Local Discriminant Canonical Correlation Analysis for Supervised PolSAR Image Classification

Xiayuan Huang, Bo Zhang, Member, IEEE, Hong Qiao, Senior Member, IEEE, and Xiangli Nie, Member, IEEE

Abstract—This letter proposes a novel multiview feature extraction method for supervised polarimetric synthetic aperture radar (PolSAR) image classification. PolSAR images can be characterized by multiview feature sets, such as polarimetric features and textural features. Canonical correlation analysis (CCA) is a well-known dimensionality reduction (DR) method to extract valuable information from multiview feature sets. However, it cannot exploit the discriminative information, which influences its performance of classification. Local discriminant embedding (LDE) is a supervised DR method, which can preserve the discriminative information and the local structure of the data well. However, it is a single-view learning method, which does not consider the relation between multiple view feature sets. Therefore, we propose local discriminant CCA by incorporating the idea of LDE into CCA. Specific to PolSAR images, a symmetric version of revised Wishart distance is used to construct the between-class and within-class neighboring graphs. Then, by maximizing the correlation of neighboring samples from the same class and minimizing the correlation of neighboring samples from different classes, we find two projection matrices to achieve feature extraction. Experimental results on the real PolSAR data sets demonstrate the effectiveness of the proposed method.

Index Terms—Canonical correlation analysis (CCA), dimensionality reduction (DR), local discriminant embedding (LDE), multiview feature extraction, supervised polarimetric synthetic aperture radar (PolSAR) image classification.

I. INTRODUCTION

POLARIMETRIC synthetic aperture radar (PolSAR) is able to provide more detailed information of targets than the single-polarization cases [1], [2]. And land cover classification of PolSAR images is a primary application of PolSAR. According to whether the discriminative information is known, classification can be divided into supervised classification and unsupervised classification. This letter focuses on supervised PolSAR image classification.

Feature extraction is a key step for supervised PolSAR image classification. Multiple distinct feature sets can be used to describe PolSAR images, such as polarimetric features (PFs) and textural features (TFs). Traditionally, PFs generated from the original PolSAR data, such as the scattering matrix, covariance matrix or coherency matrix, and numerous target decompositions, are used for PolSAR image classification [3], [4] In addition, TFs from the gray-level co-occurrence matrix (GLCM) and the Gabor filter have also been applied to PolSAR image classification [5], [6]. Moreover, a few previous studies have attempted to combine PFs and TFs for a better classification performance [7], [8]. Therefore, how to extract proper features from multiple feature sets has a great impact on the performance of PolSAR image classification [8].

Recently, various dimensionality reduction (DR) methods have been utilized to extract features for PolSAR image classification. Concretely, two classical linear methods, i.e., principle components analysis [9] and independent component analysis [10], and some nonlinear methods based on manifold learning, such as Laplacian eigenmaps [11] and supervised graph embedding [12], have been used for PolSAR image classification. Moreover, LDE proposed in [13] can preserve the local structure and the discriminative information of the data well. However, the above methods are single-view learning methods, which cannot make full use of multiview feature sets and the relation between them. Canonical correlation analysis (CCA) is a powerful tool to extract features from two views, but it is unsupervised and the label information is not exploited, which limits its application in classification or recognition [14]. To this end, discriminative CCA (DCCA) was proposed to utilize the label information for two-view feature extraction, but it does not consider the preservation of the structure of the data [15].

Motivated by those above, this letter aims to propose a novel DR method to extract features from two views, i.e., PFs and TFs by incorporating the idea of LDE into CCA. Meanwhile, the proposed method can preserve the discriminative information and the structure of the data well. First, we construct two neighboring graphs based on a symmetric version of revised Wishart (SRW) distance, which is obtained by considering the statistical distribution of the PolSAR images [16]. Then, by making neighboring samples from one class most correlated and neighboring samples from different classes least correlated, we obtain the optimization problem to find two projection matrices for feature extraction.

