high classification accuracy, even if using a rather complex classifier. Otherwise, a simple classifier still can obtain good classification result if well-separable features are provided. Therefore, it is necessary to investigate an effective dimension reduction (DR) method so that the redundancy is reduced and discriminability is enhanced.

Conventional matrix-based DR methods require the rearrangement of image into vectors. The reshaping process breaks the natural structure and correlation in the original data, without effectively utilizing the spatial relationship among neighboring pixels. Moreover, these methods suffer from the so-called curse of dimensionality problem, in which handling high-dimensional samples is computationally expensive and perform poorly with a small number of training samples available. To overcome the deficiency, tensor algebra has drawn a lot of attention and has been extensively applied for data analysis during recent years.¹⁵ Tensor algebra extends the mathematical definitions into higher-dimensional space and is very suitable for characterizing data with coupling correlation among different dimensions. In this work, our goal is to investigate multifeatures combination and incorporate spatial information together within the tensor algebra framework and to develop a pixel-based feature DR method based on the tensor learning techniques for improving the accuracy of PolSAR land cover classification.

First, multiple informative polarimetric descriptors are computed from direct measurements of PolSAR covariance matrix and several effective target decomposition theorems. Then each pixel is modeled as a third-order tensor, where the first two dimensions represent the neighborhood spatial information, and the third represents the feature dimension. Typically, the tensor is of high correlation and redundancy in both the spatial and feature dimensions. Hence, it is reasonable to expect that the tensor objects are embedded in a lower dimensional tensor subspace. A tensor learning technique, i.e., multilinear principal component analysis (MPCA),¹⁶ is applied to find a tensor subspace that captures most of the variation in the input tensor objects. This process serves as a feature extraction step, which would be beneficial for the subsequent classification. Further, all the test tensor samples undergo the same mapping with the training samples via the projection matrices. The projected training samples are utilized to train a specific classifier, such as the k -nearest neighbor (KNN).¹⁷ Subsequently, the projected test samples are fed to the trained classifier, and the classification result is obtained.

The remainder of this paper is structured as follows. Section 2 gives a brief introduction to the tensor algebra and proposes the tensor representation of the PolSAR image pixels. Section 3 introduces the theory and methodology of the proposed tensor learning using MPCA. Section 4 presents the experimental results of simulated and real SAR data, together with thorough performance discussions. Section 5 concludes this paper.

2 Tensor Representation

In this section, to avoid confusion, some fundamental definitions of tensor algebra are introduced first. Then the tensor representation of PolSAR image pixels is discussed, which provides foundation for subsequent classification scheme.

2.1 Basic of Tensor Algebra

A tensor is defined as a multidimensional array, and its number of dimensions is referred to as mode or order. The tensor definition unifies the framework for depicting the data. Scalar, vector, and matrix are special cases of the tensor, with the modes of 0, 1, and 2, respectively. Scalars are denoted by lowercase letters, e.g., x ; vectors are denoted by boldface lowercase letters, e.g., \mathbf{x} ; matrices by boldface capital letters, e.g., \mathbf{X} . For simplicity, only the third-order tensor is considered throughout this paper, which is denoted by Euler script letters, i.e., $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$,

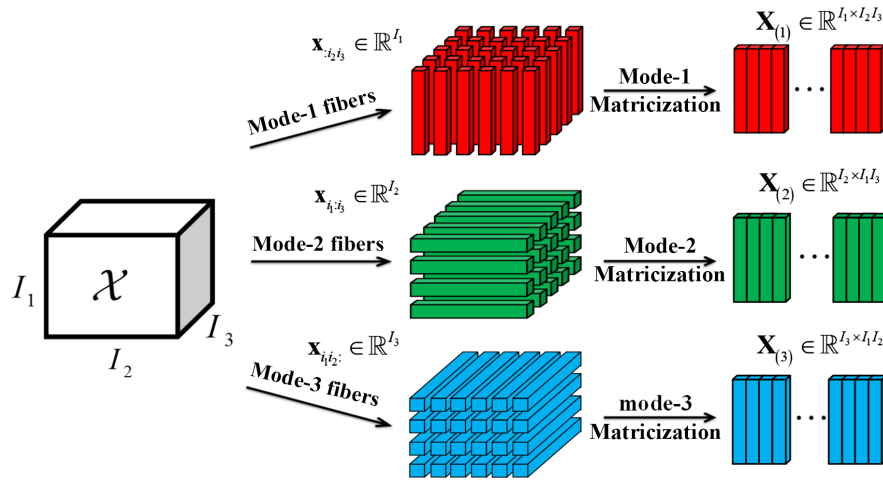


Fig. 1 Illustration of the mode- n fibers and its corresponding matricization process.

whose single entry is a real scalar and expressed as $x_{i_1 i_2 i_3}$, where $1 \leq i_n \leq I_n (1 \leq n \leq 3)$ are indexes along each mode. Similar to the Frobenius norm of a matrix, the norm of a tensor \mathcal{X} is the inner product with itself, i.e., $\|\mathcal{X}\|_F = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} x_{i_1 i_2 i_3}^2}$.

In terms of a tensor, fibers are the higher-order analogue of matrix rows and columns. The mode- n fibers are the I_n -dimensional vectors obtained by varying index i_n , while keeping other indexes fixed. Matricization or unfolding is defined as the process of reshaping a tensor into a matrix. The mode- n matricization of a tensor \mathcal{X} arranges the mode- n fibers as the columns of the resulting matrix, which is denoted by $\mathbf{X}_{(n)}$. For instance, the mode-1 matricization process transforms the original tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ into a matrix $\mathbf{X}_{(1)} \in \mathbb{R}^{I_1 \times I_2 I_3}$, whose columns consist of $I_2 I_3$ mode-1 fibers, with each of size $I_1 \times 1$. Figure 1 illustrates the mode- n matricization process of a third-order tensor.

The mode- n product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ with a real matrix $\mathbf{A} \in \mathbb{R}^{J \times I_n}$ is defined as the multiplication of mode- n unfolding matrix $\mathbf{X}_{(n)}$ and \mathbf{A} , which is expressed as Ref. 15

$$\mathcal{Y} = \mathcal{X} \times_n \mathbf{A} \Leftrightarrow \mathbf{Y}_{(n)} = \mathbf{A} \mathbf{X}_{(n)}, \tag{1}$$

where \times_n denotes the mode- n tensor-matrix product operator. $\mathbf{Y}_{(n)}$ is the mode- n matricization of the resulting third-order tensor \mathcal{Y} . For a series of multiplications, the order of the multiplication is irrelevant

$$\mathcal{Y} = \mathcal{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \Leftrightarrow \mathcal{Y} = \mathcal{X} \times_2 \mathbf{B} \times_1 \mathbf{A}. \tag{2}$$

The above definitions can be easily extended to higher-mode tensor and more mathematical details could be found in Ref. 15.

2.2 Tensor Representation of PolSAR Image Pixels

Under the assumption of monostatic case and medium reciprocity, PolSAR data are expressed as multilook complex data by spatial averaging¹

$$\mathbf{C} = \begin{bmatrix} \langle |S_{HH}|^2 \rangle & \langle \sqrt{2} S_{HH} S_{HV}^* \rangle & \langle S_{HH} S_{VV}^* \rangle \\ \langle \sqrt{2} S_{HV} S_{HH}^* \rangle & \langle 2 |S_{HV}|^2 \rangle & \langle \sqrt{2} S_{HV} S_{VV}^* \rangle \\ \langle S_{VV} S_{HH}^* \rangle & \langle \sqrt{2} S_{VV} S_{HV}^* \rangle & \langle |S_{VV}|^2 \rangle \end{bmatrix}, \tag{3}$$

where \mathbf{C} is the multilook covariance matrix and $\langle \cdot \rangle$ denotes the spatial averaging during multilook processing. S_{HH} , S_{HV} , and S_{VV} correspond to the channel wise complex scattering coefficients.

Table 1 Extracted polarimetric features.

Method	Polarimetric features		
	Correlation coeff. ρ_{12}	Correlation coeff. ρ_{13}	Correlation coeff. ρ_{23}
Measurements	C_{11}	C_{22}	C_{33}
Pauli	Pauli_a	Pauli_b	Pauli_c
Freeman	Freeman_Odd	Freeman_Dbl	Freeman_Vol
Krogager	Krogager_Ks	Krogager_Kd	Krogager_Kh
Van Zyl	VanZyl_Odd	VanZyl_Dbl	VanZyl_Vol
Yamaguchi	Yamaguchi_Odd	Yamaguchi_Dbl	Yamaguchi_Vol

Based on the covariance matrix, several parameters can be directly calculated such as its diagonal elements and correlation coefficients, which provide a good understanding of the relationships among different polarimetric channels. In addition, many effective target decomposition methods have been developed over the past few decades,¹ including the Pauli,¹⁸ Freeman,^{19,20} Krogager,²¹ Van Zyl,²² and Yamaguchi²³ methods. These decomposition theorems provide different polarimetric descriptors that could give an interpretation of canonical scattering mechanisms. Each descriptor has its own strength and weaknesses for discriminating different terrain types. Some literatures have employed different combinations of multiple features, whose experimental results indicate the significant improvement of the classification accuracy under certain experimental conditions.⁷⁻¹⁰

Therefore, multifeatures combination is investigated in this paper. The considered features in Table 1 are selected as Qi et al. did in Ref. 24, which are generally selected in a rather random manner, without further careful consideration. Totally $D = 21$ polarimetric features are calculated as listed in Table 1. All these features are calculated using PolSARpro software and the notations are consistent for better understanding.²⁵ By stacking all the extracted features, the PolSAR image could be characterized by a third-order PolSAR feature tensor $\mathcal{F} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, as shown in Fig. 2. This representation provides a global description for the PolSAR data. By employing a sliding window with the size of $W \times W$ on the feature tensor $\mathcal{F} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, each pixel could be expressed as a third-mode pixel subtensor $\mathcal{P} \in \mathbb{R}^{W \times W \times D}$, where the first two modes represent the neighboring pixels, and the third represents the polarimetric features. This tensor representation provides a local neighborhood description for PolSAR data.

