

RESEARCH

Open Access

A novel information transferring approach for the classification of remote sensing images

Jianqiang Gao^{*}, Lizhong Xu, Jie Shen, Fengchen Huang and Feng Xu

Abstract

Traditional remote sensing images classification methods focused on using a large amount of labeled target data to train an efficient classification model. However, these approaches were generally based on the target data without considering a host of auxiliary data or the additional information of auxiliary data. If the valuable information from auxiliary data could be successfully transferred to the target data, the performance of the classification model would be improved. In addition, from the perspective of practical application, these valuable information from auxiliary data should be fully used. Therefore, in this paper, based on the transfer learning idea, we proposed a novel information transferring approach to improve the remote sensing images classification performance. The main rationale of this approach is that first, the information of the same areas associated with each pixel is modeled as the intra-class set, and the information of different areas associated with each pixel is modeled as the inter-class set, and then the obtained texture feature information of each area from auxiliary is transferred to the target data set such that the inter-class set is separated and intra-class set is gathered as far as possible. Experiments show that the proposed approach is effective and feasible.

Keywords: Transfer learning; Image classification; Texture feature information; Support vector machine (SVM)

1 Introduction

Remote sensing images classification is a complex process that may be affected by many factors, such as the availability of high-quality images, proper classification method, and the analytical ability of scientists. For a particular problem, it is often difficult to identify the best classifier due to the lack of a guideline for selection and the availability of suitable classification approaches to band. Therefore, many researchers proposed all kinds of algorithms to address the remote sensing images classification problems. In [1], the authors built textural information model that use spatial information, and then proposed a wavelet-based multi-scale strategy to characterize local texture, taking the physical nature of the data into account, then the extracted textural information was used as new feature to build a texture kernel and the final kernel was the weighted sum of a kernel made with the spectral information and the texture kernel. In [2], the authors proposed applying kernels on a segmentation graph method.

Fauvel et al. [3] proposed a spatial-spectral kernel-based approach with the spatial and spectral information were jointly used for the classification. A kernel-based block matrix decomposition approach for the classification of remotely sensed images was proposed by Gao et al. [4]. Tuia et al. [5] used active learning to adapt remote sensing image classifiers. Their goal is to select these pixels in an intelligent fashion that minimizes their number and maximizes their information content. Two strategies based on uncertainty and clustering of the data space are considered to perform active selection. In [6], Dos Santos J.A. et al. proposed a method for interactive classification of remote sensing images considering multiscale segmentation. Their aim is to improve the selection of training samples using the features from the most appropriate scales of representation. They use a boosting-based active learning strategy to select regions at various scales for user's relevance feed back. However, these approaches may ignore the auxiliary data of the remote sensing images. In other words, they do not take the auxiliary data into account in the classification model. In this paper, we aim to transfer the texture feature information from the auxiliary data to

*Correspondence: jianqianggaohh@126.com
College of Computer and Information Engineering, Hohai University, Xikang Road No.1, 210098 Nanjing, China

the target data to improve the classification performance of remote sensing images.

In the traditional classification learning framework, a classification task is to first train a classification model on a labeled training data. And then, the learned model is used to classify a test data set. Hence, under such a framework, the learning method relies on the availability of a large amount of labeled data. In practice, high-quality labeled data are often hard to come by, especially for learning tasks in a new region. Labeling data in a new region involves much human labor and is time-consuming, such as [5,6]. But, fortunately, some auxiliary data such as the texture information are easy to obtain. Therefore, it is reasonable to consider that how to make full use of the valuable texture information of some auxiliary data to improve the classification performance.

Recently, transfer learning [7] has become a popular machine learning method which utilizes auxiliary data for learning. Transfer learning is concerned with adapting knowledge acquired from one source domain to solving problems in another different but related target domain [8]. Generally speaking, traditional machine learning models assume that the training samples collected previously inherit the same feature and distribution as new, incoming data samples during operation [9]. However, in many real-world cases, this assumption does not always hold. In fact, in regard to data classification in non stationary environment, it is not unlikely that the training data set follows a different data distribution as compared with the actual incoming data samples during operation. Such as, in communication channels, discrete signals generated by a specific sequence from a source could be corrupted by Gaussian noise in the transmission process; so, the received signals could deviate from the signal sequence [10]. In this case, traditional machine learning models may not be able to perform well when dealing with the new data samples in the target domain. Hence, the ability of transfer learning would greatly improve the robustness of machine learning models by transferring and adapting knowledge learned from one domain to another related, but different domain. On the other hand, a large set of data samples from a particular task normally is required to train an effective machine learning model [11]. The main principle of transfer learning is that even though the data distributions in the source and target domains are different, some common knowledge across both domains can be adapted for learning [12].

Many researchers have proposed all kinds of methods to transfer learning information or knowledge from auxiliary data. In [13], authors proposed a TrAdaBoost transfer learning framework which constructed a high-quality classification model for target domain by a small number of labeled data and auxiliary data. In [14], authors proposed an extensional method called MultiSource-

TrAdaBoost to extend the TrAdaBoost framework for solving multiple sources. In [15], authors proposed a matrix factorization framework to build two mapping matrices for the training images and the auxiliary text data. Based on the co-occurrence data, the correlative principle was introduced to transfer knowledge from text to images by Qi et al. [16]. The authors of reference [17] use an auxiliary data set to construct the pseudo text for each target image, and then, by exploiting the semantic structure of the pseudo text data, the visual features are mapped to the semantic space which respects the text structure. Generally speaking, these methods attempted to transfer information from a lot of auxiliary data to train a more effective model for target data. In our paper, we employ the texture feature information of auxiliary data set to build the similarity matrix for target data set, and then by exploiting the texture information structure of the similarity matrix, the valuable features are mapped to the spectral space and the textural space. At last, the original