II. RELATED WORKS

CCA is an unsupervised DR method for the two-view data. Given $n$ samples from two views, i.e., \{$(x_1, y_1)$, $(x_2, y_2)$, ..., $(x_n, y_n)$\}, where $x_i \in \mathbb{R}^p$, $y_i \in \mathbb{R}^q$, and two matrices $X = [x_1, x_2, ..., x_n] \in \mathbb{R}^{p \times n}$ and $Y = [y_1, y_2, ..., y_n] \in \mathbb{R}^{q \times n}$ correspond to the two views. CCA aims to search for two linear transformations $w_X$ and $w_Y$, which make the two transformed variables $w_X^T X$ and $w_Y^T Y$ most correlated, that is
\[
\max_{w_X, w_Y} w_X^T X Y^T w_Y \\
\text{s.t. } w_X^T X X^T w_X = 1, \quad w_Y^T Y Y^T w_Y = 1. \tag{1}
\]

LDE is a supervised DR method for the single-view learning. It aims to preserve the local structure and the discriminative information of the data by making neighboring samples from one class close and neighboring samples from different classes away from each other. Two neighboring graphs: the within-class neighboring graph $G$ and the between-class neighboring graph $G'$ are constructed to represent the locally discriminative information. The optimization problem is as follows:
\[
\max_W \sum_{i,j} \| W^T x_i - W^T x_j \|^2 G_{ij} \\
\text{s.t. } \sum_{i,j} \| W^T x_i - W^T x_j \|^2 G_{ij} = 1 \tag{2}
\]
where $W = [w_1, w_2, ..., w_d] \in \mathbb{R}^{p \times d}$ is the projection matrix.

III. PROPOSED METHOD

For PolSAR images, multiview feature sets can characterize them more comprehensively. And we exploit two-view feature sets, i.e., PFs and TFs in this letter. To be specific, PFs consist of: 1) elements of covariance matrix $C$ ($n_1 = 9$), elements of coherency matrix $T$ ($n_2 = 9$), and their simple transformations ($n_3 = 10$) such as span, correlation coefficient, and depolarization degree and 2) parameters of target decompositions such as Pauli ($n_4 = 3$), Krogager ($n_5 = 3$), Freeman ($n_6 = 6$), H/A/a ($n_7 = 6$), Huynen ($n_8 = 9$), and Van Zyl ($n_9 = 3$). Therefore, the total number of PFs is 58. The details about the above PFs can be found in [1]. TFs include: 1) the GLCM features, which consist of four descriptors (energy, entropy, correlation, and contrast) for a distance of one pixel and four orientations ($n_{10} = 4 \times 4 = 16$) and 2) Gabor features, which are made up of means for five scales and eight orientations ($n_{11} = 5 \times 8 = 40$). Therefore, the total number of TFs is 56. The details about the above TFs can be found in [5] and [8]. In addition, all these TFs are computed in a sliding $11 \times 11$ window.

Therefore, a pixel $i$ in a PolSAR image can be described by a PF vector $x_i \in \mathbb{R}^{58}$ and a TF vector $y_i \in \mathbb{R}^{56}$. Given $n$ training samples with the two-view feature sets, i.e., \{$(x_1, y_1)$, $(x_2, y_2)$, ..., $(x_n, y_n)$\}, two matrices $X = [x_1, x_2, ..., x_n] \in \mathbb{R}^{58 \times n}$ and $Y = [y_1, y_2, ..., y_n] \in \mathbb{R}^{56 \times n}$ correspond to the PF set and the TF set, respectively. The labels of $n$ samples are denoted as \{$l_1, l_2, ..., l_n$\}, where $l_i \in \{1, 2, ..., L\}$. Since the two feature sets are distinctly different, the proposed method aims to find two projection matrices to extract low-dimensional features. Assuming that the reduced dimensionality of the two-view feature sets is $d$, the two projection matrices can be denoted as follows:
\[
W_x = [w_{1x}, w_{2x}, ..., w_{dx}] \in \mathbb{R}^{58 \times d} \quad (d < 58) \tag{3}
\]
\[
W_y = [w_{1y}, w_{2y}, ..., w_{dy}] \in \mathbb{R}^{56 \times d} \quad (d < 56). \tag{4}
\]

