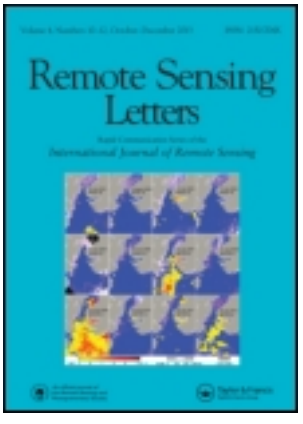


Remote Sensing Letters
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/trsl20

PolSAR image classification using a semi-supervised classifier based on hypergraph learning

Binghui Wei[a], Jun Yu[b], Cheng Wang[a], Hongyi Wu[a] & Jonathan Li[ac]

[a] Department of Computer Science, Xiamen University, Xiamen, China

[b] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China

[c] Faculty of Environment, University of Waterloo, Waterloo, Canada

Published online: 29 Apr 2014.

To cite this article: Binghui Wei, Jun Yu, Cheng Wang, Hongyi Wu & Jonathan Li (2014) PolSAR image classification using a semi-supervised classifier based on hypergraph learning, Remote Sensing Letters, 5:4, 386-395, DOI: 10.1080/2150704X.2014.912765

To link to this article: http://dx.doi.org/10.1080/2150704X.2014.912765

Remote Sensing Letters, 2014
Vol. 5, No. 4, 386–395, http://dx.doi.org/10.1080/2150704X.2014.912765

Taylor & Francis
Taylor & Francis Group

PolSAR image classification using a semi-supervised classifier based on hypergraph learning

Binghui Wei[a], Jun Yu[b], Cheng Wang[a]*, Hongyi Wu[a], and Jonathan Li[a,c]

[a]*Department of Computer Science, Xiamen University, Xiamen, China;* [b]*School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China;* [c]*Faculty of Environment, University of Waterloo, Waterloo, Canada*

(*Received 10 December 2013; accepted 30 March 2014*)

This letter presents a novel semi-supervised method based on hypergraph learning for polarimetric synthetic aperture radar (PolSAR) image classification. Compared with the classic support vector machine, simple-graph learning, *k*-nearest neighbour (*k*-NN) and semi-supervised discriminant analysis (SDA) classifiers, the proposed method achieves better performance with fewer labelled points for PolSAR imagery. A hyperspectral image is used for comparison with use of PolSAR imagery, and the proposed method is found to be inferior to *k*-NN and SDA for the hyperspectral image. The performance of our method is evaluated in single, dual and full-polarization cases, respectively. The results demonstrate that the performance of our method in the full-polarization case is superior to that in either single or dual-polarization case.

1. Introduction

Currently, there is widespread interest in the development of polarimetric synthetic aperture radar (PolSAR) for earth observation because of its all-weather day and night capabilities. How to accurately and robustly classify PolSAR imagery is one of the most active research topics nowadays. Several classification algorithms have been developed in recent years, including expectation-maximization (Beaulieu and Touzi 2004), support vector machines (SVMs) (Lardeux et al. 2009), spectral graph partitioning (Ersahin, Cumming, and Ward 2010), wavelet texture models (De Grandi, Lee, and Schuler 2007), etc. In general, PolSAR classification algorithms can be characterized as supervised, semi-supervised or unsupervised, according to whether there is training data to participate in learning. PolSAR classification algorithms also can be characterized according to methods based on target decomposition or statistical property by feature expression.

In supervised classification, training data sets for each class are selected based on reference data maps or scattering contrast differences in PolSAR imagery. The unsupervised classification method classifies images automatically by locating clusters based on a specific criterion. The final class identification must be inferred manually (Lee and Pottier 2009). The misclassification ratio is more difficult to define. The semi-supervised method is a compromise between the supervised and unsupervised methods. Supervised classification for SAR data usually encounters a problem of time consumption and it is costly to obtain the reference data; unsupervised classification encounters trouble locating the final class identification. Therefore, the highly desirable method is the one that achieves better performance with fewer training samples, without the trouble of locating the final class

*Corresponding author. Email: cwang@xmu.edu.cn

identification. For PolSAR classification, the emphasis is on the latter taxonomy. For the method based on target decomposition, each class is determined by diverse after-scattering and polarization features. For the method based on a statistical property, the featured expression for training and testing data is extracted from the statistical property.

In this letter, we focus on the semi-supervised classification method and develop a novel method based on hypergraph learning (HGL) (Zhou, Huang, and Scholkopf 2006; Agarwal, Branson, and Belongie 2006; Huang et al. 2010; Yu, Rui, and Tao 2014; Yu et al. 2014; Yu, Rui, and Chen 2014). Our proposed method is a general transductive learning method (Wu and Scholkopf 2007) that explores labelled and unlabelled data simultaneously (Yu et al. 2011). Our method regularizes the hyperedge weights and optimizes the labels and hyperedge weights simultaneously.