Under traditional matrix algebra framework, the PolSAR feature tensor $\mathcal{F} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is reshaped into a feature matrix $\mathbf{F}_{(3)} \in \mathbb{R}^{I_3 \times I_1 I_2}$, with each column represents the feature vector of a pixel $f \in \mathbb{R}^{I_3}$. It is worth noting that this process is equivalent to the mode-3 matricization of the PolSAR feature tensor. However, this matrix representation uses only the 21 polarimetric attributes of one pixel, which ignores the spatial correlation of scattering coefficients among

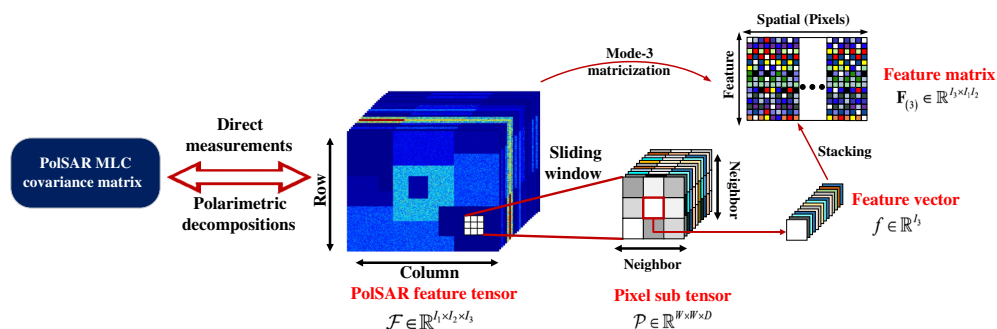


Fig. 2 Global and local tensor representation of a pixel in PolSAR data.

neighboring pixels. By employing a sliding window with the size of $W \times W$ on the feature tensor $\mathcal{F} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, each pixel could be expressed as a third-mode pixel subtensor $\mathcal{P} \in \mathbb{R}^{W \times W \times I_3}$, where the first two modes represent the neighborhood pixels, and the third represents the polarimetric features. Inclusion of polarization diversity captures abundant structural and textural information of the medium. On the contrary, if we only consider spatial correlation without considering polarimetric information, the classification accuracy is rather poor. This is consistent with the fact that single-polarization SAR data are not good candidates for land cover classification.²⁶

Unlike traditional matrix-based representation, the proposed single-pixel tensor representation $\mathcal{P} \in \mathbb{R}^{W \times W \times I_3}$ not only combines multiple features, but also takes the spatial homogeneity into consideration. This local tensor representation conserves the natural data structure and would be beneficial for analysis of local scattering mechanisms, and is the model foundation for subsequent processing.

However, this tensor representation with high dimensionality requires larger memory storage, and thus increases computation complexity. Meanwhile, the redundancy existed in both spatial and feature dimension may pose a hindrance for accurate classification. Therefore, it is necessary to find a tensor subspace that captures most of the variation in the original tensor object, and thus improves the accuracy and efficiency of the classification scheme. In the following, we will provide solutions for this issue.

3 Theory and Methodology of Tensor Learning

In this section, MPCA is introduced as the solution to realize DR for the tensor object. Then the differences between conventional matrix-based methods and proposed tensor-based method are clarified. Further, the flowchart of overall classification scheme is presented.

3.1 Multilinear Principal Component Analysis

The goal of MPCA is to reduce the dimensionality of a tensor consisting of a large number of interrelated variables, while retaining as much as possible the variation present in the original tensor. Given a training set $\{\mathcal{P}_m \in \mathbb{R}^{I_1 \times I_2 \times I_3}, m = 1, \dots, M\}$ consists of M tensors of size $I_1 \times I_2 \times I_3$, and a set of projections matrices $\{\mathbf{U}^{(n)} \in \mathbb{R}^{I_n \times J_n}, n = 1, 2, 3\}$, the projected subtensors $\{\mathcal{Y}_m \in \mathbb{R}^{J_1 \times J_2 \times J_3}, m = 1, \dots, M\}$ are expressed as

$$\mathcal{Y}_m = \mathcal{P}_m \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \times_3 \mathbf{U}^{(3)T}. \tag{4}$$

Equivalently, Eq. (4) can be rewritten in the form of mode- n matricization

$$\mathbf{Y}_{m(1)} = \mathbf{U}^{(1)} \cdot \mathbf{P}_{m(1)} \cdot [\mathbf{U}^{(2)} \otimes \mathbf{U}^{(3)}]^T, \tag{5}$$

$$\mathbf{Y}_{m(2)} = \mathbf{U}^{(2)} \cdot \mathbf{P}_{m(2)} \cdot [\mathbf{U}^{(1)} \otimes \mathbf{U}^{(3)}]^T, \tag{6}$$

$$\mathbf{Y}_{m(3)} = \mathbf{U}^{(3)} \cdot \mathbf{P}_{m(3)} \cdot [\mathbf{U}^{(1)} \otimes \mathbf{U}^{(2)}]^T, \tag{7}$$

where $\mathbf{Y}_{m(n)}$ is the mode- n matricization of the m 'th projected subtensor \mathcal{Y}_m . $\mathbf{U}^{(n)}$ is the n 'th projection matrix. $\mathbf{P}_{m(n)}$ is the mode- n matricization of the m 'th training tensor \mathcal{P}_m . \otimes denotes the Kronecker product.

MPCA seeks to find the best projection matrices $\mathbf{U}^{(n)}$ such that the projected subtensors have maximum energy, which is essentially an optimization problem¹⁶

$$\begin{cases} \max_{\mathbf{U}^{(n)}} & \Psi_{\mathcal{Y}} = \sum_{m=1}^M \|\mathcal{Y}_m - \bar{\mathcal{Y}}_m\|_F^2 \\ \text{s.t.} & \mathbf{U}^{(n)T} \mathbf{U}^{(n)} = \mathbf{I}, \quad n = 1, 2, 3, \end{cases} \tag{8}$$

where $\bar{\mathcal{Y}}_m = \sum_{m=1}^M \mathcal{Y}_m / M$ is the mean tensor of all the projected samples. $\Psi_{\mathcal{Y}}$ denotes the total energy of projected tensors. \mathbf{I} is the identity matrix.

Based on Eq. (4), the energy of projected tensors $\Psi_{\mathcal{Y}}$ can be expanded as

$$\Psi_{\mathcal{Y}} = \sum_{m=1}^M \|(\mathcal{P}_m - \bar{\mathcal{P}}_m) \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \times_3 \mathbf{U}^{(3)T}\|_F^2, \quad (9)$$

where $\bar{\mathcal{P}}_m = \sum_{m=1}^M \mathcal{P}_m / M$ is the mean tensor of all the training samples. According to Eq. (5), $\Psi_{\mathcal{Y}}$ in Eq. (9) can be expressed equivalently in mode-1 matrix representation

$$\begin{aligned} \Psi_{\mathcal{Y}} &= \sum_{m=1}^M \|\mathbf{U}^{(1)T} \cdot [\mathbf{P}_{m(1)} - \bar{\mathbf{P}}_{m(1)}] \cdot [\mathbf{U}^{(2)} \otimes \mathbf{U}^{(3)}]\|_F^2 \\ &= \sum_{m=1}^M \text{tr} \left\{ \mathbf{U}^{(1)T} \cdot [\mathbf{P}_{m(1)} - \bar{\mathbf{P}}_{m(1)}] \cdot [\mathbf{U}^{(2)} \otimes \mathbf{U}^{(3)}] \cdot [\mathbf{U}^{(2)} \otimes \mathbf{U}^{(3)}]^T \cdot [\mathbf{P}_{m(1)} - \bar{\mathbf{P}}_{m(1)}]^T \cdot \mathbf{U}^{(1)} \right\} \\ &= \text{tr}[\mathbf{U}^{(1)T} \cdot \Phi^{(1)} \cdot \mathbf{U}^{(1)}], \end{aligned} \quad (10)$$

where $\text{tr}(\cdot)$ is the trace operator. $\Phi^{(1)} = \sum_{m=1}^M [\mathbf{P}_{m(1)} - \bar{\mathbf{P}}_{m(1)}] \cdot [\mathbf{U}^{(2)} \otimes \mathbf{U}^{(3)}] \cdot [\mathbf{U}^{(2)} \otimes \mathbf{U}^{(3)}]^T \cdot [\mathbf{P}_{m(1)} - \bar{\mathbf{P}}_{m(1)}]^T$.

Similarly, based on Eqs. (6) and (7), $\Psi_{\mathcal{Y}}$ in Eq. (9) also can be rewritten in terms of the mode-2 and mode-3 matricizations:

$$\begin{cases} \Psi_{\mathcal{Y}} = \text{tr}[\mathbf{U}^{(2)T} \cdot \Phi^{(2)} \cdot \mathbf{U}^{(2)}] \\ \Phi^{(2)} = \sum_{m=1}^M [\mathbf{P}_{m(2)} - \bar{\mathbf{P}}_{m(2)}] \cdot [\mathbf{U}^{(1)} \otimes \mathbf{U}^{(3)}] \cdot [\mathbf{U}^{(1)} \otimes \mathbf{U}^{(3)}]^T \cdot [\mathbf{P}_{m(2)} - \bar{\mathbf{P}}_{m(2)}]^T, \end{cases} \quad (11)$$

$$\begin{cases} \Psi_{\mathcal{Y}} = \text{tr}[\mathbf{U}^{(3)T} \cdot \Phi^{(3)} \cdot \mathbf{U}^{(3)}] \\ \Phi^{(3)} = \sum_{m=1}^M [\mathbf{P}_{m(3)} - \bar{\mathbf{P}}_{m(3)}] \cdot [\mathbf{U}^{(1)} \otimes \mathbf{U}^{(2)}] \cdot [\mathbf{U}^{(1)} \otimes \mathbf{U}^{(2)}]^T \cdot [\mathbf{P}_{m(3)} - \bar{\mathbf{P}}_{m(3)}]^T. \end{cases} \quad (12)$$

Therefore, the original optimization problem can be formulated as three-equivalent optimization problems

$$\begin{cases} \max_{\mathbf{U}^{(n)}} \text{tr}[\mathbf{U}^{(n)T} \cdot \Phi^{(n)} \cdot \mathbf{U}^{(n)}] \\ \text{subject to } \mathbf{U}^{(n)T} \mathbf{U}^{(n)} = \mathbf{I}, n = 1, 2, 3 \end{cases}. \quad (13)$$

The formulation of the optimization problem in Eq. (13) is the same with that of the matrix PCA.²⁷ Therefore, the optimal solution to $\mathbf{U}^{(n)}$ is obtained by applying the eigenvalue decomposition on $\Phi^{(n)}$ and assigning the eigenvectors corresponding to the largest J_n eigenvalues as the columns of $\mathbf{U}^{(n)}$. However, the optimal solution to $\mathbf{U}^{(n)}$ depends on other projection matrices, and it is rather difficult to solve all the projection matrices simultaneously. Therefore, an alternative least square (ALS) scheme is applied to iteratively solve the projection matrices. The projection matrices $\mathbf{U}^{(n)}$ define the mapping from the original high-dimensional training tensor into an intrinsic low-dimensional tensor subspace. This tensor subspace is assumed to capture most of the variation in the training tensors set and would benefit for the classification process.