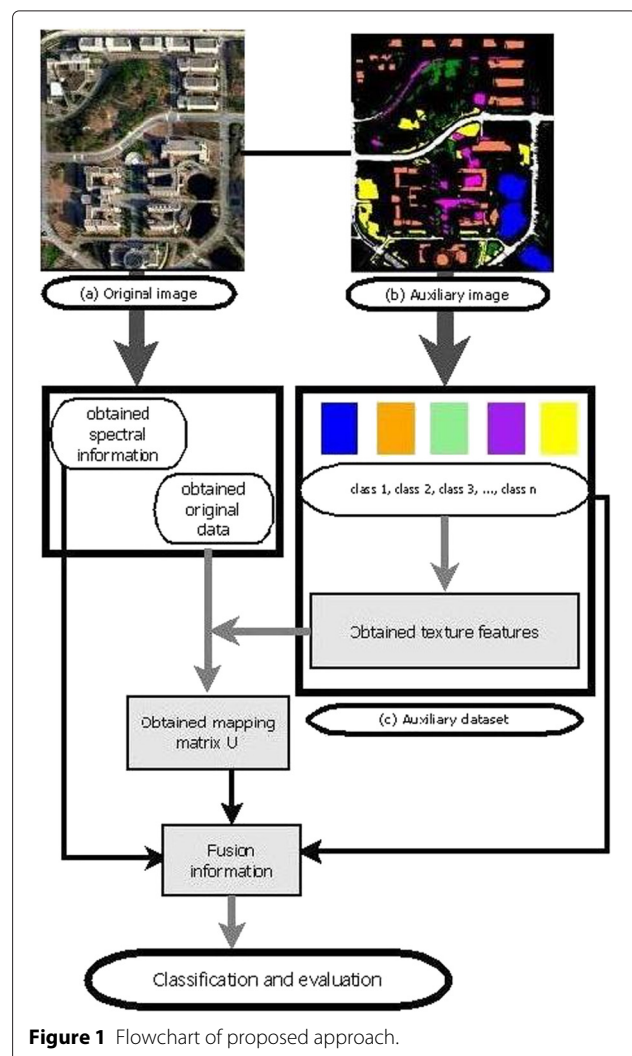


Figure 1 Flowchart of proposed approach.

Table 1 Information classes and training and test samples for PUD

Class No	Samples			
	Name	Train	Test	Auxiliary data
1	Asphalt	548	6304	300
2	Meadow	540	18146	300
3	Gravel	392	1815	300
4	Tree	524	2912	300
5	Metal sheet	265	1113	300
6	Bare soil	532	4572	300
7	Bitumen	375	981	300
8	Bricks	514	3364	300
9	Shadow	231	795	300
-	-	3921	40002	2700

spectral information is combined with texture information to improve the performance of classification model. In order to solve the shortcomings of scale sensitive and more time consuming, Zhang et al. [22] proposed a potential support vector machine (PSVM) algorithm, which uses a novel objective function to overcome the problem of scale sensitivity in SVM.

The remainder of this paper is organized as follows. Section 2 briefly reviews the formulations of relevant knowledge. In Section 3, the derivation process of the proposed method is described in detail. The effectiveness of the proposed method is demonstrated in Section 4 by experiments on remote sensing images. Finally, Section 5 concludes this paper.

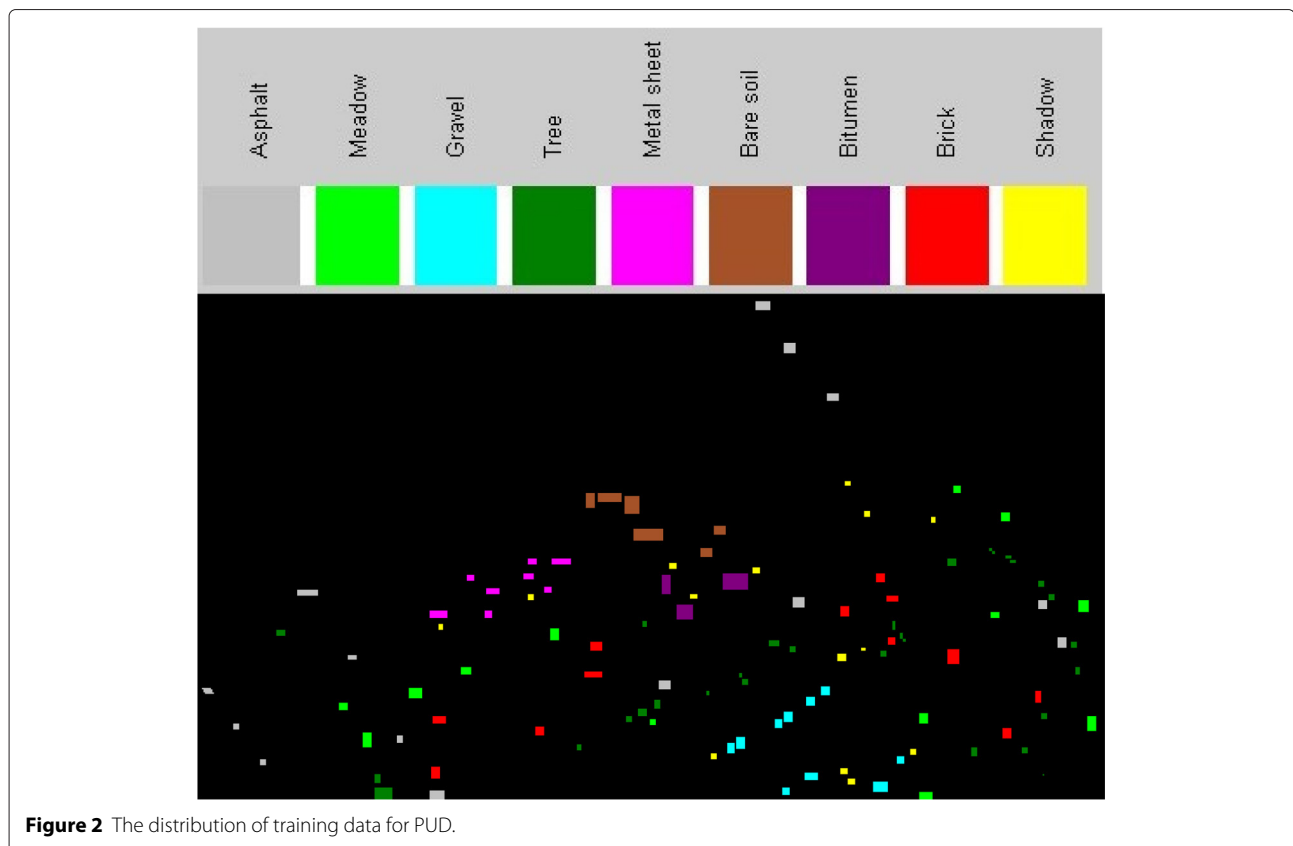


Figure 2 The distribution of training data for PUD.