2. Method

2.1. Feature description

In this letter, we describe a two-step feature. The first step is derived from Pauli decomposition; the second takes into account speckle reduction.

Because its physical meaning is explicit, target decomposition is the main method used to extract features from PolSAR data. Thus, much research is devoted to target decomposition. The simple case is that of Pauli decomposition:

$$\boldsymbol{x} = \left(|S_{\mathrm{hh}} + S_{\mathrm{vv}}| \quad |S_{\mathrm{hh}} - S_{\mathrm{vv}}| \quad 2|S_{\mathrm{hv}}| \right)^{\mathrm{T}}, \tag{1}$$

where S_{pq} denotes the scattering matrix element corresponding to the pq polarization of a receiving–transmitting wave (p and q referring to horizontal h or vertical v linear polarization); the notation $|\cdot|$ denotes the amplitude of the wave after scattering; the superscript T denotes matrix transposition, and the notation $\boldsymbol{x}$ is a PolSAR feature vector. S_{hv} is equal to S_{vh} for reciprocity in most cases according to Pauli decomposition.

Then, we set that, $\boldsymbol{x}_{\mathrm{s}}$, $\boldsymbol{x}_{\mathrm{d}}$, $\boldsymbol{x}_{\mathrm{f}}$ are the feature vectors of single-polarization, dual-polarization and full-polarization, respectively, as follows:

$$\begin{aligned} \boldsymbol{x}_{\mathrm{f}} &= \left(<|S_{\mathrm{hh}} + S_{\mathrm{vv}}|> \quad <|S_{\mathrm{hh}} - S_{\mathrm{vv}}|> \quad 2<|S_{\mathrm{hv}}|> \right)^{\mathrm{T}} \\ \boldsymbol{x}_{\mathrm{d}} &= \left(<|S_{\mathrm{hh}}|> <|S_{\mathrm{vv}}|> \right)^{\mathrm{T}} \\ \boldsymbol{x}_{\mathrm{s}} &= \left(<|S_{\mathrm{hh}}|> \mid <|S_{\mathrm{vv}}|> \right) \end{aligned}, \tag{2}$$

where the notation $< \cdot >$ stands for the spatial averaging over a 3×3 (or other size) neighbourhood. Speckle appearance in SAR images is due to the coherent interference of waves reflected from many elementary scatters. This effect causes a pixel-to-pixel variation in intensities, and the variation manifests itself as a granular noise pattern in SAR images. Speckles reduce the image classification accuracy. Here, we take simple averaging to reduce the speckle noise (Lee, Grunes, and De Grandi 1999).

2.2. Hypergraph learning

We start from a simple graph whose vertices represent samples, and each edge connects two related vertices. The simple graph denotes pairwise relationships and is created based upon pairwise distance. We classify samples with graph-based learning in a feature space.

However, simple graph learning (SGL) fails to capture high-order information, such as the information of clustering, which causes the development of hypergraph learning (HGL). A hyperedge in a hypergraph links several (two or more) vertices that show the clustering information.

A brief description of the hypergraph construction is introduced. Given c categories of PolSAR image, including m training PolSAR image points $(\boldsymbol{x}_1, \boldsymbol{y}_1)$, …, $(\boldsymbol{x}_m, \boldsymbol{y}_m)$ and n testing data points $(\boldsymbol{x}_{m+1}, \boldsymbol{0})$, …, $(\boldsymbol{x}_{m+n}, \boldsymbol{0})$, where $\mathbf{x}_i \in \mathrm{R}^d, 1 \le i \le m+n$ is the input space; $\mathbf{y}_i = [0, \cdots, 1, \cdots, 0]^{\mathrm{T}} \in \Re^{c\times 1}, 1 \le i \le m$ is the label vector of $\boldsymbol{x}_i$, where the gth component is 1 if $\boldsymbol{x}_i$ belongs to the gth category, otherwise, 0; and $\boldsymbol{0}$ is a vector with c components 0.