3.2 Comparison with Conventional Matrix-Based Methods

Under the matrix algebra framework, traditional matrix-based DR methods such as PCA,²⁷ independent component analysis (ICA),²⁸ factor analysis (FA),²⁹ and linear discriminate analysis (LDA)²⁹ require to reshape the third-order PolSAR feature tensor $\mathcal{F} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ to the two-dimensional feature matrix $\mathbf{F}_{(3)} \in \mathbb{R}^{I_3 \times I_1 I_2}$, as shown in Fig. 2. Then the feature matrix is multiplied by a transformation matrix so that a reduced feature matrix is obtained. The generative model of the linear matrix-based DR method can be expressed as

$$\mathbf{Y} = \mathbf{V}^T \mathbf{F}_{(3)} \Leftrightarrow \mathcal{Y} = \mathcal{F} \times_3 \mathbf{V}^T, \quad (14)$$

where $\mathbf{Y} \in \mathbb{R}^{p \times l_1 l_2}$ is the projected low-dimensional data matrix, and $\mathbf{V}^T \in \mathbb{R}^{p \times l_3}$ is the projection matrix. It is worth noting that the reduced feature matrix \mathbf{Y} can also be expressed as the mode-3 matricization of a third-order tensor. Therefore, the matrix-based DR methods can be unified into the tensor algebra framework using the definition of tensor–matrix product, as shown in the right part of Eq. (14).

However, the projection in Eq. (14) does not take the spatial relations among neighboring pixels into consideration. It is worth noting that the reduced feature matrix \mathbf{Y} can also be expressed as the mode-3 matricization of a third-order tensor. Comparing Eq. (4) with Eq. (14), we can easily tell the difference between the matrix-based and tensor-based learning techniques. For matrix-based DR methods, only mode-3 feature information is utilized, while ignores the spatial information in mode-1 and mode-2. For the proposed tensor-based scheme, the pixel subtensor $\mathcal{P} \in \mathbb{R}^{w \times w \times l_3}$ is processed directly without further reshaping process. Therefore, a dimensionality reduction algorithm operating directly on a tensor object rather than its vectorized version is desirable. Owing to the iterative ALS procedure, the proposed scheme takes into account the cross dependence between each mode, which means any projection along a given mode depends on the projections along all other modes. The projection matrix for feature extraction is estimated based on the update of the spatial-mode information. In summary, the main differences between the proposed tensor-based method with other matrix-based methods reside in two aspects: data representation and projection matrix optimization. This explains how the spatial information is used in the proposed tensor representation and learning scheme.

3.3 Flowchart of Classification Scheme

Figure 3 summarizes the flowchart of the proposed supervised classification scheme. After obtaining the tensor representation of all the pixels as shown in Sec. 2, a set of training samples and a set of test samples are determined according to the prior knowledge of the ground truth.

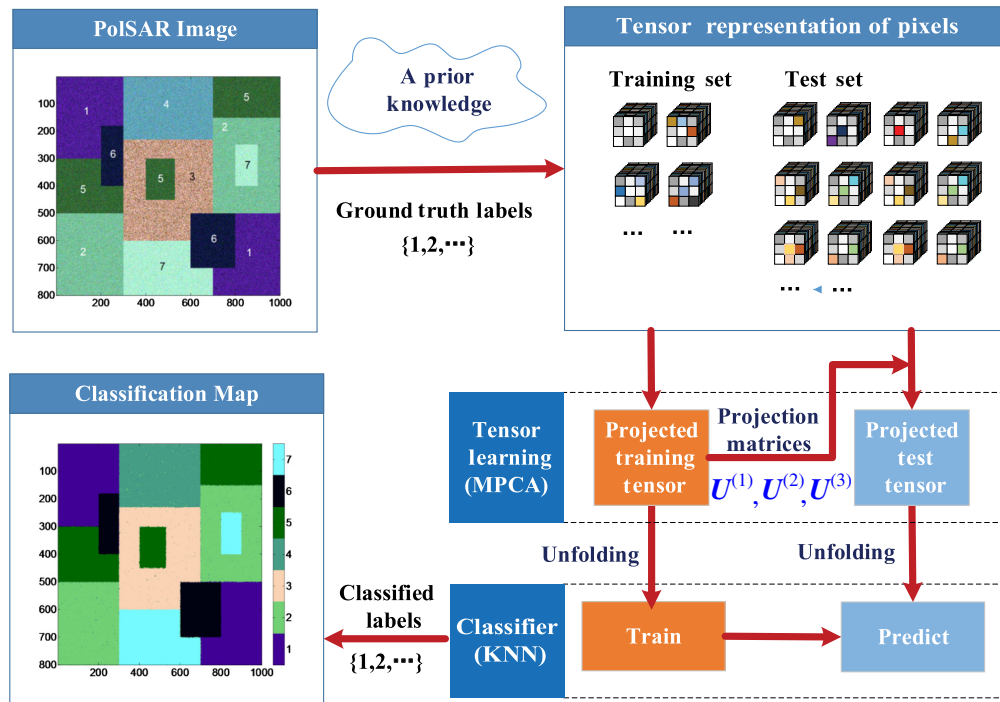


Fig. 3 Flowchart of supervised classification scheme based on tensor representation and learning.

Then the tensor learning technique, i.e., MPCA, is applied on the training tensors set to explore the data structure and find projection matrices along each mode. Further, all the test tensor samples undergo the same mapping with the training samples via the projection matrices. The projected training samples are utilized to train a specific classifier, such as the KNN classifier.¹⁷ Subsequently, the projected test samples are fed to the trained classifier, and the classification result is obtained.

Indeed, advanced classifiers could improve the classification performance to some extent. However, without well-selected discriminative features, it would be difficult to obtain high classification accuracy, even if using a rather complex classifier. Otherwise, a simple classifier still can obtain a good classification result if well-separable features are given. Since our work mainly focuses on the DR process, the rationality behind the choice of KNN is to show that the extracted features by the proposed method are separable so that even a simple classifier can achieve a satisfactory result.

4 Experimental Results and Discussions

Aforementioned sections introduced the theory and flowchart of the proposed classification scheme. In this section, both simulated and real data are used to evaluate the performance of the proposed scheme.

4.1 Results of Simulated Data

For PolSAR data, it is common to assume that the multilook covariance matrix is complex Wishart distributed, whose probability density function is expressed as Ref. 2

$$f(\mathbf{C}_{\text{Wishart}}; l, \Sigma) = \frac{l^d |\mathbf{C}_l|^{l-d} \exp[-l \cdot \text{tr}(\Sigma^{-1} \mathbf{C}_{\text{Wishart}})]}{I(l, d) |\Sigma|^l}, \quad (15)$$

where $\mathbf{C}_{\text{Wishart}}$ is a realization of the covariance matrix, and Σ is the mean covariance matrix. l is the number of looks, d is the dimension of the scattering vector and satisfies $l > d$; $|\cdot|$ is the matrix determinant; $I(l, d) = \pi^{\frac{d(d-1)}{2}} \prod_{i=1}^d \Gamma(l - i + 1)$, and $\Gamma(\cdot)$ is the Gamma function.

In order to well-characterize heterogeneity of the scattering medium, a more refined non-Gaussian product model that incorporating the texture variation and speckle is defined as Refs. 5 and 6:

$$\mathbf{C}_{K\text{Wishart}} = t \cdot \mathbf{C}_{\text{Wishart}}, \quad (16)$$

where t is the positive texture variable with mean value of 1, which defines the spatial variation in the mean backscatter due to target variability. If t follows a gamma distribution, then Eq. (16) defines the K -Wishart model. It is noted that the K -Wishart model degenerates into the Wishart model in the case in the case that t is a constant. Therefore, the K -Wishart model in Eq. (16) is a more general model and it is utilized to simulate data to verify the proposed scheme.

Based on Eq. (16), if given the mean covariance matrices Σ , the number of looks L and the texture parameter t , one can simulate different realizations of the PolSAR covariance matrix.¹ In this experiment, four-look fully polarimetric data are simulated for illustration. It is 800×1000 pixels in size and has seven types of land covers. The mean covariance matrices Σ of each type are extracted from the real datasets of AIRSAR Flevoland dataset.³⁰

Figure 4(a) shows the color-coded Pauli image of the simulated data with different land-cover types and their corresponding class labels. The number of samples for each class is listed in Table 2. All the classes are supposed to follow Gaussian distributed, except that the class 3 is more non-Gaussian distributed considering the graininess-like texture variation. This can be verified from Fig. 4(b), which illustrates the non-Gaussianity of the simulated data using the relative kurtosis as defined in Ref. 5. A larger relative kurtosis value indicates a relatively larger degree of non-Gaussianity. The presence of several straight lines results from the heterogeneous boundary between different classes.

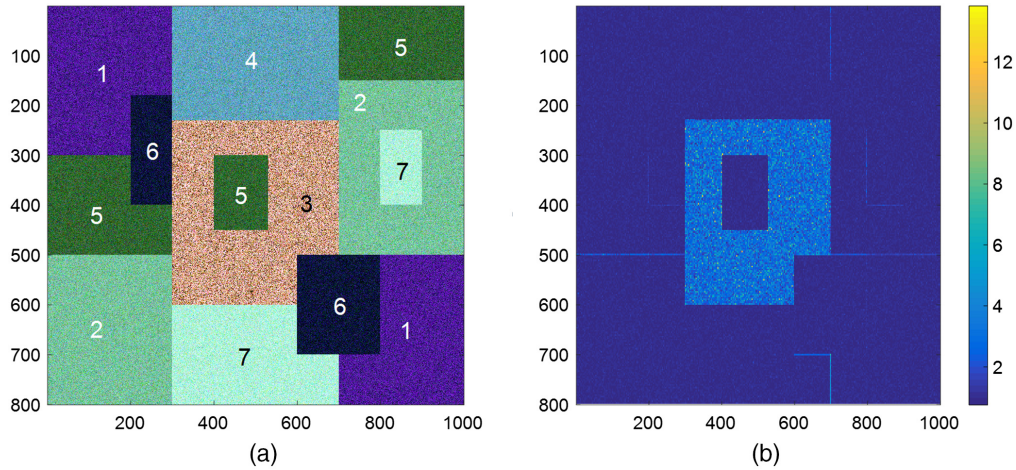


Fig. 4 Simulated data with corresponding class labels: (a) color-coded Pauli image of the original simulated data and (b) illustration of non-Gaussianity using relative kurtosis.

Table 2 Samples for each class in simulated data.

Class label	Total samples	Training samples
1	148,000	1480
2	179,900	1799
3	118,500	1185
4	92,000	920
5	114,500	1145
6	62,000	620
7	85,100	851

According to Table 1, a total of 21 polarimetric features are extracted. To balance the contributions of each feature, all the features are scaled in the range of 0 to 1 with respect to their maximum values. Then each pixel is formulated as a third-order tensor $\mathcal{P} \in \mathbb{R}^{W \times W \times 21}$ as illustrated in Fig. 2. In this simulation, the sliding neighboring window size is set as $W = 5$, and the dimension of projected features is set as $p = 3$ for illustration. Only 1% of the samples is selected as training samples as shown in Table 2.