2 Relevant knowledge

2.1 Transferring knowledge of feature representations

A new case of clustering problems, known as self-taught clustering, was proposed by Dai et al. [18]. Self-taught clustering (STC) is an instance of unsupervised transfer learning, which aims at clustering a small collection of unlabeled data in the target domain with the help of a large amount of unlabeled data in the source domain. STC tries to learn a common feature space across domains, which helps in clustering in the target domain. The objective function of STC is shown as follows:

$$J(\tilde{X}_T, \tilde{X}_S, \tilde{Z}) = I(X_T, Z) - I(\tilde{X}_T, \tilde{Z}) + \lambda [I(X_S, Z) - I(\tilde{X}_S, \tilde{Z})], \tag{1}$$

where X_S and X_T are the source and target domain data, respectively. Z is a shared feature space by X_S and X_T , and $I(\cdot, \cdot)$ is the mutual information between two random variables. Suppose that there exist three clustering functions $C_{X_T} : X_T \rightarrow \tilde{X}_T$, $C_{X_S} : X_S \rightarrow \tilde{X}_S$, and $C_Z : Z \rightarrow \tilde{Z}$, where \tilde{X}_T, \tilde{X}_S , and \tilde{Z} are corresponding clusters of X_T, X_S , and Z , respectively. The aim of STC is to learn \tilde{X}_T by solving the optimization problem (1):

$$\arg \min_{\tilde{X}_T, \tilde{X}_S, \tilde{Z}} J(\tilde{X}_T, \tilde{X}_S, \tilde{Z}). \tag{2}$$

An iterative algorithm for solving the optimization function (2) was given in [18].

2.2 Fisher linear discriminant analysis (FLDA)

The main goal of FLDA is to perform dimension reduction while preserving as much information as possible. Linear discriminant analysis aims to find the optimal transformation matrix such that the class structure of the original high-dimensional space is preserved in the low-dimensional space. But in hyperspectral remote sensing images classification problem, generally dimension of the feature vectors is very high with respect to the number of feature vectors. In this subsection, we briefly review the two-dimension Fisher discriminant analysis (2DFLDA) method by Kong et al. [19] proposed to handle the reduce dimensional problem. The main content can be summarized as follows:

Let c be the number of classes, N_i be the number of selected samples from i th class, N be the number of total selected samples from each class, A_j^i be the j th image from i th class, and m_i be the mean image of i th class. $N = \sum_{i=1}^c N_i$, $m_i = \frac{1}{N} \sum_{j=1}^{N_i} A_j^i$, ($i = 1, \dots, c$). The optimal projection matrix $G = [g_1, g_2, \dots, g_l]$ can be found in 2DFLDA. Where l is at most $\min(c - 1, N)$. We can obtain the optimal projection matrix by maximizing the following criterion:

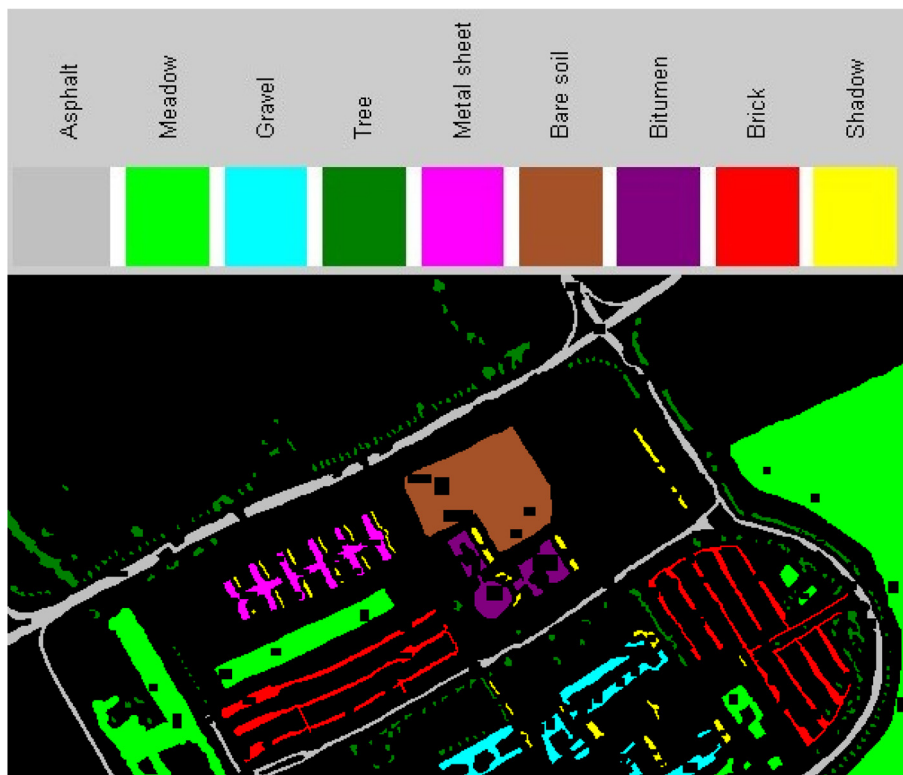


Figure 3 The distribution of test data for PUD.

$$J(G) = \frac{G^T S_b G}{G^T S_w G}, \tag{3}$$

where S_b and S_w are the inter-class and intra-class scatter matrices, respectively. $S_b = \sum_{i=1}^c (m_i - m_0)^T (m_i - m_0)$, $S_w = \sum_{i=1}^c \sum_{j=1}^{N_i} (A_j^i - m_i)^T (A_j^i - m_i)$. $m_0 = \frac{1}{c} \sum_{i=1}^c m_i$ is the global mean image of all classes.