The hypergraph, $\mathbf{H} = (\mathbf{x}, \boldsymbol{\varepsilon})$, is formed by the vertex set, $\mathbf{x}$, and the hyperedge set, $\boldsymbol{\varepsilon}$, An incidence matrix, $\mathbf{A}$, whose size is $|\mathbf{x}| \times |\boldsymbol{\varepsilon}|$, denotes the hypergraph with the following elements:

$$\mathbf{A}(i,j) = \begin{cases} 1, \text{if } \boldsymbol{x}_i \in \boldsymbol{\varepsilon}_j \\ 0, \text{if } \boldsymbol{x}_i \notin \boldsymbol{\varepsilon}_j \end{cases}, \tag{3}$$

where $\boldsymbol{\varepsilon}_j$ denotes the jth subset of the hyperedge set. The distance between two data points is

$$\begin{aligned} \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}_j) &= \exp\left(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{\sigma^2}\right) \\ \sigma &= \sqrt{\frac{1}{m+n-1}\sum_{\delta=1}^{m+n}\left\|\boldsymbol{x}_\delta - \overline{\boldsymbol{x}}\right\|^2} \quad \overline{\boldsymbol{x}} = \frac{1}{m+n}\sum_{\delta=1}^{m+n}\boldsymbol{x}_\delta \end{aligned}. \tag{4}$$

We use $\boldsymbol{\varphi}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$ to denote diagonal matrices of vertex degrees, hyperedge degrees and hyperedge weights, respectively. $\boldsymbol{\varphi}_i$ denotes the entry (i, i) of matrix $\boldsymbol{\varphi}$; $\boldsymbol{\omega}_i$ and $\boldsymbol{\gamma}_i$ have similar meanings. Then, the initial weight, $\boldsymbol{\omega}_i$, is

$$\boldsymbol{\omega}_i = \sum_{\boldsymbol{x}_j \in \boldsymbol{\varepsilon}_i} \operatorname{dist}(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{5}$$

Based on $\mathbf{A}$, the ith vertex degree, $\boldsymbol{\varphi}_i$, is

$$\boldsymbol{\varphi}_i = \sum_{\boldsymbol{\varepsilon}_i \in \boldsymbol{\varepsilon}} \boldsymbol{\omega}_j \mathbf{A}(i,j), \tag{6}$$

and the ith hyperedge degree, $\boldsymbol{\gamma}_i$, is

$$\boldsymbol{\gamma}_i = \sum_{\boldsymbol{x}_j \in \boldsymbol{x}} \mathbf{A}(j,i). \tag{7}$$

In this letter, we adopt the regularization framework proposed in Yu, Tao, and Wang (2012), i.e.,

$$\begin{aligned} &\arg\min_{\mathbf{F},\boldsymbol{\omega}} \sum_{\eta=1}^{c} \left(\boldsymbol{f}_\eta^{\mathrm{T}} \mathbf{L} \boldsymbol{f}_\eta\right) + \lambda \left\|\boldsymbol{f}_\eta - \mathbf{Y}_\eta\right\|^2 + \mu \|\mathrm{diag}(\boldsymbol{\omega})\|^2 \\ &\text{s.t.} \sum_{j=1}^{l} \boldsymbol{\omega}_j = 1, \boldsymbol{\omega}_j \geq 0,\ j = 1, \cdots, l \end{aligned}, \tag{8}$$

where $\mathbf{L} = \mathbf{I} - \boldsymbol{\varphi}^{-(1/2)} \mathbf{A} \boldsymbol{\gamma}^{-1} \boldsymbol{\omega} \mathbf{A}^{\mathrm{T}} \boldsymbol{\varphi}^{-(1/2)}$; l is the number of hyperedges; diag($\boldsymbol{\omega}$) is the diagonal vector of $\boldsymbol{\omega}$, i.e., ($\boldsymbol{\omega}_1$, $\boldsymbol{\omega}_2$, …, $\boldsymbol{\omega}_l$); c is the number of classes; $\mathbf{F}$ is a matrix, $\mathbf{F} = (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_c) \in \Re^{(m+n)\times c}$, $\boldsymbol{f}_\eta$ is the classification function for the data belonging to the ηth category; $\lambda > 0$ and $\mu > 0$ are two trade-off parameters to balance the empirical loss, the weight and the regularization; $\mathbf{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_m, \mathbf{0}, \ldots, \mathbf{0})^{\mathrm{T}}$, $\mathbf{Y} \in \Re^{(m+n)\times c}$, $\mathbf{Y}_\eta$ is the ηth column of $\mathbf{Y}$. The third term is introduced to avoid a degenerate solution caused by the former two terms existing only in the regularization. Here we add two constraints, one to fix the summation of the weights, $\sum_{j=1}^{l} \boldsymbol{\omega}_j = 1$ and one to avoid negative weights, $\boldsymbol{\omega}_j \geq 0$

Because the classifier function is not jointly convex with respect to $\mathbf{F}$ and $\boldsymbol{\omega}$, we solve one variable by fixing another variable.