First, several classical matrix-based linear DR methods, i.e., PCA,²⁷ ICA,²⁸ FA,²⁹ and LDA,²⁹ are applied to obtain a reduced feature set, respectively. It is worth noting that nonlinear methods including ISOMAP, LPP, and Laplacian Eigenmaps require constructing neighborhood graph, which has extensive computation and is very time consuming. Therefore, these nonlinear techniques are not considered for comparison here.

The scatter plots of the reduced feature set by different matrix-based DR methods are illustrated in Figs. 5(a)–5(d). Each dot represents a training sample after DR, and different colors indicate different land-cover types. From Figs. 5(a)–5(c), it is shown that the unsupervised DR methods, i.e., PCA, ICA, and FA, fail to distinguish the samples belonging to different classes, and almost all the samples are mixed together in the feature space. Unsupervised learning cannot properly model underlying structures and characteristics of different classes. LDA is the supervised approach to learn discriminant subspaces by utilizing the *a priori* label information, where the between-class scatter of samples is maximized and the within-class scatter is minimized at the same time. Hence the samples are more discriminative in Fig. 5(d) than that in Figs. 5(a)–5(c). However, the scatter plots in Figs. 5(a)–5(d) indicate that the projected features

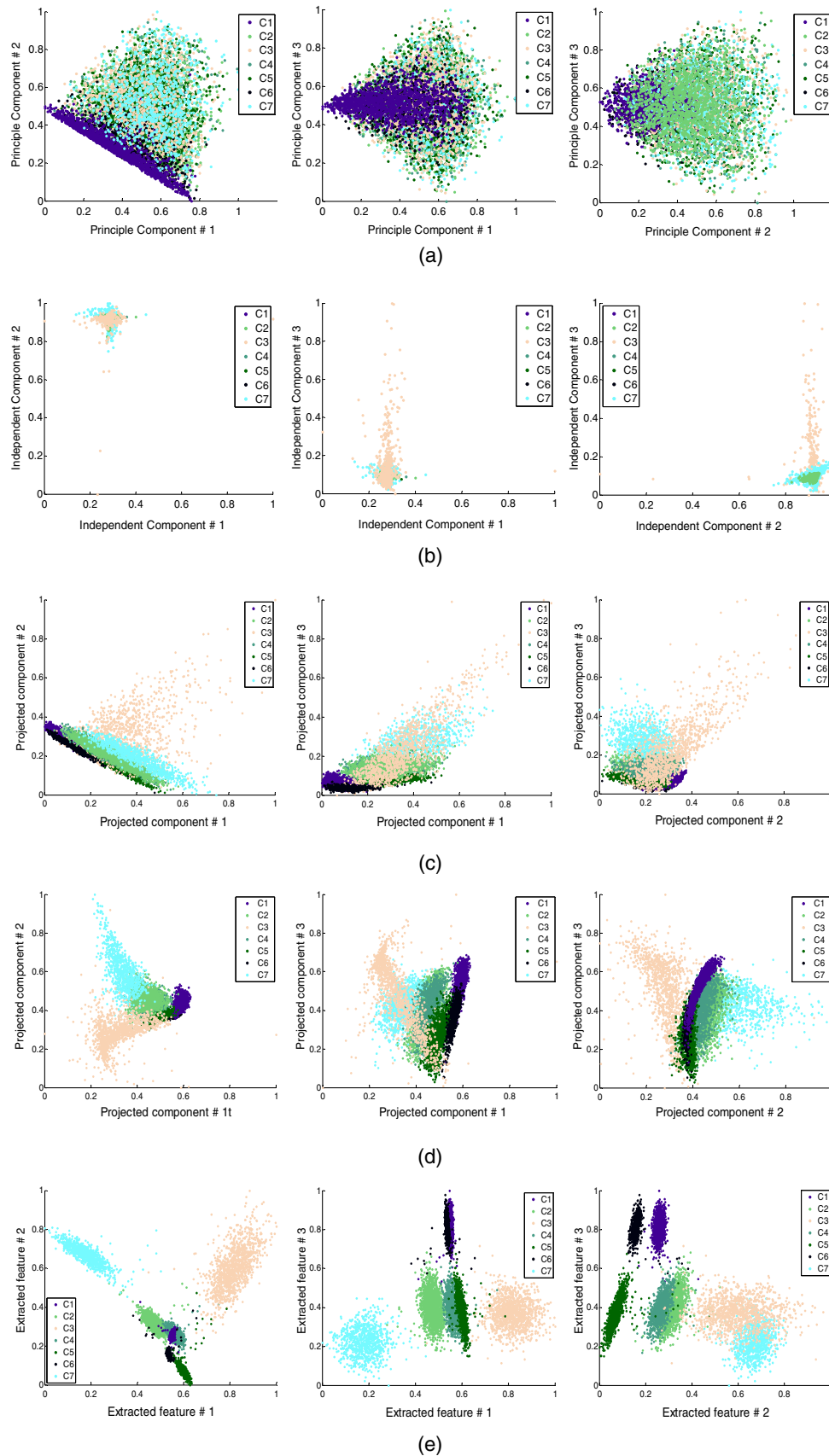


Fig. 5 Scatter plots of extracted intrinsic features obtained by different methods: (a) PCA, (b) ICA, (c) FA, (d) LDA, and (e) MPCA.

are not distinguishable enough, which would pose hindrance for accurate classification even with a rather complex classifier.

Further, the proposed tensor-learning technique, i.e., MPCA, is applied and its resulting scatter plot is presented in Fig. 5(e). It is shown that each class is much more concentrated and different classes are more distinguishable, i.e., the samples have low within-class variation and large between-class discriminability. The proposed method extracts the most salient information embedded in the local tensor object by exploiting the mutual correlation between the spatial and feature dimension. This indicates that the redundancy among features is minimized and a simpler classifier should have rather high classification accuracy.

First, the complex Wishart classifier (CWC)² is applied as an evaluation benchmark. The CWC only utilizes the covariance matrix, which classifies the PolSAR data by exploiting statistical properties under the Bayesian maximum likelihood rule.² A particular result of CWC is shown in Fig. 6(a). It is shown that the CWC suffers from the adverse impact of speckle, and misclassifies many samples in a homogeneous area, especially in class 3.

Then after applying various DR techniques, i.e., PCA, ICA, FA, LDA, and proposed MPCA, one can obtain a low-dimensional intrinsic feature set, respectively. Further, these samples are fed to the KNN classifier as illustrated in Sec. 3.3. For the KNN classifier, the number of nearest neighbors k is chosen by employing fivefold cross validation.

Figures 6(b)–6(e) shows a particular classification result for PCA + KNN, ICA + KNN, FA + KNN, and LDA + KNN, respectively. It is shown that all these matrix-based DR methods lead to very messy classification results because they fail to obtain distinguishable feature sets and suffers from the influence of speckle noise. Under the situation with extremely limited training samples, i.e., only 1% of all the samples, the conventional matrix-based mapping has large generalization error. Although the spatial information is considered, but without well extracted, the redundancy may pose an adverse impact on the classification accuracy.

Figure 6(f) shows a particular classification result for MPCA + KNN. It is shown that many more pixels are correctly classified as a class in the homogeneous area. The tensor representation and learning copes with the problem of insufficient training samples, which is more realistic in practical applications.

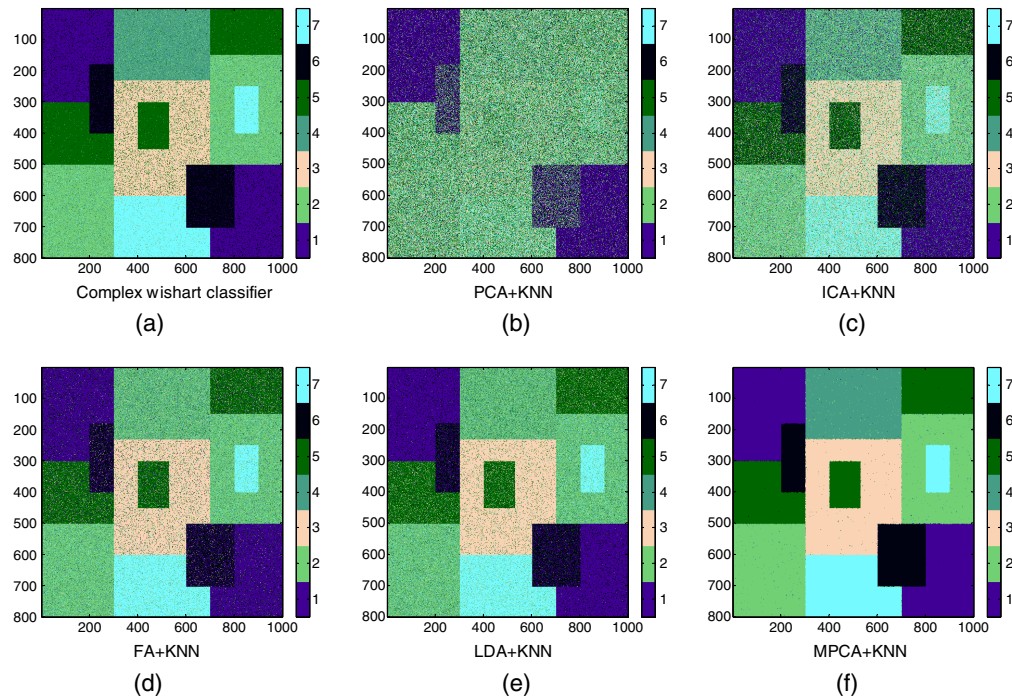


Fig. 6 A particular classification result using different methods: (a) CWC, (b) PCA + KNN, (c) ICA + KNN, (d) FA + KNN, (e) LDA + KNN, and (f) proposed: MPCA + KNN.

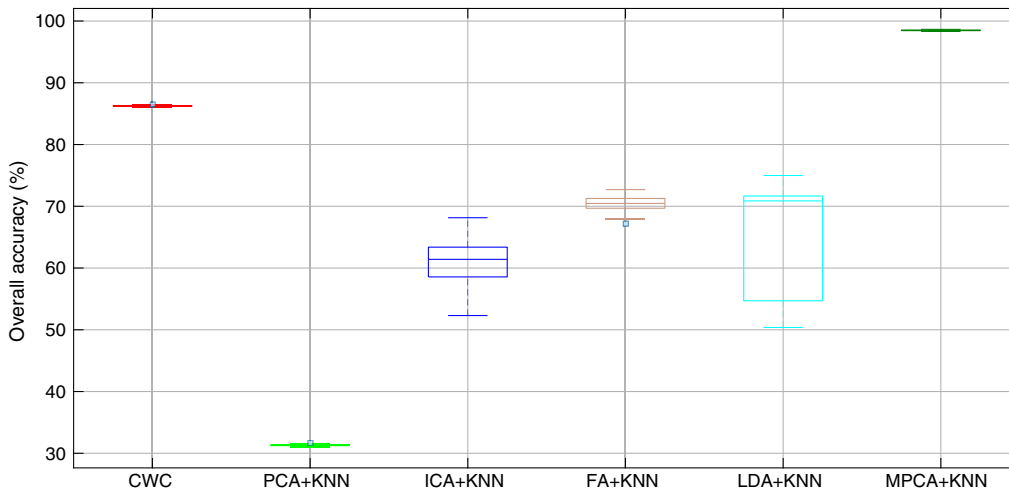


Fig. 7 Comparison of OA values for different combinations of DR method and KNN classifier.