A. Constructing Two Neighboring Graphs

Similar to LDE, we first construct two neighboring graphs: the between-class graph $G'$ and the within-class graph $G$ to represent the locally discriminative information [13]. The similarity measure is a key factor for constructing neighboring graphs. Usually, the Euclidean distance is used to search for neighboring samples and compute weights between two samples. However, for PolSAR images, the Wishart distance is the most commonly used similarity measure for classification, because the Wishart distance is derived from the log-likelihood of Wishart distribution. Note that each pixel in the PolSAR image can be represented as the covariance matrix $C$, which can be modeled by a complex Wishart distribution. Furthermore, to meet the four conditions (nonnegativity, definiteness, symmetry, and triangle inequality) for metric, an SRW distance has been proposed [16]. For two $m \times m$ covariance matrices $C_i$ and $C_j$, the SRW distance is defined as follows:
\[
d_{SRW}(C_i, C_j) = \frac{1}{2} \big( \text{tr}(C_i^{-1}C_j) + \text{tr}(C_j^{-1}C_i) \big) - m. \tag{5}
\]

Then $G'$ and $G$ are constructed as follows:
\[
G_{ij} = \begin{cases} 
\frac{d_{SRW}(C_i, C_j)}{m} & \text{if } C_i \in O(k, C_j) \\
\frac{d_{SRW}(C_i, C_j)}{m} & \text{if } C_j \in O(k, C_i) \\
0 & \text{otherwise}
\end{cases} \tag{6}
\]
\[
G'_{ij} = \begin{cases} 
\frac{d_{SRW}(C_i, C_j)}{m} & \text{if } C_i \in O(k, C_j) \\
\frac{d_{SRW}(C_i, C_j)}{m} & \text{if } C_j \in O(k, C_i) \\
0 & \text{otherwise}
\end{cases} \tag{7}
\]
where $l_i$ denotes the label of sample $i$, $O(k, C_i)$ denotes the $k$ nearest samples of $C_i$, and the parameter $t$ is used to adjust the weight.

B. Solving for Two Projection Matrices

Next, we maximize the correlation of neighboring samples from the same class and minimize the correlation of neighboring samples from different classes to maintain the locally discriminative information. For an arbitrary column vector $w_x$ of $W_x$ and an arbitrary column vector $w_y$ of $W_y$, the optimization problem is as follows:
\[
\max_{w_x, w_y} W_x^T M_w w_y - \eta w_x^T M_b w_y \\
\text{s.t. } W_x^T X X^T w_x = 1, \quad W_y^T Y Y^T w_y = 1 \tag{8}
\]
where $M_w$ is the correlation of neighboring samples from one class, i.e., $M_w = \sum_{i,j} x_i y_j^T G_{i,j} = X G Y^T$, and $M_b$ is the
correlation of neighboring samples from different classes, i.e.,

\[ M_b = \sum_{i,j} x_i y_j^T G_{i,j} = X' G' Y^T \]

Based on the maximum margin criterion [17], \( \eta \) can be set to 1. So the optimization problem (8) is formulated as

\[
\begin{align*}
\max_{w_x, w_y} & \quad w_x^T X G Y^T w_y - w_y^T X G' Y^T w_y \\
\text{s.t.} & \quad w_x^T X X^T w_x = 1, w_y^T Y Y^T w_y = 1.
\end{align*}
\]

Then the Lagrange multiplier method is applied to solve problem (9) and the corresponding Lagrange function is

\[
L(w_x, w_y, \lambda) = w_x^T X (G - G') Y^T w_y - \lambda \left( w_x^T X X^T w_x - 1 \right) - \lambda \left( w_y^T Y Y^T w_y - 1 \right).
\]