First we initialize $\boldsymbol{\omega}$ with Equation (6), so the solution of $\mathbf{F}$ becomes

$$\mathbf{F} = \frac{\lambda}{\lambda + 1} \left(\mathbf{I} - \frac{\boldsymbol{\varphi}^{-(1/2)} \mathbf{A} \boldsymbol{\gamma}^{-1} \boldsymbol{\omega} \mathbf{A}^{\mathrm{T}} \boldsymbol{\varphi}^{-(1/2)}}{1 + \lambda}\right)^{-1} \mathbf{Y}. \tag{9}$$

Then we update the weights, $\boldsymbol{\omega}$, with an iterative coordinate descent method as follows:

$$\begin{cases} \boldsymbol{\omega}_i^{t+1} = 0, \boldsymbol{\omega}_j^{t+1} = \boldsymbol{\omega}_i^t + \boldsymbol{\omega}_j^t, \text{ if } 2\mu\left(\boldsymbol{\omega}_i^t + \boldsymbol{\omega}_j^t\right) + \left(s_j - s_i\right) \leq 0 \\ \boldsymbol{\omega}_i^{t+1} = \boldsymbol{\omega}_i^t + \boldsymbol{\omega}_j^t, \boldsymbol{\omega}_j^{t+1} = 0, \text{ if } 2\mu\left(\boldsymbol{\omega}_i^t + \boldsymbol{\omega}_j^t\right) + \left(s_i - s_j\right) \leq 0 \\ \boldsymbol{\omega}_i^{t+1} = \frac{2\mu\left(\boldsymbol{\omega}_i^t + \boldsymbol{\omega}_j^t\right) + \left(s_j - s_i\right)}{4\mu}, \boldsymbol{\omega}_j^{t+1} = \boldsymbol{\omega}_i^t + \boldsymbol{\omega}_j^t - \boldsymbol{\omega}_i^{t+1}, \text{otherwise} \end{cases}, \tag{10}$$

where $s_i = \sum_{z=1}^{c} -\left(\boldsymbol{r}_i^z\right)^2 \left(\boldsymbol{\gamma}_i\right)^{-1}$, $\boldsymbol{r}^z = [\boldsymbol{f}_z^{\mathrm{T}} \boldsymbol{\varphi}^{-1/2} \mathbf{A}$, $\boldsymbol{r}_i^z$ is the ith component of $\boldsymbol{r}^z$; $\boldsymbol{\omega}_i^t$ denotes the tth iteration of $\boldsymbol{\omega}_i$ and the initial iteration t is 0. Based on the coordinate descent method, an iterative process alternately updates the labels and the weights.

In the next iteration, we calculate the new $\mathbf{F}$ with new $\boldsymbol{\omega}$. The iteration ceases at a given state. After obtaining $\mathbf{F}$, we set the gth class to the ith data point if the gth component is the maximum in the ith row of F. A more detailed solution of Equation (9) can be found in Yu, Tao, and Wang (2012).

3. Experiments

The experiments were performed with four data sets, including two Airborne Synthetic Aperture Radar (AIRSAR) data sets, a Radarsat-2 data set and an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) hyperspectral data set. In order to assess the performance of the proposed method with polarization data, we compared our method with the classic SVM, SGL, k-nearest neighbour (k-NN) (Duda, Hart, and Stork 2000) and semi-supervised discriminant analysis (SDA) (Cai, He, and Han 2007) classifications. Further, we show the performance for the single-, dual- and full-polarization cases.

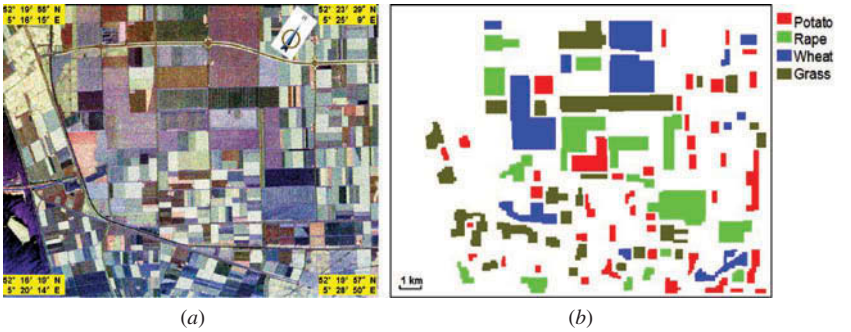
3.1. Data sets

3.1.1. AIRSAR data

Two PolSAR images acquired by the National Aeronautics and Space Administration Jet Propulsion Laboratory (NASA/JPL) AIRSAR L-band system were downloaded from http://earth.eo.esa.int/polsarpro/index.html. One image covers a part in the region of Flevoland in the Netherlands and the other covers an area in San Francisco, California, in the USA.