Figure 7 compares the overall accuracy (OA) values for different combinations of DR method and KNN classifier, which is calculated over 100 independent Monte Carlo runs. The dimension of projected features is set as $p = 3$, neighborhood window size $W = 5$, and ratio of the training samples $\rho = 1\%$. These matrix-based methods suffer from the curse of dimensionality, i.e., it is difficult to obtain high OA accuracy with a small number of training samples available. PCA + KNN has the worst performance. The very large deviations of OAs for ICA and LDA indicate their poor generalization performance to the test samples. This phenomenon attributes to the fact that they are not deterministic methods and the projections are highly dependent on the training samples.²⁸ The proposed MPCA has the highest OA values, which indicate its projected features are more discriminative. Even compared with the CWC, the accuracy is improved as much as $\sim 12\%$. In addition, its robustness is proved by rather small deviation of OAs.

The dimensionality of the projected feature p , the ratio of training samples ρ , and the window size W are critical parameters for the DR process. To provide a thorough analysis on the performance of the proposed tensor learning technique, some results on different experimental conditions are presented in the following.

4.1.1 Influence of projected feature dimension

The dimensionality of the projected feature defines a feature space that characterizes the data samples. High-dimensional features would increase the computation burden, which leads to an inefficient classification process. Moreover, for a classifier, a high-dimensional space requires much more training samples to obtain a better classification performance. To provide a comprehensive evaluation, Fig. 8 presents the variation of mean OAs under different

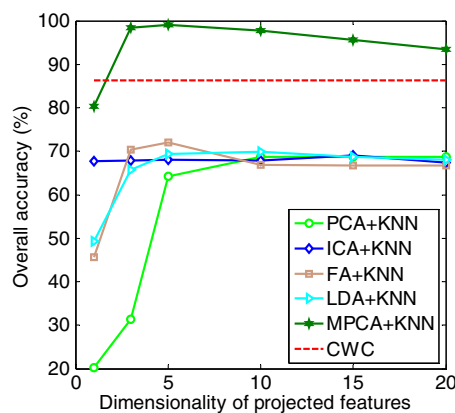


Fig. 8 Simulated data: variation of OA with various dimensions of projected features.

dimensionalities of projected features. It is shown that the mean OA increases with the dimensionality of projected features at first and becomes almost stable subsequently. This proves the validity of the assumption that the significant information of the original higher dimensional is embedded in a low-dimensional feature space. The proposed MPCA-based method has a superior performance.

4.1.2 Influence of training samples size

The training samples are representatives of the data and are used to determine the structure of the classifier. The variability of the training data number has an influence on the classification accuracy. In practical application, the size of the training samples may be limited. Therefore, the performance of the proposed scheme is investigated for different ratios of training samples ranging from 0.1% to 20%. In this comparison, the projected feature dimension is set as $p = 3$ and the neighboring window is set as $W = 5$.

Figures 9(a)–9(d) shows the performance of the conventional matrix-based methods. It is shown that these methods perform poorly when the raining samples are very limited. This attributes to the fact that they suffer from the curse of dimensionality because of the high-dimensional nature of the feature vector. As more training samples are available, these methods have significant performance improvement. This plot highlights the importance of labeled training samples.

As shown in Fig. 9(e), with the number of training points increase, the mean value curve of OAs has a trend of increasing with no $>1\%$. Meanwhile, the deviation decreases because more training samples would mitigate the ambiguity of the classifier. Even when the number of training samples is rather small, i.e., 0.1%, the OA of proposed method is close to 98%, which is superior to the conventional matrix-based methods with 90% training samples available. This demonstrates that the features learned by the proposed scheme are of high generality. The increase of training samples would complicate the training process and slow down the entire classification step. For the proposed scheme, a small training set is able to obtain rather accurate classification result, which could accelerate the training phase and improve efficiency.

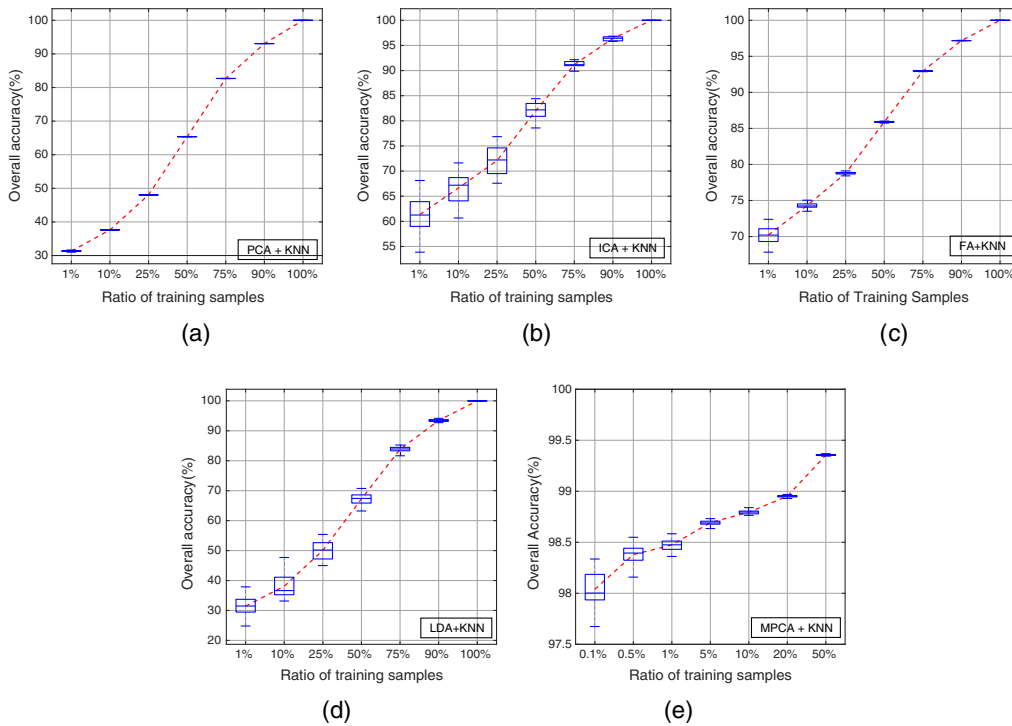


Fig. 9 Simulated data: variation of OA with various percentages of training samples: (a) PCA, (b) ICA, (c) FA, (d) FA, and (e) proposed: MPCA.

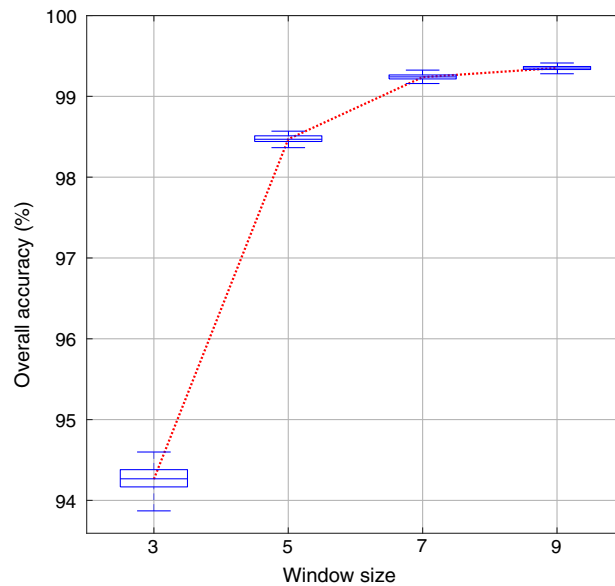


Fig. 10 Simulated data: variation of OA with various window sizes.

The proposed method could cope with the curse of dimensionality and could exploit the available training dataset more effectively.

4.1.3 Influence of neighboring window size

Another important adjustable parameter is the window size that used for tensor representation. In this comparison, the projected feature dimension is set as $p = 3$ and the ratio of training samples is set as $\rho = 1\%$. Figure 10 shows the OAs under various window sizes of the proposed tensor-based methods. It is shown that OA increases with the enlargement of neighboring window size. This phenomenon demonstrates that the tensor representation plays a significant role in improving the classification performance by considering more spatial information. However, a larger window requires more memory storage and also more computation time, which definitely reduces the efficiency. Moreover, a larger window may not be suitable for analyzing scattering areas with high heterogeneity. Therefore, there is a trade-off between efficiency and accuracy.

From the above analysis, it is shown that the conventional matrix-based methods are very sensitive to the ratio of training samples, and the projected feature dimension and neighboring window size have little impact. Under the situation without sufficient training samples, the proposed method is capable of coping with the curse of dimensionality and still has superior generalization performance. The proposed tensor-based classification scheme is less sensitive to the tuning parameters than other methods and has a superior performance under the same parameter configuration.

4.2 Results of Real EMISAR Data

In this part, a real dataset is provided to test the validity of the proposed scheme. This dataset is acquired over the Foulum area by Danish EMISAR.³¹ A subarea is extracted from this dataset with a size of 286×337 pixels. Figure 11(a) shows the color-coded Pauli image of the area. Based on Ref. 5, we manually specify nine types of terrain in this area and the corresponding ground-truth map is shown in Fig. 11(b).

Next, the real Foulum dataset undergoes the same processing flow with the simulated data, as illustrated in Sec. 3.3. Each pixel is represented as a third-mode tensor $\mathcal{P} \in \mathbb{R}^{5 \times 5 \times 21}$, and a training set is determined based on *a priori* ground-truth map. Then different DR methods are applied on the training set to obtain an intrinsic low-dimensional

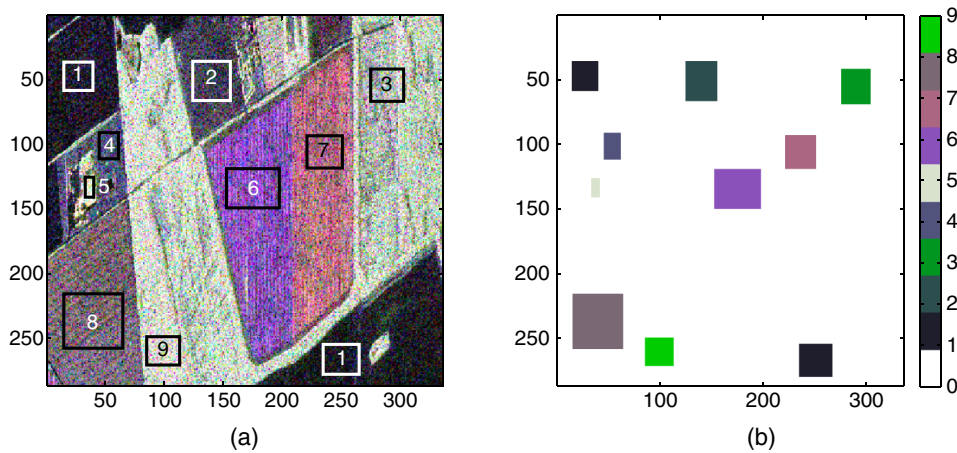


Fig. 11 EMISAR L-band fully polarimetric data of Foulum, Netherlands: (a) color-coded Pauli image and (b) ground-truth map based on Ref. 5.

feature set. Furthermore, the projected intrinsic features are proceeding to the KNN classifier for supervised classification. The experimental settings are the same with the simulated data in Sec.4.1.