Sequently, set \( \partial L / \partial w_x = 0 \) and \( \partial L / \partial w_y = 0 \), that is

\[
\begin{align*}
\frac{\partial L}{\partial w_x} &= X (G - G') Y^T w_y - 2 \lambda X X^T w_x = 0 \quad (11) \\
\frac{\partial L}{\partial w_y} &= Y (G - G') X^T w_x - 2 \lambda Y Y^T w_y = 0. \quad (12)
\end{align*}
\]

Based on (11) and (12), problem (8) is transformed into a generalized eigenvalue problem

\[
A w = 2 \lambda B w
\]

where

\[
A = \begin{bmatrix} 0 & X (G - G') Y^T \\ Y (G - G') X^T & 0 \end{bmatrix}, \quad (14)
\]

\[
B = \begin{bmatrix} X X^T & 0 \\ 0 & Y Y^T \end{bmatrix}, \quad (15)
\]

\[
w = \begin{bmatrix} w_x \\ w_y \end{bmatrix}. \quad (16)
\]

The optimal \( w \) is the eigenvector corresponding to the maximum eigenvalue of \( B^{-1} A \), which is equivalent to the solution of the optimization problem in the ratio form, i.e., \( \max_w (w^T A w) / \max_w (w^T B w) \). To avoid computing the inverse of a matrix, we compute \( w \) from the difference form \( \max_w w^T A w - \gamma w^T B w \), where the parameter \( \gamma \) can be adjusted [18]. That is, \( w \) is the eigenvector corresponding to the maximum eigenvalue of \( A - \gamma B \).

After obtaining the \( d \) eigenvectors corresponding to the largest \( d \) eigenvalues, these eigenvectors form the projection matrix

\[
\begin{bmatrix} W_x \\ W_y \end{bmatrix} = \begin{bmatrix} w_{x1} & w_{x2} & \cdots & w_{dx} \\ w_{y1} & w_{y2} & \cdots & w_{dy} \end{bmatrix}. \quad (17)
\]

C. Obtaining Low-Dimensional Features

After obtaining the project matrices \( W_x \) and \( W_y \), the low-dimensional features for two views are \( W_x^T X \) and \( W_y^T Y \), respectively. We stake the two feature sets together like

\[
\begin{bmatrix} W_x^T X \\ W_y^T Y \end{bmatrix}
\]

which is used as features for classification. The algorithm is shown in Algorithm 1.

### Algorithm 1 Proposed Method

**Input:** Polarmetric feature set \( X = [x_1, x_2, \ldots, x_n] \), textural feature set \( Y = [y_1, y_2, \ldots, y_n] \), the labels of training samples \( \{l_1, l_2, \ldots, l_n\} \), the retained dimensionality \( d \).

1. Construct two graphs \( A \) and \( A' \) by cross-validation. The supervised Wishart classifier (SWC), for the equity of experiments and its parameters are chosen by cross-validation. The supervised Wishart classifier (SWC), which is derived from the complex Wishart distribution of

**Output:** The stacked low-dimensional feature sets from two views like \( \begin{bmatrix} W_x^T X \\ W_y^T Y \end{bmatrix} \)

![Fig. 1. Pauli RGB images of two real PolSAR data sets. (a) Flevoland data set. (b) San Francisco Bay data set.](image)

## IV. EXPERIMENTS

The proposed method is evaluated on two real PolSAR data sets: the Flevoland data set and the San Francisco Bay data set, which are public resources from https://earth.esa.int/web/polsarpro. They are the fully PolSAR images of an agricultural area from Flevoland in the Netherlands and the San Francisco Bay in the USA, respectively. Fig. 1 shows their Pauli RGB images and their original sizes are of 750 × 1024 pixels and 900 × 1024 pixels, respectively. Since speckle noise greatly affects the accuracy of image classification [19], the two images are denoised simply by the refined Lee filter [11] with a 7 × 7 window for the subsequent classification. For each class, we select 100 pixels as the training set and the rest of the pixels are used as the testing set.