3 Learning for information transferring

In this section, based on gray level co-occurrence matrix (GLCM), we first obtain the texture feature information of an image. In addition, the feature matrix of auxiliary data for remote sensing images can be obtained, as described in the following. According to Equation 3, compute matrices

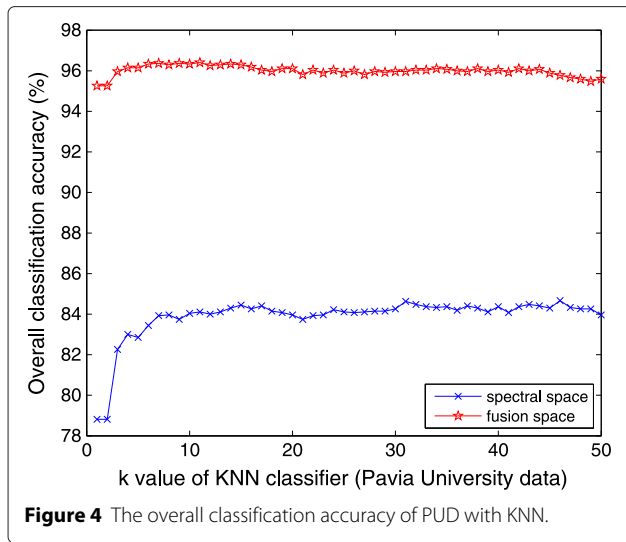
S_b and S_w , and solve the optimal projection matrix G . Let λ_i , ($i = 1, 2, \dots, l$) be the absolute values of the diagonal elements of the G corresponds to matrix. The value of k is determined such that E is at least some fixed percentage of the whole energy of the image. In our following experiments, we choose $E = 99.99\%$:

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^l \lambda_i} \geq E, \tag{4}$$

Figure 1 shows a block diagram of our simple system. In the next, we will introduce the proposed approach which can be summarized as follows.

Table 2 Confusion matrices, κ and time (s) of PUD

Class no.	1	2	3	4	5	6	7	8	9	UA (%)
Spectral space										
1	5,244	37	130	22	19	16	368	460	8	83.19
2	0	12,230	0	2,223	0	3,675	0	18	0	67.40
3	29	8	1,194	0	0	3	1	580	0	65.79
4	0	36	0	2,858	1	17	0	0	0	98.15
5	0	1	2	2	1,105	0	0	0	3	99.28
6	5	118	0	40	99	4,216	0	24	0	92.21
7	96	0	1	0	0	0	872	12	0	88.89
8	31	19	185	3	0	31	8	3,087	0	91.77
9	21	0	7	0	0	0	0	0	767	96.48
PA(%)	96.65	97.69	78.60	55.52	90.28	52.98	69.82	73.83	98.59	
OA(%) = 78.93										
$\kappa = 0.7340$										
t=1.26s										
Fusion space										
1	5,286	5	4	41	3	1	403	547	14	83.85
2	0	18,145	0	1	0	0	0	0	0	99.99
3	0	0	1,782	0	0	33	0	0	0	98.18
4	0	0	0	2,911	1	0	0	0	0	99.97
5	0	0	0	4	1,109	0	0	0	0	99.64
6	0	0	11	26	33	4,502	0	0	0	98.47
7	95	0	0	0	0	0	872	14	0	88.89
8	32	0	0	0	0	0	8	3,324	0	98.81
9	21	0	0	0	0	0	0	3	771	96.98
PA(%)	97.28	99.97	99.17	97.59	96.77	99.25	67.97	85.49	98.22	
OA(%) = 96.75										
$\kappa = 0.9562$										
t=0.51s										



3.1 Notations

In this paper, we consider two data sets. One is the target data set (*viz.* original image) which only includes spectral information. The other is the auxiliary data set which consists of texture information (Please consult Figure 1). Both the two data sets include c classes. Let \mathbb{R}^k and \mathbb{R}^m be the spectral information and texture information feature spaces. And without loss of generality, we use $S^{(t)}$ and $S^{(a)}$ to represent target data set and auxiliary data set, respectively. Denote the feature matrix of target data set as $\mathbf{X}^{(t)} \in \mathbb{R}^{k \times n^{(t)}}$, the feature matrix of spectral information of auxiliary data set as $\mathbf{X}^{(a)} \in \mathbb{R}^{k \times n^{(a)}}$, and the texture feature information matrix in auxiliary data set as $\mathbf{T}^{(a)} \in \mathbb{R}^{m \times n^{(a)}}$. For target data set, we assume that each sample corresponds to particular auxiliary information. We use $S^{(t)}$ to represent the target data as below Equation 5:

$$S^{(t)} = \left\{ \left(\mathbf{x}_i^{(t)}, \widehat{\mathbf{f}}_i^{(t)}, y_i^{(t)} \right) \mid 1 \leq i \leq n^{(t)} \right\}, \quad (5)$$

where $\mathbf{x}_i^{(t)} \in \mathbb{R}^k$ is the column vector of $\mathbf{X}^{(t)}$, $\widehat{\mathbf{f}}_i^{(t)} \in \mathbb{R}^m$ is the feature vector of the pseudo texture feature information of target data, and $y_i^{(t)} \in \{1, 2, \dots, c\}$ is the class label of target data. Similarly, we use $S^{(a)}$ to represent the auxiliary data as below Equation 6:

$$S^{(a)} = \left\{ \left(\mathbf{x}_j^{(a)}, \mathbf{f}_j^{(a)}, y_j^{(a)} \right) \mid 1 \leq j \leq n^{(a)} \right\}, \quad (6)$$

where $\mathbf{x}_j^{(a)} \in \mathbb{R}^k$ is the column vector of $\mathbf{X}^{(a)}$, $\mathbf{f}_j^{(a)} \in \mathbb{R}^m$ is the feature vector of the texture feature information of target data, and $y_j^{(a)} \in \{1, 2, \dots, c\}$ is the class label of auxiliary data. In addition, we use C_w and C_b to represent the relationship of $\mathbf{x}_i^{(t)}$ and $\mathbf{x}_j^{(t)}$ as follows in Equations 7 and 8, respectively:

$$C^{(w)} = \left\{ \left(\mathbf{x}_i^{(t)}, \mathbf{x}_j^{(t)} \right) \mid y_i^{(t)} = y_j^{(t)} \right\}, \quad (7)$$

$$C^{(b)} = \left\{ \left(\mathbf{x}_i^{(t)}, \mathbf{x}_j^{(t)} \right) \mid y_i^{(t)} \neq y_j^{(t)} \right\}. \quad (8)$$

Equation 7 shows that $\mathbf{x}_i^{(t)}$ and $\mathbf{x}_j^{(t)}$ are in the same class in target data set. And then, Equation 8 shows that $\mathbf{x}_i^{(t)}$ and $\mathbf{x}_j^{(t)}$ are in the different classes in target data set.