Figure 12(a) presents the classification result of CWC. Many misclassification points occur in the homogeneous areas such as type 1 (i.e., right bottom corner) due to the influence of speckle noise. Figures 12(b)–12(e) illustrate the results after applying the PCA + KNN, ICA + KNN, FA + KNN, and LDA + KNN, respectively. It is shown that PCA + KNN has the worst performance, even type 6 and type 7 are mixed together. The LDA + KNN has a comparatively better performance than other methods, because it is a supervised DR method. All these methods including CWC seem to suffer from the adverse impact of speckle noise. Figure 12(f)

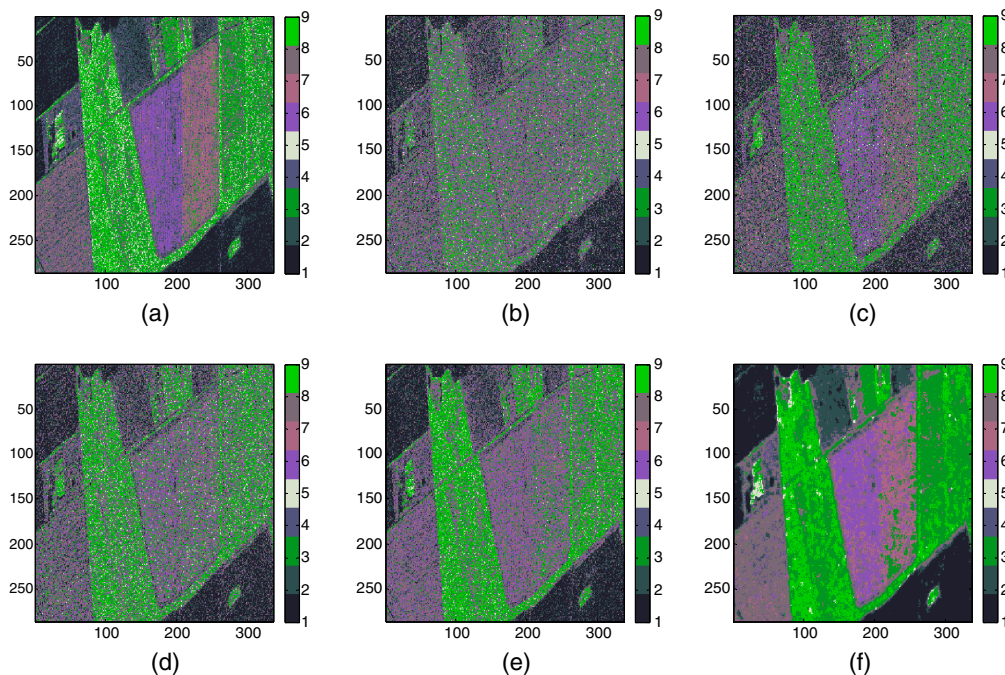


Fig. 12 Real EMISAR data: a particular classification result using different methods: (a) CWC, (b) PCA + KNN, (c) ICA + KNN, (d) FA + KNN, (e) LDA + KNN, and (f) proposed: MPCA + KNN.

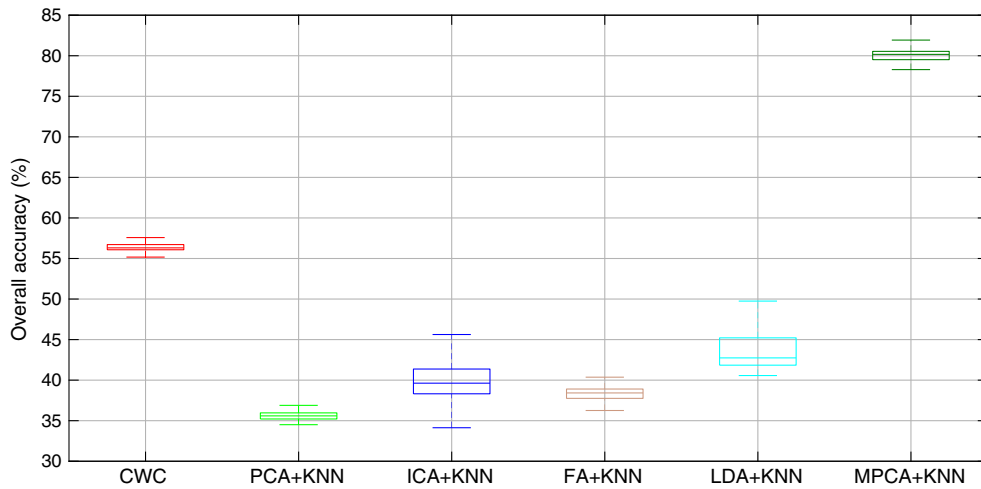


Fig. 13 Real EMISAR data: comparison of OA values for KNN classifier.

shows the result after realizing the proposed MPCA + KNN. It is shown that the classification result is much smoother in homogeneous areas such as type 1. It implies that the introduction of spatial neighboring information could alleviate the speckle noise to some extent, which facilitates the classification and improves the accuracy. Further, the OA values are calculated over 100 independent Monte Carlo runs, as shown in Fig. 13. It is worth noting that only 10% of samples in the ground-truth map are used for training, and all other samples are used to evaluate the OA. Since the total samples of simulated data are almost 9 times than the real data, and thus here the training ratio is set as 10%, not set 1% as in simulated data. According to Fig. 13, it is shown that the proposed scheme has the best performance with a rather large lead in the OA values.

Similarly, as in Sec. 4.1, some further performance discussions are provided regarding with the influence of the projected dimension, ratio of training samples, and window size. Figure 14 plots the variation of OAs with various projected dimensions. Figure 15 shows the variation of OAs for different methods with various percentages of training samples. Figure 16 shows the variation of proposed method with various window sizes. Similar observations with the simulated dataset can be drawn from these figures.

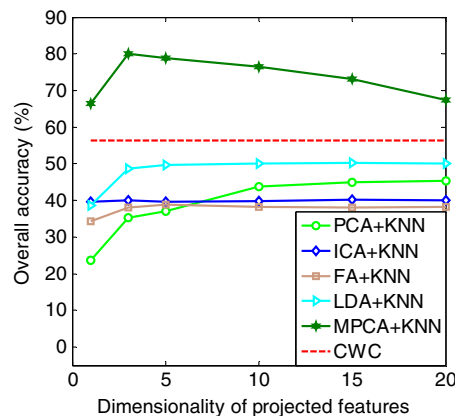


Fig. 14 Real EMISAR data: variation of OA with various dimensions of projected features for KNN classifier.

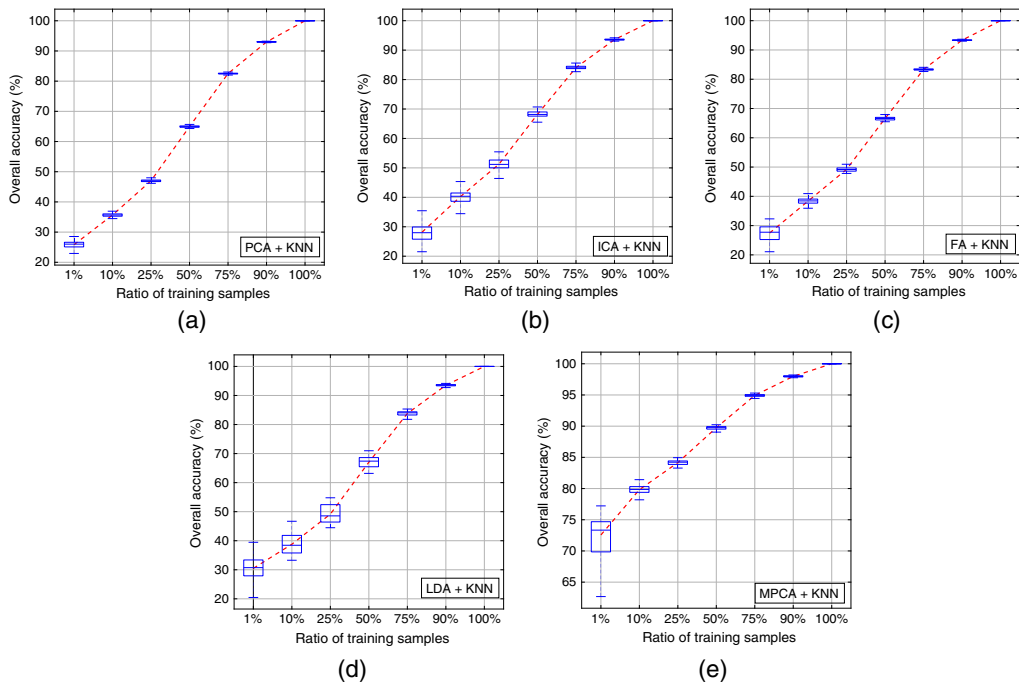


Fig. 15 Real EMISAR data: variation of OA with various percentages of training samples: (a) LDA, (b) ICA, (c) PCA, (d) FA, and (e) proposed: MPCA.

4.3 Results of Real AIRSAR Data

In order to further evaluate the effectiveness of the proposed data, other real data are also tested. This dataset is acquired over Flevoland in Netherlands by the National Aeronautics and Space Administration (NASA)/Jet Propulsion Laboratory (JPL) AIRSAR on August 16, 1989. The PolSAR image used here has a size of 270×250 pixels. Figure 17(a) shows the color-coded Pauli image of the area. The PolSAR image is impaired with speckle noise, which poses a hindrance for image interpretation. Based on Ref. 32, there are totally six terrain types in this area, and the corresponding ground-truth map is shown in Fig. 17 (b).