To demonstrate the effects of the two types of features on image classification, a DR method for single-view features, LDE is used to extract features based on PFs, TFs, and the two types of features (PF+TF). Here, PF+TF denotes that PF and TF are stacked as the vectors for LDE. Furthermore, our method is compared with other multiview learning methods, CCA and DCCA. A support vector machine (SVM) is a very effective classifier. After obtaining extracted features, we use the same classifier, SVM to complete the classification for the equity of experiments and its parameters are chosen by cross-validation. The supervised Wishart classifier (SWC), which is derived from the complex Wishart distribution of
TABLE I

<table>
<thead>
<tr>
<th>Feature</th>
<th>Method</th>
<th>Potatoes</th>
<th>Grass</th>
<th>Beet</th>
<th>Lucerne</th>
<th>Wheat I</th>
<th>Wheat II</th>
<th>Stem beans</th>
<th>Bare soil</th>
<th>Rapeseed</th>
<th>Total accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>LDE</td>
<td>0.9132</td>
<td>0.8515</td>
<td>0.8831</td>
<td>0.7824</td>
<td>0.7099</td>
<td>0.7883</td>
<td>0.9100</td>
<td>0.9352</td>
<td>0.8596</td>
<td>0.8420</td>
</tr>
<tr>
<td>TF</td>
<td>LDE</td>
<td>0.9931</td>
<td>0.9271</td>
<td>0.9964</td>
<td>0.4118</td>
<td>0.6822</td>
<td>0.9593</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.8268</td>
<td>0.8777</td>
</tr>
<tr>
<td>PF+TF</td>
<td>LDE</td>
<td>0.9901</td>
<td>0.9674</td>
<td>0.9949</td>
<td>0.5162</td>
<td>0.7680</td>
<td>0.9904</td>
<td>0.9835</td>
<td>1.0000</td>
<td>0.7895</td>
<td>0.9088</td>
</tr>
<tr>
<td></td>
<td>CCA</td>
<td>0.9879</td>
<td>0.7950</td>
<td>0.9895</td>
<td>0.5347</td>
<td>0.7409</td>
<td>0.9773</td>
<td>0.9967</td>
<td>1.0000</td>
<td>0.7961</td>
<td>0.8632</td>
</tr>
<tr>
<td></td>
<td>DCCA</td>
<td>0.9851</td>
<td>0.7802</td>
<td>0.9826</td>
<td>0.8325</td>
<td>0.8709</td>
<td>0.9773</td>
<td>0.9967</td>
<td>1.0000</td>
<td>0.8947</td>
<td>0.9030</td>
</tr>
<tr>
<td></td>
<td>Our</td>
<td>0.9945</td>
<td>0.8454</td>
<td>0.9949</td>
<td>0.9126</td>
<td>0.9509</td>
<td>0.9988</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9496</td>
<td>0.9451</td>
</tr>
<tr>
<td>SWC</td>
<td></td>
<td>0.9814</td>
<td>0.8636</td>
<td>0.9971</td>
<td>0.7524</td>
<td>0.9013</td>
<td>0.9880</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.9803</td>
<td>0.9268</td>
</tr>
</tbody>
</table>

Fig. 2. Classification maps of the four methods on the Flevoland data set. (a) Denoised subimage. (b) Ground truth. (c) PF/LDE. (d) TF/LDE. (e) PF+TF/LDE. (f) CCA. (g) DCCA. (h) SWC. (i) Proposed method.

PolSAR data, is usually used as a baseline to evaluate PolSAR image classification methods. In our experiments, we set the retained dimensionality $d$ to be 10 for a good performance. Based on experience, we set $k = 10$ and $t = 10$. $\gamma$ is usually a very small number and is set to be 0.001.