3.2 Construct the similarity matrix of $S^{(t)}$ and $S^{(a)}$

As we all know, there are same spectrum and texture information for the same region (or field). Therefore, the similarity matrix with very important information for target data set is constructed based on the similarities between samples in $S^{(t)}$ and $S^{(a)}$. For the sample $\mathbf{x}_i^{(t)}$ in $S^{(t)}$, the most similar sample in $S^{(a)}$ is defined as Equation 9:

$$f(\mathbf{x}_i^{(a)}) = \min_{\mathbf{x}_j^{(a)}} d(\mathbf{x}_i^{(t)}, \mathbf{x}_j^{(a)}), \forall j, y_i^{(t)} = y_j^{(a)}, (j = 1, 2, \dots, n^{(a)}), \quad (9)$$

where $d(\cdot, \cdot)$ is the Euclidean distance in \mathbb{R}^k . The Equation 9 shows that the auxiliary data set corresponding to $\mathbf{x}_j^{(a)}$ can approximately reflect the similarity relationship of

Table 3 Information classes and training and test samples for HUD

Class No	Samples			
	Name	Train	Test	Auxiliary data
1	Road	488	64,650	100
2	Roof	242	56,560	100
3	Tree	266	24,765	100
4	Bare soil	230	36,650	100
5	Water	182	36,600	100
6	Shadow	304	21,220	100
-	-	8,220	240,445	600

$\mathbf{x}_i^{(t)}$. So, we can obtain the $\hat{\mathbf{f}}_i^{(t)} = \mathbf{x}_j^{(a)}$. In the following steps, we will obtain the similarity matrix of intra-class \mathbf{W}_w and the similarity matrix of inter-class \mathbf{W}_b by computing Equations 10 and 11:

$$\mathbf{W}_w = \begin{cases} w_{ij}^{(w)} = d(\hat{\mathbf{f}}_i^{(t)}, \mathbf{f}_j^{(a)}), & y_i^{(t)} = y_j^{(a)} \\ 0, & \text{otherwise} \end{cases}, \quad (10)$$

$$\mathbf{W}_b = \begin{cases} w_{ij}^{(b)} = d(\hat{\mathbf{f}}_i^{(t)}, \mathbf{f}_j^{(a)}), & y_i^{(t)} \neq y_j^{(a)} \\ 0, & \text{otherwise} \end{cases}. \quad (11)$$

In Equations 10 and 11, $w_{ij}^{(w)}$ and $w_{ij}^{(b)}$ are the elements of \mathbf{W}_w and \mathbf{W}_b , respectively. $d(\cdot, \cdot)$ is the Euclidean distance

between two feature vectors with very important texture information. For \mathbf{W}_w and \mathbf{W}_b , in order to simplify the calculation, we have done the approximate calculation. The specific steps are as follows:

Firstly, we build feature matrices of similarity matrix $S^{(t)}$ by using $\hat{\mathbf{f}}_i^{(t)}$ ($i = 1, 2, \dots, n^{(t)}$) and auxiliary data set matrix $S^{(a)}$ by using $\mathbf{f}_j^{(a)}$ ($j = 1, 2, \dots, n^{(a)}$), respectively. Viz:

$$S_{\mathbf{f}}^{(t)} = [\hat{\mathbf{f}}_1^{(t)}, \hat{\mathbf{f}}_2^{(t)}, \dots, \hat{\mathbf{f}}_{n^{(t)}}^{(t)}], \quad (12)$$

$$S_{\mathbf{f}}^{(a)} = [\mathbf{f}_1^{(a)}, \mathbf{f}_2^{(a)}, \dots, \mathbf{f}_{n^{(a)}}^{(a)}]. \quad (13)$$



Figure 5 The airborne remote sensing digital image of HUD.

Secondly, we build the similarity matrix of intra-class \mathbf{W}_w and the similarity matrix of inter-class \mathbf{W}_b by using the feature vector of each sample of $S_{\mathbf{f}_i}^{(t)}$ and $S_{\mathbf{f}_j}^{(a)}$. At last, the \mathbf{W}_w and \mathbf{W}_b as Equations 14 and 15.

$$\mathbf{W}_w = \sum_{i=1}^{n^{(t)}} \sum_{j=1}^{n^{(a)}} (S_{\mathbf{f}_i}^{(t)} - S_{\mathbf{f}_j}^{(a)})^T (S_{\mathbf{f}_i}^{(t)} - S_{\mathbf{f}_j}^{(a)}), \quad (14)$$

$$\mathbf{W}_b = \sum_{i=1}^{n^{(t)}} \sum_{j=1}^{n^{(a)}} (\bar{S}_{\mathbf{f}_i}^{(t)} - \bar{S}_{\mathbf{f}_j}^{(a)})^T (\bar{S}_{\mathbf{f}_i}^{(t)} - \bar{S}_{\mathbf{f}_j}^{(a)}), \quad (15)$$

where $\bar{S}_{\mathbf{f}_i}^{(t)}$ is the i th row mean value of $S_{\mathbf{f}_i}^{(t)}$ and $\bar{S}_{\mathbf{f}_j}^{(a)}$ is the j th row mean value of $S_{\mathbf{f}_j}^{(a)}$.