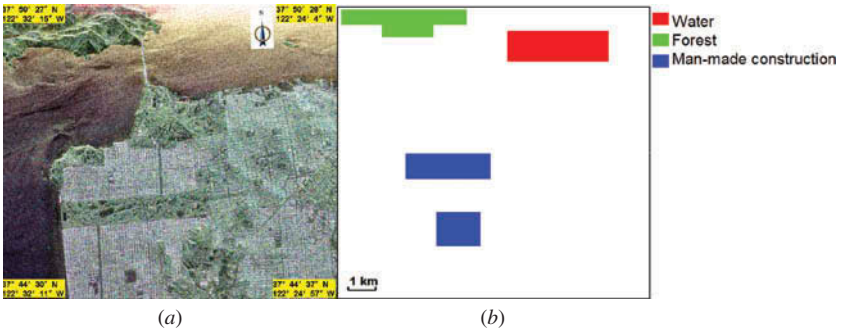
The image of Flevoland was used for crop classification study. As shown in Figure 1, this image has a size of 1024 × 750 pixels. The pixel size is 6.6 m in the slant range direction and 12.1 m in the azimuth direction. The incidence angles are 20°at near range and 44° at far range. The ground of this image is located at 52.31°N, 5.33°E . Four land-use classes, such as potato, wheat, rape and grass, were extracted, and the samples' amount of each class in our experiment is 500.

The image of San Francisco has a size of 1024 × 900 pixels, as shown in Figure 2. Its incident angles span from 10° to 60°. The ground of this image is located at 49.13°N, 122.45°W, in San Francisco, California, United States. Three land covers, such as vegetation, man-made construction and water, were used for training and testing, and the samples' amount of each class in our experiment is 1000. The two data sets were labelled manually by referring to Lee and Pottier (2009).



(a) (b)

Figure 1. (a) AIRSAR false colour image of Flevoland, based on Pauli decomposition, with $|S_{hh} + S_{vv}|$ as blue, $2|S_{hv}|$ as green, and $|S_{hh} - S_{vv}|$ as red. (b) Training and evaluation polygons.



(a) (b)

Figure 2. (a) AIRSAR false colour image of San Francisco, based on Pauli decomposition, with $|S_{hh} + S_{vv}|$ as blue, $2|S_{hv}|$ as green, and $|S_{hh} - S_{vv}|$ as red. (b) Training and evaluation polygons.

3.1.2. Radarsat-2 data

The third data set, collected by C-band Radarsat-2, is shown in Figure 3. The image has a size of 2020 × 13,299 pixels. The pixel size is 4.87 m in the line-spacing and 4.73 m in the pixel-spacing. The acquisition-start-time is 14:38, 15 April 2008. The beam-mode is Fine Quad-Pol 2. The incidence angles are 19° at near range and 22° at far range. The slant-range-near-edge is 842.87 km. The ground of this image is located at 49.13°N, 123.10°W, in Vancouver, British Columbia, Canada. Four land-cover classes, such as man-made construction, forest, water and low vegetation including bare soil, were selected for training and testing, and the samples' amount of each class in our experiment is 700. The labels were created manually by referring to Google Earth.

3.1.3. AVIRIS data

The fourth data set, acquired by an AVIRIS over the Kennedy Space Center, Orlando, Florida, United States, in March 1996, is shown in Figure 4. The image has the size of 614 × 512 pixels. The pixel size is 18 m. There are 224 spectral bands spanning from 400 nm to 2500 nm and 176 bands are used for the study by removing the bands absorbed

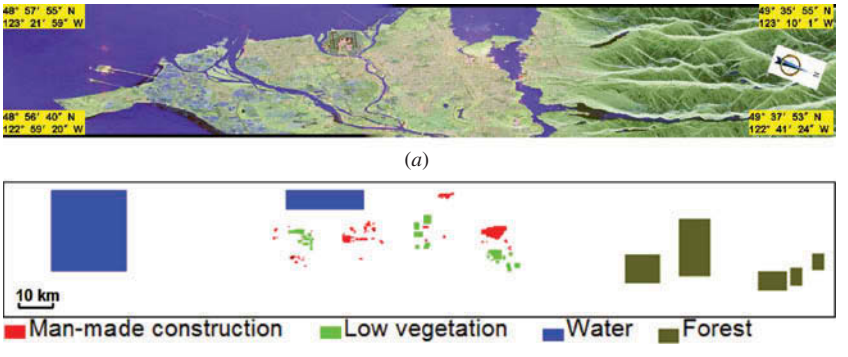


Figure 3. (*a*) Radarsat-2 false colour image of Vancouver, Canada, based on Pauli decomposition, with $|S_{hh} + S_{vv}|$ as blue, $2|S_{hv}|$ as green, and $|S_{hh} - S_{vv}|$ as red. (*b*) Training and evaluation polygons.