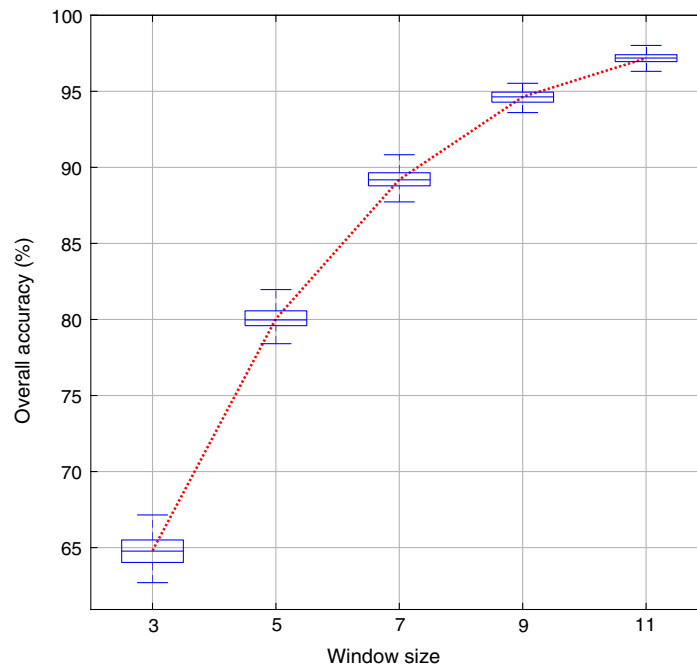


Fig. 16 Real EMISAR data: variation of OA with various window sizes for the proposed method.

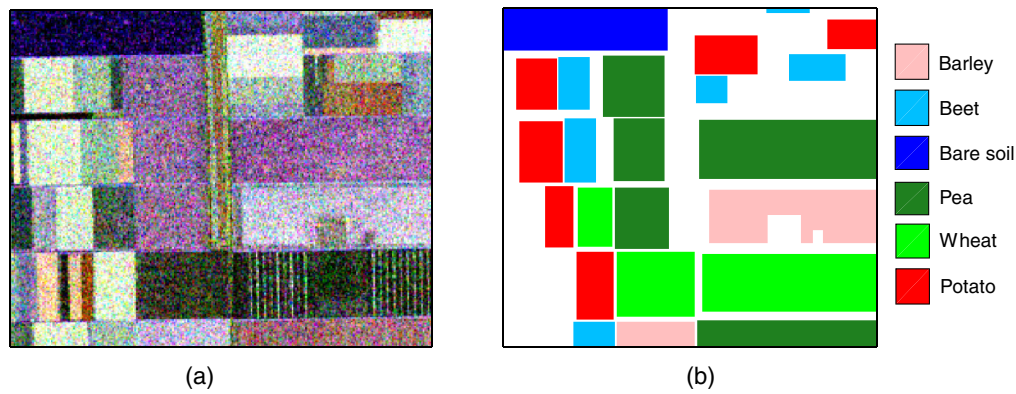


Fig. 17 AIRSAR L-band fully polarimetric data of Flevoland, Netherlands: (a) Color-coded Pauli image and (b) ground-truth map.

In Secs. 4.1 and 4.2, the experiments of simulated data and real EMISAR data are conducted for nonfiltered PolSAR data. In this part, in order to evaluate the effects of filtering on the classification performance, the real AIRSAR data are filtered using Lee refined filter.³³ The projected feature dimension is set as $p = 3$, and the ratio of training samples is set as 10%.

Figure 18 compares the classification results of using existing matrix-based methods and proposed methods. Figures 19 and 20 provide detailed quantitative performance comparisons in terms of the projected feature dimension and training samples size. Meanwhile, Fig. 21 plots the variation of overall classification accuracy with the neighboring window size. From these figures, it is shown that after filtering preprocessing, the OA of proposed method is still the best, even better than the benchmark CWC for nearly 4%. This also indicates the superiority of the proposed tensor-based processing scheme.

4.4 Discussion on Computation Time

Table 3 compares the computational time for each DR method for the simulated data and real data. In this comparison, the projected dimension is fixed as $p = 3$, the ratio of training samples

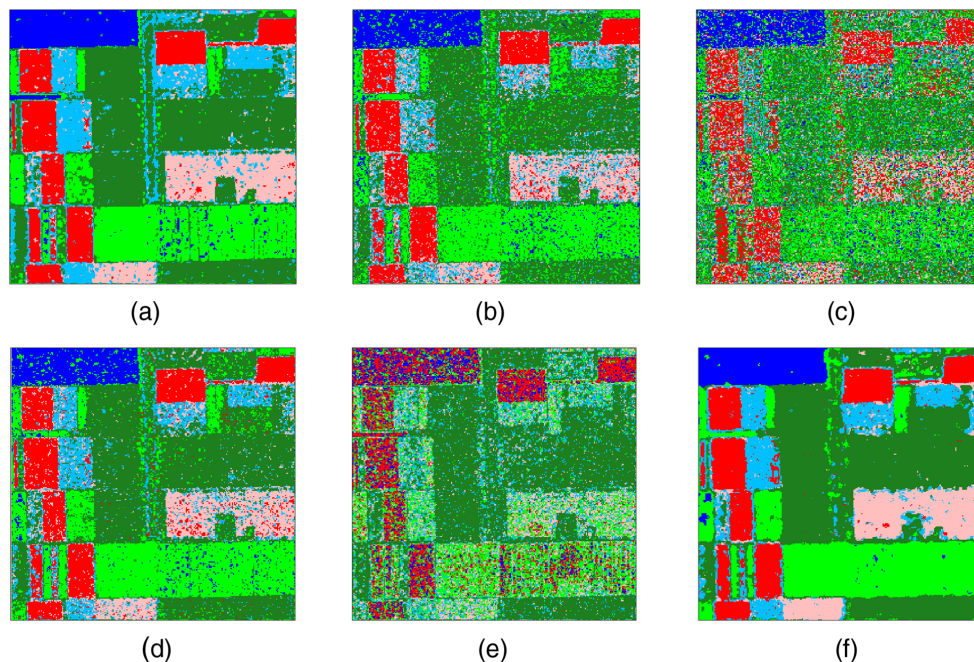


Fig. 18 Real AIRSAR data: a particular classification result using different methods: (a) CWC, (b) PCA + KNN, (c) ICA + KNN, (d) FA + KNN, (e) LDA + KNN, and (f) proposed: MPCA + KNN.

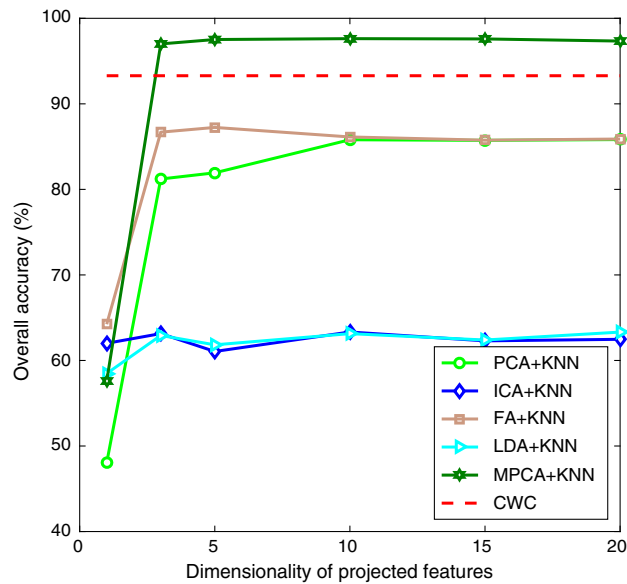


Fig. 19 Real AIRSAR data: variation of OA with various dimensions of projected features for KNN classifier.

is $\rho = 10\%$, and the neighboring window size is $W = 5$. This comparison is realized in a personal laptop with Intel[®] Core i7-6820 using MATLAB R2013a software. The time is averaged under 100 independent runs. It is shown that matrix-based methods are quite efficient, the proposed tensor-based method costs much more time than these matrix-based methods. This is reasonable because the proposed method involves alternative optimizing along the spatial and feature dimension, whereas the matrix-based methods only processing along the feature dimension. The number of samples in the simulated data is larger than that of the real data, and thus

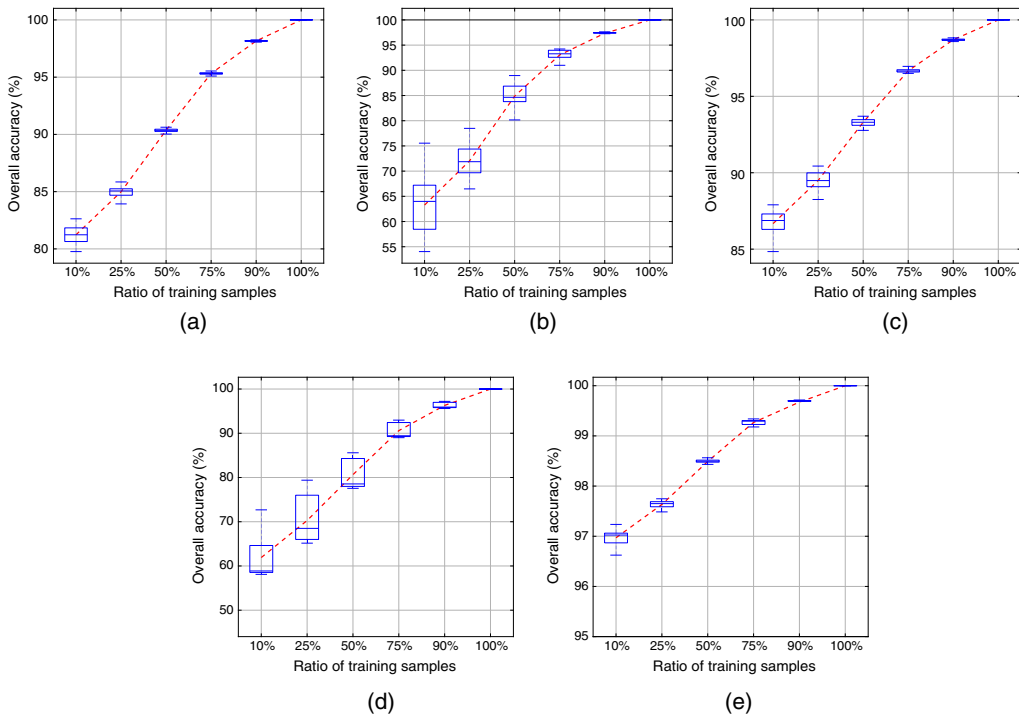


Fig. 20 Real AIRSAR data: variation of OA with various percentages of training samples: (a) LDA, (b) ICA, (c) PCA, (d) FA, and (e) proposed: MPCA.