3.3 Information transferring of auxiliary data

In this paper, our goal is to learn an optimal linear mapping matrix $\mathbf{U} \in \mathbb{R}^{k \times m}$ which project the texture information from auxiliary data set to the target data set. That is because the texture information of an image is very

important; meanwhile, it can enhance the image detail by introducing the texture information of auxiliary data. We formulate the regularization framework for information transferring of auxiliary data as follows:

$$\min_{\mathbf{U}} F(\mathbf{U}) = \left\| \mathbf{U}^T \mathbf{X}^{(a)} - \mathbf{T}^{(a)} \right\|_F^2 + \Omega(\mathbf{U}), \quad (16)$$

where $\|\cdot\|_F^2$ is the Frobenius norm, and $\Omega(\cdot)$ is the regularization constraint on $S^{(t)}$. In this framework, we project the texture feature information in $S^{(a)}$ from the auxiliary data set space to the target data set space. Meanwhile, the constraint on $S^{(t)}$ is taken into account. In this paper, we define $\Omega(\cdot)$ as follows:

$$\Omega(\mathbf{U}) = \alpha \Psi_w(\mathbf{U}) - (1 - \alpha) \Psi_b(\mathbf{U}), \quad (17)$$

where Ψ_w is the similarity constraints on C_w , Ψ_b is the diversity constraints on C_b , $\alpha (0 < \alpha < 1)$ is regularization parameter for balancing the tradeoff between

Table 4 Confusion matrices, κ and time (s) of HUD

Class no.	1	2	3	4	5	6	UA (%)
Spectral space							
1	44,609	19,748	0	188	0	105	69.00
2	16,283	37,422	250	777	4	1824	66.16
3	184	1,160	22,429	11	662	319	90.57
4	189	4,576	52	31,832	0	1	86.85
5	33	4	47	0	36,484	32	99.68
6	176	3041	1,165	2	4,358	12,478	58.80
PA(%)	72.57	56.74	93.68	97.02	87.90	84.55	
OA(%) = 77.05							
$\kappa = 0.7145$							
t=18s							
Fusion space							
1	64,634	16	0	0	0	0	99.98
2	0	56,560	0	0	0	0	100.0
3	0	638	22,843	0	662	622	92.24
4	0	1,250	0	35,400	0	0	96.59
5	0	33	46	0	36,485	36	99.69
6	0	267	1,457	0	4,366	15,130	71.30
PA(%)	100.0	96.25	93.83	100.0	87.89	95.83	
OA(%) = 96.09							
$\kappa = 0.9515$							
t=8s							

within-class and between-class constraints. Specifically, Ψ_w is formulated as follows:

$$\begin{aligned}\Psi_w(\mathbf{U}) &= \sum_{(\mathbf{x}_i^t, \mathbf{x}_j^t) \in C_w} w_{ij}^{(w)} \left\| \mathbf{U}^T \mathbf{x}_i^{(t)} - \mathbf{U}^T \mathbf{x}_j^{(t)} \right\|_F^2 \\ &= \text{tr} \left(\mathbf{U}^T \mathbf{X}^{(t)} \mathbf{P}_w (\mathbf{X}^{(t)})^T \mathbf{U} \right)\end{aligned}\quad (18)$$

where $\mathbf{P}_w = \mathbf{I} - \mathbf{D}_w^{-\frac{1}{2}} \mathbf{W}_w \mathbf{D}_w^{-\frac{1}{2}}$ the normalized Laplacian matrix, \mathbf{I} is a unit matrix, and $\mathbf{D}_w = \text{diag}(\mathbf{W}_w \cdot \mathbf{1})$ is a weight matrix whose diagonal elements are $\mathbf{D}_w^{ii} = \sum_{j=1}^{n^{(t)}} w_{ij}^{(w)}$, and $\text{tr}(\cdot)$ denotes the trace function. Similarly, Ψ_b can be formulated as Equation 19

$$\begin{aligned}\Psi_b(\mathbf{U}) &= \sum_{(\mathbf{x}_i^t, \mathbf{x}_j^t) \in C_b} w_{ij}^{(b)} \left\| \mathbf{U}^T \mathbf{x}_i^{(t)} - \mathbf{U}^T \mathbf{x}_j^{(t)} \right\|_F^2 \\ &= \text{tr} \left(\mathbf{U}^T \mathbf{X}^{(t)} \mathbf{P}_b (\mathbf{X}^{(t)})^T \mathbf{U} \right)\end{aligned}\quad (19)$$

where $\mathbf{P}_b = \mathbf{I} - \mathbf{D}_b^{-\frac{1}{2}} \mathbf{W}_b \mathbf{D}_b^{-\frac{1}{2}}$ the normalized Laplacian matrix, $\mathbf{D}_b = \text{diag}(\mathbf{W}_b \cdot \mathbf{1})$ is a weight matrix whose diagonal elements are $\mathbf{D}_b^{ii} = \sum_{j=1}^{n^{(t)}} w_{ij}^{(b)}$.

Through the above analysis, the objective function in Equation 16 can be rewritten as follows:

$$\begin{aligned}\min_{\mathbf{U}} F(\mathbf{U}) &= \left\| \mathbf{U}^T \mathbf{X}^{(a)} - \mathbf{T}^{(a)} \right\|_F^2 + \Omega(\mathbf{U}) \\ &= \left\| \mathbf{U}^T \mathbf{X}^{(a)} - \mathbf{T}^{(a)} \right\|_F^2 + \alpha \Psi_w(\mathbf{U}) - (1 - \alpha) \Psi_b(\mathbf{U}) \\ &= \left\| \mathbf{U}^T \mathbf{X}^{(a)} - \mathbf{T}^{(a)} \right\|_F^2 + \alpha \text{tr} \left(\mathbf{U}^T \mathbf{X}^{(t)} \mathbf{P}_w (\mathbf{X}^{(t)})^T \mathbf{U} \right) \\ &\quad - (1 - \alpha) \text{tr} \left(\mathbf{U}^T \mathbf{X}^{(t)} \mathbf{P}_b (\mathbf{X}^{(t)})^T \mathbf{U} \right).\end{aligned}\quad (20)$$

It is obvious that the above optimization is a convex problem, which can be achieved using existing convex optimization packages, such as **fminunc** and **fmincon** functions [20], **SeDuMi** [21]. The detailed description of the overall pseudo algorithm process is given in Algorithm 1.

Algorithm 1 Proposed approach

Input: Datasets $S^{(t)}$, $S^{(a)}$, regularization parameter α .

First stage:

Step 1. The value of k is determined by using Equations 3 and 4;

Step 2. Construct three matrices $\mathbf{X}^{(t)}$, $\mathbf{X}^{(a)}$ and $\mathbf{T}^{(a)}$, meanwhile, calculate matrices \mathbf{P}_w and \mathbf{P}_b ;

Step 3. Set initialize $\mathbf{U}_0 = \mathbf{0}$;

Step 4. According to the gradient descent method, solve the \mathbf{U} by using optimization packages;

Step 5. Until convergence or maximum iteration number achieves;

Output: The mapping matrix \mathbf{U} .