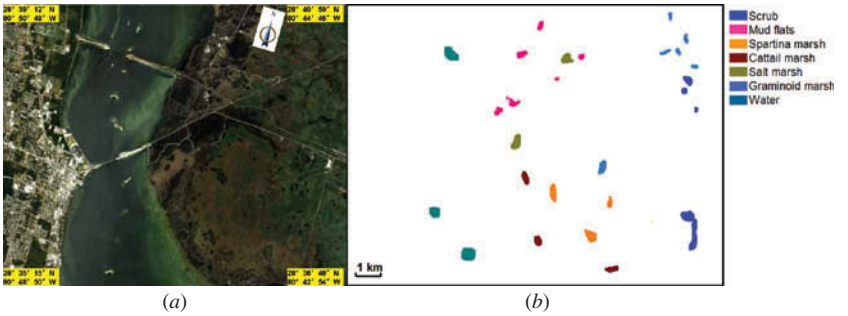


Figure 4. (*a*) AVIRIS false colour image of Kennedy Space Center, based on Pauli decomposition, with $|S_{hh} + S_{vv}|$ as blue, $2|S_{hv}|$ as green, and $|S_{hh} - S_{vv}|$ as red. (*b*) Training and evaluation polygons.

by atmospheric moisture and those with low signal-to-noise ratio. Seven land covers are chosen for illustration, including scrub (761 samples), graminoid marsh (503 samples), spartina marsh (520 samples), cattail marsh (404 samples), salt marsh (419 samples), mud flats (431 samples) and water (927 samples). The ground of this image is located at 28.61°N, 80.80°W, in Orlando, United States.

3.2. Results

The proposed method was applied to the four data sets. The classification accuracy is given by the mean producer's accuracy (MPA) that represents the mean of the producer's accuracy (expressed in per cent) of several classes. Each testing correct ratio is averaged over 20 trials, and the training data points are randomly extracted from the data sets.

We compared the proposed approach with two supervised classifications: classic SVM and k-NN, and two semi-supervised classifications: SGL and SDA.

SVM: Classic SVM is the most robust and accurate method among all the well-known data mining algorithms, and it has been developed rapidly both in theory and practice. It aims at finding a hyperplane to separate two classes of given samples with a maximal margin. Here, we use it in multiclassification with the library Libsvm written by Chih-Chung Chang and Chih-Jen Lin. The parameters are default.

SGL: SGL addresses the pair-wise relationships between any two vertices in the correlated Laplacian graph. We use the regulation, which is similar to the first two items of Equation (9). However, SGL has difficulty tuning the parameters of λ and μ. Here, we set λ to be the norm variance of the data points, and μ to be $\lambda/4$.

***k*-NN:** The aim of k-NN Classification is to locate a group of k points in the training set that are closest to the labelled point and base the assignment of a label on the labelled point in this neighbourhood. Here, k is set to be 5.

SDA: It aims to find a projection which respects the discriminant structure inferred from the training data. The training data, combined with unlabelled data, are used to build a Laplacian graph which provides a discrete approximation to the local geometry of the data manifold and can be incorporated into objective function. It can preserve the manifold structure.

Figure 5 shows the experiment results for the four data sets based on the proportion of the training data points to all data points, which spanned from 1% to 15%. Our method achieves better performance than that of both SVM and SGL for the four data sets. For SVM, the training data points are too fewer to present the structure of data set. SGL can't take into account high-order information. The MPA of three methods, HGL, k-NN and SDA, converges to stabilization when the training proportion closes to 5%. HGL achieves better performance than that of k-NN and SDA for the two AIRSAR data sets, and it is similar for Radarsat-2 data set. However, it is inferior to k-NN and SDA heavily for AVIRIS data set. And then, we performed the proposed method by tuning the parameters λ and μ from 0.1 to 10 in step 0.1. There is a little bit change with different λ and μ, and the optimal values are $\lambda = 1$ and $\mu = 1$.