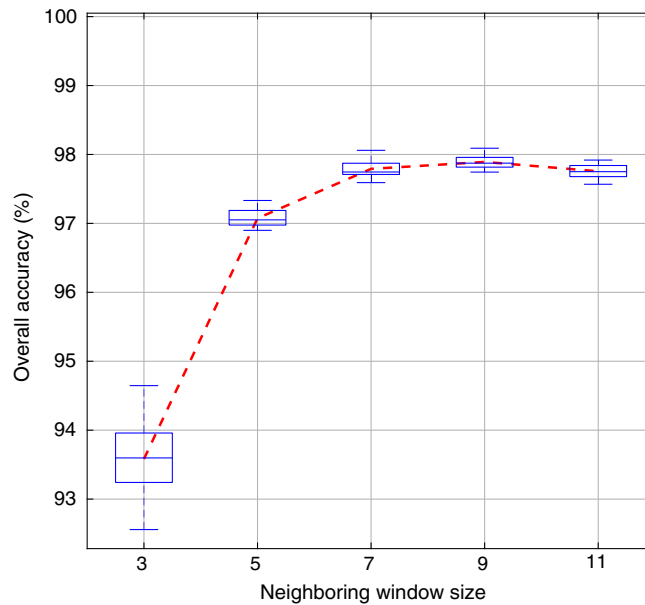


Fig. 21 Real AIRSAR data: variation of OA with various window sizes for the proposed method.

Table 3 Comparison of computation time for each DR method.

Dataset	Method				
	Matrix-based methods				Proposed
	PCA (s)	ICA (s)	FA (s)	LDA (s)	
Simulated data (800 × 1000)	0.0843	0.2373	5.3859	0.1104	201.05
Real EMISAR data (268 × 337)	0.0041	0.0125	0.0489	0.0052	2.82
Real AIRSAR data (270 × 250)	0.0038	0.0115	0.0452	0.0048	2.61

its average time is much higher. The solving process of FA involves an iterative expectation–maximization–optimization under maximum likelihood criterion. As the number of samples increases, more iterations are needed for convergence. Therefore, the cost time of FA value is much higher than normal trend. Although the samples of simulated data are ~ 9 times the real data, the matrix-based method still is quite efficient. The computation time of the proposed technique grows significantly with the increase of processing samples. However, the extra time cost is worthy considering the great improvement of classification accuracy.

5 Conclusions

This paper addresses the land cover classification of PolSAR data within the tensor algebra framework. The novelty of the proposed method lies in two aspects: tensor representation and tensor-based DR. Under the tensor algebra framework, each pixel is modeled as a third-order tensor object. The proposed tensor representation conserves the natural structure of data, in which incorporates the spatial correlation among neighboring pixels and the variation among multiple polarimetric features. The tensor-based DR determines a multilinear projection onto a tensor subspace of lower dimensionality that captures most of the variation present in the original tensorial representation. It improves the discriminability among different classes by jointly considering the polarimetric features and neighboring spatial information.

Performance comparisons with several classical matrix-based DR algorithms on both the simulated and real datasets demonstrate that the proposed classification scheme could greatly improve the classification accuracy while with the ability to alleviate the adverse impacts of speckle noise. The proposed tensor-based classification scheme has a superior performance even when the number of training samples is limited, which is more realistic in many practical applications.

The performance is verified in both airborne simulated and real AIRSAR and EMISAR datasets. Thorough performance comparisons with several classical matrix-based DR algorithms demonstrate that the extracted features are more discriminative in the feature space, and the classification accuracy is significantly improved by at least 10% compared with other existing methods. The proposed method involves alternative optimizing along the spatial and feature dimensions, while the matrix-based methods only process along the feature dimension. Therefore, the computational complexity of the proposed method is greater than other existing matrix-based methods. Similarly, for actual polarimetric satellite SAR data with higher quality in terms of noise equivalent sigma zero (NESZ) and resolution, such as Radarsat-2 and Alos-PalSAR-2, the proposed method approach could also improve the classification accuracy by jointly analyzing the inherent connection between the polarimetric features and neighboring spatial information.

The tensor-based learning approach is a very promising tool for PolSAR data classification. More advanced and efficient tensor learning techniques remain to be investigated. In the future, we would like to investigate the possibility of applying tensor-based techniques for multitemporal PolSAR image.

Acknowledgments

This work was supported by the Science, Technology, and Innovation Commission of Shenzhen Municipality under Grant No. JCYJ20170306154716846. This work was also supported by the National Natural Science Foundation of China under Grant Nos. 61801390 and 61701414, National Postdoctoral Program for Innovative Talents under Grant No. BX201700199, and China Postdoctoral Science Foundation under Grant Nos. 2018M631123 and 2017M623240. This work was also supported by the Fundamental Research Funds for the Central Universities under Grant No. 3102017jg02014.

References

1. J. S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, CRC Press, Boca Raton, Florida (2009).
2. J. S. Lee, M. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sens.* **15**(11), 2299–2311 (1994).
3. W. Gao, J. Yang, and W. Ma, "Land cover classification for polarimetric SAR images based on mixtures models," *Remote Sens.* **6**, 3770–3790 (2014).
4. Y. Wang, T. L. Anisworth, and J. S. Lee, "On characterizing high-resolution SAR imagery using kernel-based mixture speckle models," *IEEE Geosci. Remote Sens. Lett.* **12**, 968–972 (2015).
5. A. P. Doulgeris, S. N. Anfinsen, and T. Eltoft, "Classification with a non-Gaussian model for PolSAR data," *IEEE Trans. Geosci. Remote Sens.* **46**, 2999–3009 (2008).
6. T. Eltoft, S. N. Anfinsen, and A. P. Doulgeris, "A multitexture model for multilook polarimetric synthetic aperture radar data," *IEEE Trans. Geosci. Remote Sens.* **52**, 2910–2919 (2014).
7. A. Buono et al., "Classification of the Yellow River delta area using fully polarimetric SAR measurements," *Int. J. Remote Sens.* **38**, 6714–6734 (2017).
8. L. Zhang et al., "Fully polarimetric SAR image classification via sparse representation and polarimetric features," *IEEE J. Sel. Top. Appl. Earth Obs.* **8**, 3923–3932 (2015).
9. F. Zhang et al., "Nearest-regularized subspace classification for PolSAR imagery using polarimetric feature vector and spatial information," *Remote Sens.* **9**, 1114 (2017).

10. H. Dong et al., "Gaofen-3 PolSAR image classification via XGBoost and polarimetric spatial information," *Sensors* **18**, 611 (2018).
11. S. De et al., "A novel technique based on deep learning and a synthetic target database for classification of urban areas in PolSAR data," *IEEE J. Sel. Top. Appl. Earth Obs.* **11**, 154–170 (2018).
12. H. Kim and A. Hirose, "Unsupervised hierarchical land classification using self-organizing feature codebook for decimeter-resolution PolSAR," *IEEE Trans. Geosci. Remote Sens.* **12**, 1–12 (2018).
13. W. Chen et al., "Classification of PolSAR images using multilayer autoencoders and a self-paced learning approach," *Remote Sens.* **10**, 1–17 (2018).
14. X. Ma et al., "Polarimetric-spatial classification of SAR images based on the fusion of multiple classifiers," *IEEE J. Sel. Top. Appl. Earth Obs.* **7**, 961–971 (2014).
15. T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Rev.* **51**, 455–500 (2009).
16. H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Networks* **19**, 18–39 (2018).
17. P. T. Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sensors* **18**, E18 (2018).
18. S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geosci. Remote Sens.* **34**, 498–518 (1996).
19. A. Freeman and L. Durden, "A three-component scattering model for polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.* **36**, 963–973 (1998).
20. J. S. Lee, T. L. Anisworth, and Y. Wang, "Generalized polarimetric model-based decompositions using incoherent scattering models," *IEEE Trans. Geosci. Remote Sens.* **52**, 2474–2491 (2014).
21. E. Krogager, "New decomposition of the radar target scattering matrix," *Electron. Lett.* **26**, 1525–1527 (1990).
22. J. J. van Zyl, M. Arii, and Y. Kim, "Model-based decomposition of polarimetric SAR covariance matrices constrained for nonnegative eigenvalues," *IEEE Trans. Geosci. Remote Sens.* **49**, 3452–3459 (2011).
23. Y. Yamaguchi et al., "Four-component scattering model for polarimetric SAR image decomposition," *IEEE Trans. Geosci. Remote Sens.* **43**, 1699–1706 (2005).
24. Z. Qi et al., "A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data," *Remote Sens. Environ.* **118**, 21–39 (2012).
25. European Space Agency, "ESA PolSARPro V4. 2," 2011, <http://earth.esa.int/polsarpro> (02 January 2018).
26. J. S. Lee, M. R. Grunes, and E. Pottier, "Quantitative comparison of classification capability: fully polarimetric versus dual and single-polarization SAR," *IEEE Trans. Geosci. Remote Sens.* **39**, 1347–1351 (2001).
27. F. D. Torre, "A least-squares framework for component analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 1041–1055 (2012).
28. A. Hyvärinen, "Independent component analysis: recent advances," *Philos. Trans. R. Soc. A* **371**, 1–19 (2013).
29. A. Sarveniazi, "An actual survey of dimensionality reduction," *Am. J. Comput. Math.* **04**, 55–72 (2014).
30. European Space Agency, "AIRSAR Flevoland Dataset," 2011, <http://earth.esa.int/web/polsarpro/data-sources/sample-datasets> (02 January 2018).
31. European Space Agency, "EMISAR Foulum Dataset," 2011, <https://earth.esa.int/web/polsarpro/data-sources/sample-datasets> (02 January 2018).
32. S. Wang et al., "Unsupervised classification of fully polarimetric SAR images based on scattering power entropy and copolarized ratio," *IEEE Geosci. Remote Sens. Lett.* **10**, 622–626 (2013).
33. J. S. Lee, "Refined filtering of image noise using local statistics," *Comput. Graphics Image Process.* **15**(4), 380–389 (1981).

Mingliang Tao received his BEng and his PhD degrees in signal processing from Xidian University in 2016. In July 2016, he joined Northwestern Polytechnical University (NPU) as an associate professor. His research interests are synthetic aperture radar imaging and data interpretation. He was the recipient of the Young Scientist Award from the International Union of Radio Science (URSI). He was a recipient of the National Postdoctoral Innovation Talent Support Program in China and also a recipient of the Excellent Doctoral Dissertation Award by China Education Society of Electronics in 2017.

Jia Su received his BEng degree in communication engineering and his MEng degree in optical communication both from Guilin University of Electronic Technology, Guilin, China, in 2008 and 2011, respectively, and received his PhD in signal and information processing from Xidian University, Xi'an, China, in 2015. Currently, he is an assistant professor at Northwestern Polytechnical University, Xi'an. His research interests include radar signal processing and time-frequency analysis.

Ling Wang received his BSc, MSc, and PhD degrees in electronic engineering from Xidian University, Xi'an, China, in 1999, 2002, and 2004, respectively. From 2004 to 2007, he worked at Siemens and Nokia Siemens Networks. Since 2007, he has been with the School of Electronic and Information, Northwestern Polytechnical University, Xi'an, and he was promoted to a professor in 2012. His current research interests include array processing and smart antennas, wideband communications, and cognitive radio.