Second stage:

Step 1. For training and test datasets, the valuable texture features are mapped to the spectral space, then combined original spectral space. Viz. $[\mathbf{U} * \mathbf{T}^{(a)}; \mathbf{X}^{(a)}]$ and $[\mathbf{U} * \mathbf{T}^{(s)}; \mathbf{X}^{(t)}]$, where, $\mathbf{T}^{(s)}$ is pseudo texture matrix of test dataset.

Step 2. The effective classification model can be obtained by using SVM or K-Nearest Neighbor (KNN) classifier to train.

4 Experimental results and analysis

In this section, we demonstrate the effectiveness of the proposed approach on remote sensing images classification tasks. The available data set, namely Pavia University data set and Hohai University data set, are used for experiments. In order to evaluate the efficiency of proposed method, the Gaussian radial basis kernel function is employed in our experiment as Equation 21. And then, the penalty term C and the width of kernel g are need to be tuned. In addition, the two parameters were set using five-fold cross validation strategy. Each original data set was scaled between $[-1, 1]$ by using a per band range stretching method.

$$k_{\sigma}(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) = \exp \left(-g \cdot \|x_i - x_j\|^2 \right), \quad (21)$$

4.1 Pavia University data set (PUD)

Pavia dataset is around the Engineering School at the University of Pavia. It is 610×340 pixels. The spatial resolution is 1.3 m per pixel. Twelve channels have been removed due to noise. The remaining 103 spectral channels are processed. Nine classes of interest are considered: asphalt, meadow, gravel, tree, metal sheet, bare soil, bitumen, bricks, and shadow. The training and test sets for each class are given in Table 1.

In our experiments, the product's accuracy (PA) and the user's accuracy (UA) are defined as Equations 22 and 23, respectively:

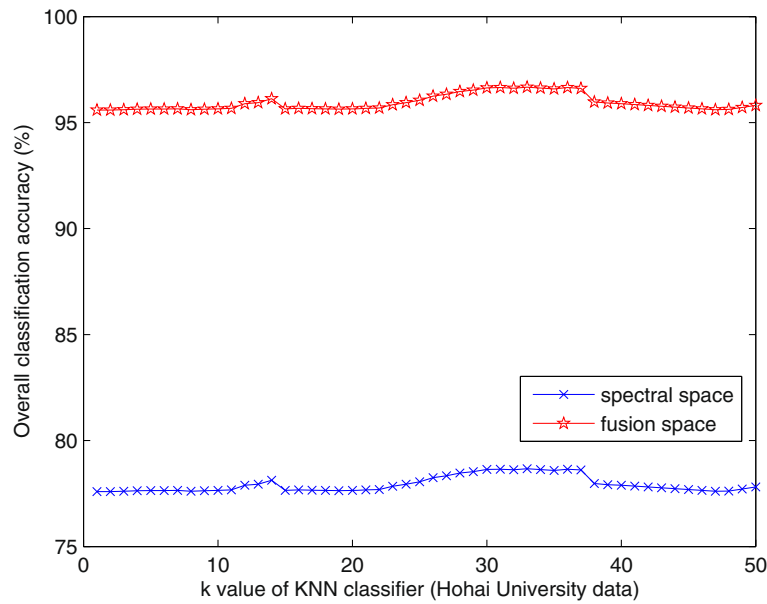


Figure 6 The overall classification accuracy of HUD with KNN.

$$PA_i = \frac{x_{i,i}}{x_{+i}}, \tag{22}$$

$$UA_i = \frac{x_{i,i}}{x_{i+}}, \tag{23}$$

where $x_{i,i}$ is the value on the major diagonal of the i th row in the confusion matrix, x_{i+} is the total number of the i th row, and x_{+i} is the total number of the i th column. To measure the agreement between the classification and the reference data, we compute the kappa coefficient (κ) based on the following equation, where N is the number of total pixels.

$$\kappa = \frac{\left[N \sum_{i=1}^k x_{i,i} - \sum_{i=1}^k (x_{i+} \times x_{+i}) \right]}{\left[N^2 - \sum_{i=1}^k (x_{i+} \times x_{+i}) \right]}. \tag{24}$$

The distribution of training data and test data are listed in Figures 2 and 3, respectively. All the algorithms are

tested in MATLAB (2010b) running on a PC with Intel Core 2 Celeron (2.40 GHz) with 2 GB of RAM. The two parameters C and g (from 2^{-10} to 2^{10} , the step is $2^{0.5}$) are determined by fivefold cross-validation strategy. According to the experiments, we can see that $C = 64, g = 8$ is the best choice in spectral space, while $C = 16, g = 16$ is the best choice in fusion space. In addition, we also found that the trained model is more efficient in fusion space than only in spectral space by using SVM. The confusion matrices of PUD were shown in Table 2.

According to Table 2, we found that the proposed approach with the fusion space gives better results as compared to the spectral space applied on PUD. In addition, the proposed method gives more overall accuracy (OA) (96.63%) and kappa value (0.9621) as compared to the method original spectral space. So, the proposed method can improve the overall classification accuracy and kappa value. In addition, it is worth noting that the elapsed time of proposed method is less than the original method.

Table 5 OA (%), κ and time (s) for SVM, PSVM, Mbsvd, Mbqrcp and proposed method with SVM

Method	PUD			HUD		
	OA	κ	Time	OA	κ	Time
SVM [22]	78.93	0.7340	1.26	77.05	0.7145	18
PSVM [22]	95.36	0.9466	0.91	91.88	0.9345	11
Mbsvd [4]	95.91	0.9489	1.01	92.03	0.9366	28
Mbqrcp [4]	96.60	0.9548	0.80	94.74	0.9411	10
Proposed	96.75	0.9562	0.51	96.09	0.9515	8

Figure 4 will show the overall classification accuracy by using KNN classifier. According to Figure 4, we can see that the proposed method gives the best results.

4.2 Hohai University data set

The data is the airborne remote sensing digital orthophoto map images acquired in February 2012, at the location of Jiangning campus of Hohai University, Nanjing city, Jiangsu province, P.R. China. This data set is at a spatial resolution of 0.5 m, and the size of image is $1,400 \times 1,024$ pixels. In this data set, we only considered six classes such as road, roof, tree, bare soil, water, and shadow to characterize this area. The class definitions and the number of samples for each experiment is listed in Table 3.