Further, we performed the experiment with full-polarization, dual-polarization and single-polarization. The results, shown in Table 1, indicate that the MPA obtained for the three test data sets for training samples ranges from 5% to 10% for full-polarization, dual-polarization and single-polarization. The first column is the proportion of training data to all data, and the last row is the average MPA. Obviously, the performance of full-polarization exceeds that of dual-polarization, which, because of more information, is better than that of single-polarization. For single-polarization, the overall accuracy of VV-polarization is close to that of HH-polarization.

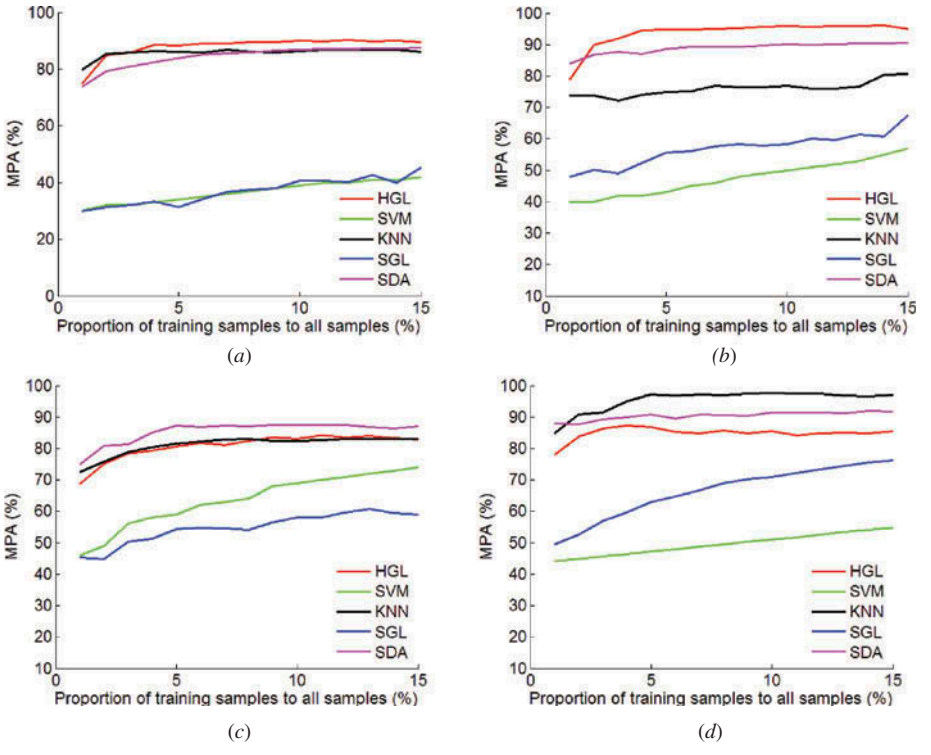


Figure 5. Comparison of HGL to classic SVM, SGL and *k*-NN on four data sets of (*a*) Flevoland, (*b*) San Francisco, (*c*) Vancouver, and (*d*) Kennedy Space Center.

4. Conclusions

This letter addressed the performance of a HGL algorithm for the classification of polarimetric SAR data. We have conducted experiments on four data sets, three PolSAR data sets and a hyperspectral data set. We compared the proposed HGL with classic SVM, SGL, *k*-NN and SDA. The proposed algorithm achieves a better overall accuracy than that of SVM, SGL, *k*-NN and SDA. Then, a hyperspectral data was conducted on the proposed method. The proposed method could not achieve a better result in this case. Finally, we conducted the experiments with single, dual-, and full-polarization SAR images. The results for full-polarization and dual-polarization are superior to those for single-polarization. The result for full-polarization is much better than that for dual-polarization. Based on our proposed method, we also conclude that the multi-polarization SAR system is superior to single polarization. However, using the proposed method to construct the hypergraph is time-consuming, especially when faced with large data sets. Therefore, in the future, comparable analyses must be performed.

Acknowledgement

The AIRSAR images were provided by the European Space Agency. The authors would like to thank Dr. Xingquan Wang for his constructive discussions and to thank the anonymous reviewers for their comments and suggestions.

Table 1. Mean producer's accuracy (MPA) of polarization based on HGL. First row lists the proportion of labelled points to overall samples.

	MPA of HGL (%)											
	Full PolSAR			Dual PolSAR			Single PolSAR					
	Pauli-decomposition			HH + VV			VV			HH		
Proportion (%)	San Francisco	Vancouver	Flevoland	San Francisco	Vancouver	Flevoland	San Francisco	Vancouver	Flevoland	San Francisco	Vancouver	Flevoland
5	95	82	88	79	75	84	62	62	53	52	58	51
6	95	82	89	82	74	85	63	65	59	58	46	56
7	95	83	89	81	75	86	64	68	58	61	48	58
8	95	83	90	82	75	86	64	69	62	63	48	58
9	96	82	90	83	74	85	65	69	62	66	65	61
10	96	83	90	82	74	86	64	70	62	66	65	60
Average	95.3	82.5	89.3	81.5	74.5	85.3	63.7	67.2	59.3	61	55	57.3

Funding

This work was supported in part by the National Natural Science Foundation of China [61100104], in part by the Program for New Century Excellent Talents in University [NCET-12-0323], and in part by the Hong Kong Scholar Programme [XJ2013038].