For the Flevoland data set, we select a subimage with $200 \times 320$ pixels for experiments and its denoised subimage is shown in Fig. 2(a). This subimage consists of nine different types of fields: stem beans, potatoes, lucerne, winter wheat I, winter wheat II, bare soil, sugar beat, rapeseed, and grass. Fig. 2(b) shows the corresponding ground truth. The classification accuracies of the comparing methods based on different features are shown in Table I, and the classification maps of four methods are displayed in Fig. 2.

For the San Francisco Bay data set, the denoised image is shown in Fig. 3(a). Four classes of land covers are considered, consisting of sea, mountains, grass, and buildings. Some areas from the four classes are used for experiments as shown in Fig. 3(a) and the corresponding ground truth is shown in Fig. 3(b).

From Tables I and II, we can see that TF is a more effective feature for image classification on the Flevoland data set and PF is a more effective feature on the San Francisco data set. For the single-view learning method, i.e., LDE, PF+TF performs better than PF and TF. In addition, DCCA performs better than CCA, because DCCA exploits the label information. Therefore, we can conclude that the combination of two types of features, especially with multiview learning methods and the discriminative information, indeed helps to improve the performance of classification. The fact that SWC performs well demonstrates that the Wishart distance is indeed a superior similarity measure for PolSAR image classification.

The superiority of the proposed method mainly owes to the following three factors.

TABLE II

<table>
<thead>
<tr>
<th>Feature</th>
<th>Method</th>
<th>Sea</th>
<th>Mountains</th>
<th>Grass</th>
<th>Buildings</th>
<th>Total accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF</td>
<td>LDE</td>
<td>0.9718</td>
<td>0.9075</td>
<td>0.4619</td>
<td>0.8839</td>
<td>0.8390</td>
</tr>
<tr>
<td>TF</td>
<td>LDE</td>
<td>0.9050</td>
<td>0.8578</td>
<td>0.5285</td>
<td>0.6246</td>
<td>0.7445</td>
</tr>
<tr>
<td>PF+TF</td>
<td>LDE</td>
<td>0.9510</td>
<td>0.7580</td>
<td>0.7441</td>
<td>0.8178</td>
<td>0.8401</td>
</tr>
<tr>
<td></td>
<td>CCA</td>
<td>0.9755</td>
<td>0.6761</td>
<td>0.6673</td>
<td>0.8317</td>
<td>0.8241</td>
</tr>
<tr>
<td></td>
<td>DCCA</td>
<td>0.9692</td>
<td>0.6835</td>
<td>0.7518</td>
<td>0.8719</td>
<td>0.8507</td>
</tr>
<tr>
<td></td>
<td>Our</td>
<td>0.9781</td>
<td>0.8338</td>
<td>0.7907</td>
<td>0.9303</td>
<td>0.9040</td>
</tr>
<tr>
<td>SWC</td>
<td></td>
<td>0.9999</td>
<td>0.5231</td>
<td>0.9228</td>
<td>0.8824</td>
<td>0.8691</td>
</tr>
</tbody>
</table>
1) Two neighboring graphs are constructed based on the SRW distance, which takes the statistical distribution of the PolSAR images into account.

2) More types of features are exploited to describe PolSAR images more comprehensively. And the multiview learning method is utilized to extract features rather than the single-view learning method, which ignores the relation between different views.

3) The idea of LDE is applied to CCA, which results in achieving multiview feature extraction meanwhile preserving the local structure and the discriminative information of the data.

V. CONCLUSION

This letter proposes a novel multiview feature extraction method by incorporating the idea of LDE into CCA for supervised PolSAR image classification. Two-view feature sets, i.e., PFs and TFs are used to describe PolSAR images. In addition, the SRW distance is utilized to construct the between-class and within-class neighboring graphs. By maximizing the correlation of samples from two views in one class and minimizing the correlation of samples from two views in different classes, we obtain the two projection matrices corresponding to the two views to compute the low-dimensional features. Experimental results on the two real PolSAR data sets demonstrate that the proposed method performs better than other compared methods.

REFERENCES