The airborne remote sensing digital image of Hohai University data set (HUD) is shown in Figure 5. According to the experiments, we can see that $C = 2^{0.5} \approx 1.4142$, and $g = 2^{7.5} \approx 181.0193$ is the best choice in fusion space for HUD.

Table 4 shows the confusion matrices, κ values, and elapsed time obtained for different space.

According to Table 4, the OA in classification accuracy obtained by proposed approach (96.09%) was much higher than that obtained by original method (77.05%). In addition, it is worth noting that the elapsed time of proposed algorithm is less than original algorithms. Meanwhile, we also found that in the fusion space, the result is better as compared to in the spectral space in terms of the accuracy of each class classification and κ .

Figure 6 will display the overall classification accuracy by using KNN classifier.

From Figure 6, we can obtain that the proposed approach gives a better result with respect to the OA by using KNN classifier. In order to demonstrate the effectiveness of the proposed approach on remote sensing image classification task, the comparison with other techniques proposed in the literature is implemented in the following experiment. Table 5 gives the overall accuracy and κ value of different data sets. The best results are reported in Table 5 according to different approaches.

From Table 5, we found that the proposed method shows better performance as compared to other approaches in terms of OA, κ , and running time. This is because the valuable texture information is employed in classification process. Hence, in the classification phase, the classification performance is improved.

5 Conclusions

In this paper, we proposed an information transferring approach to enhance remote sensing images classification performance. The main idea of the proposed method is that the texture feature information of auxiliary data set is transferred to the target data set, and then, the

classification model is trained by using SVM or KNN classifier. And finally, experimental results show our approach is feasible.

In addition, the authors realize that more work must be done to improve the classification results in the further. Such as, how to choose a suitable method in classification tasks for remote sensing images. In addition, how to avoid negative transfer is an important open issue that is attracting more and more attention in the future. Of course, in this paper, how to determine the parameter λ , how to transfer other valuable information. This will be an interesting open issue.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work is supported partly by the National Natural Science Foundation of PR China (No. 61271386) and by the Graduates' Research Innovation Program of Higher Education of Jiangsu Province of PR China (No. CXZZ13-0239), and the Industrialization Project of Universities in Jiangsu Province PR China (No. JH10-9).

Received: 30 April 2014 Accepted: 2 April 2015

Published online: 24 April 2015

References

1. G Mercier, F Girard-Ardhuin, Partially supervised oil-slick detection by SAR imagery using kernel expansion. *IEEE Trans. Geosci. Remote Sensing*. **44**(10), 2839–2846 (2006)
2. Z Harchaoui, F Bach, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Image classification with segmentation graph kernels, (2007), pp. 1–8
3. M Fauvel, J Chanussot, JA Benediktsson, A spatial-spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognit.* **45**(1), 381–392 (2012)
4. J Gao, L Xu, A Shi, F Huang, A kernel-based block matrix decomposition approach for the classification of remotely sensed images. *Appl. Math. Comput.* **228**, 531–545 (2014)
5. D Tuia, E Pasollib, WJ Emeryc, Using active learning to adapt remote sensing image classifiers. *Remote Sensing Environ.* **115**(9), 2232–2242 (2011)
6. JA Dos Santos, PH Gosselin, PF Sylvie, RDS Torres, AX Falcao, Interactive multiscale classification of high-resolution remote sensing images. *Selected Topics Appl. Earth Observations Remote Sensing, IEEE J.* **99**, 1–15 (2013)
7. SJ Pan, Q Yang, A survey on transfer learning. *Knowledge and Data Engineering. IEEE Trans.* **22**(10), 1345–1359 (2010)
8. G Boutsioukiss, I Partalas, I Vlahavas, in *Recent advances in reinforcement learning, 9th European Workshop EWRL*. Transfer learning in multi-agent reinforcement learning domains (Athens, Greece, 2011)
9. B Kocor, A Arslan, Genetic transfer learning. *Expert Syst. Appl.* **37**, 6997–7002 (2010)
10. DI Ostry, Synthesis of accurate fractional Gaussian noise by filtering. *IEEE Trans. Inf. Theory.* **52**(4), 1609–1623 (2006)
11. Z Xu, S Sun, in *Neural information processing, 19th International Conference ICONIP*. Multi-source transfer learning with multiview Adaboost (Doha, Qatar, 2012)
12. S Yang, M Lin, C Hou, C Zhang, Y Wu, A general framework for transfer sparse subspace learning. *Neural Comput. Appl.* **21**(7), 1801–1817 (2012)
13. Y Yao, G Doretto, in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE*. Boosting for transfer learning with multiple sources, (2010), pp. 1855–1862
14. W Dai, Y Chen, G Xue, Q Yang, Y Yu, in *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. Translated learning: transfer learning across different feature spaces, (2008), pp. 353–360
15. Y Zhu, Y Chen, Z Lu, et al, in *Special Track on AI and the Web, associated with The Twenty-Fourth AAI Conference on Artificial Intelligence (AAAI)*. Heterogeneous Transfer Learning for Image Classification, (2011)

16. GJ Qi, C Aggarwal, T Huang, in *Proceedings of the 20th international conference on World wide web*. Towards semantic knowledge propagation from text corpus to web images (ACM, 2011), pp. 297–306
17. Y Wei, Y Zhao, Z Zhu, Y Xiao, in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP) 2012 Eighth International Conference on*. Knowledge transferring for Image Classification (IEEE, 2012), pp. 347–350
18. W Dai, Q Yang, G Xue, Y Yu, in *Proc. 25th Int'l Conf. Machine Learning*. Self-Taught Clustering, (2008), pp. 200–207
19. H Kong, EK Teoh, JG Wang, R Venkateswarlu, Two dimensional fisher discriminant analysis: Forget about small sample size problem. *Proc. IEEE Intern. Conf. Acoustics Speech, Signal Process.* **2**, 761–764 (2005)
20. MathWorks (2013). [Online]. Available: <http://www.mathworks.com>
21. JF Sturm, Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods Softw.* **11**(1–4), 625–653 (1999)
22. R Zhang, J Ma, An improved SVM method P-SVM for classification of remotely sensed data. *Int. J. Remote Sensing.* **29**(20), 6029–6036 (2008)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