References

Agarwal, S., K. Branson, and S. Belongie. 2006. "Higher Order Learning with Graphs." In *Proceedings International Conference Machine Learning*, 17–24. New York: ACM.

Beaulieu, J. M., and R. Touzi. 2004. "Segmentation of Textured Polarimetric SAR Scenes by Likelihood Approximation." *IEEE Transactions on Geoscience and Remote Sensing* 42: 2063–2072. doi:10.1109/TGRS.2004.835302.

Cai, D., X. He, and J. Han. 2007. "Semi-Supervised Discriminant Analysis." IEEE 11th international conference on computer vision, Rio de Janeiro, October 14–21, 1–7.

De Grandi, G. D., J. S. Lee, and D. L. Schuler. 2007. "Target Detection and Texture Segmentation in Polarimetric SAR Images Using a Wavelet Frame: Theoretical Aspects." *IEEE Transactions on Geoscience and Remote Sensing* 45: 3437–3453. doi:10.1109/TGRS.2007.905103.

Duda, R. O., P. E. Hart, and D. G. Stork. 2000. *Pattern Classification*. 2nd ed. New York: Wiley.

Ersahin, K. I., G. Cumming, and R. K. Ward. 2010. "Segmentation and Classification of Polarimetric SAR Data Using Spectral Graph Partitioning." *IEEE Transactions on Geoscience and Remote Sensing* 48: 164–174. doi:10.1109/TGRS.2009.2024303.

Huang, Y., Q. Liu, S. Zhang, and D. Metaxas. 2010. "Image Retrieval Via Probabilistic Hypergraph Ranking." IEEE conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, June 13–18, 3376–3383.

Lardeux, C., P. L. Frison, C. Tison, J. C. Souyris, B. Stoll, B. Fruneau, and J. P. Rudant. 2009. "Support Vector Machine for Multifrequency SAR Polarimetric Data Classification." *IEEE Transactions on Geoscience and Remote Sensing* 47: 4143–4152. doi:10.1109/TGRS.2009.2023908.

Lee, J. S., M. R. Grunes, and G. De Grandi. 1999. "Polarimetric SAR Speckle Filtering and Its Implication for Classification." *IEEE Transactions on Geoscience and Remote Sensing* 37 (5): 2363–2373. doi:10.1109/36.789635.

Lee, J. S., and E. Pottier. 2009. *Polarimetric Radar Imaging: From Basics to Applications*. New York: CRC Press.

Wu, M., and B. Scholkopf. 2007. "Transductive Classification Via Local Learning Regularisation." International conference on artificial intelligence and statistics, Puerto Rico, March 21–24, 628–635.

Yu, J., D. Liu, D. Tao, and H.-S. Seah. 2011. "Complex Object Correspondence Construction in Two-Dimensional Animation." *IEEE Transactions on Image Processing* 20: 3257–3269. doi:10.1109/TIP.2011.2158225.

Yu, J., Y. Rui, and B. Chen. 2014. "Exploiting Click Constraints and Multiview Features for Image Reranking." *IEEE Transactions on Multimedia* 16: 159–168. doi:10.1109/TMM.2013.2284755.

Yu, J., Y. Rui, Y. Tang, and D. Tao. 2014. "High-Order Distance Based Multiview Stochastic Learning in Image Classification." *IEEE Transactions on Cybernetics* PP: 1–12. doi:10.1109/TCYB.2014.2307862.

Yu, J., Y. Rui, and D. Tao. 2014. "Click Prediction for Web Image Reranking Using Multimodal Sparse Coding." *IEEE Transactions on Image Processing* 23: 2019–2032. doi:10.1109/TIP.2014.2311377.

Yu, J., D. Tao, and M. Wang. 2012. "Adaptive Hypergraph Learning and Its Application in Image Classification." *IEEE Transactions on Image Processing* 21: 3262–3272. doi:10.1109/TIP.2012.2190083.

Zhou, D., J. Huang, and B. Scholkopf. 2006. "Learning with Hypergraphs: Clustering, Classification, and Embedding." *Advances in Neural Information Processing Systems* 19: 1601–1608.