Journal of Applied Remote Sensing

Support vector machine-based decision tree for snow cover extraction in mountain areas using high spatial resolution remote sensing image

Liujun Zhu
Pengfeng Xiao
Xuezhi Feng
Xueliang Zhang
Zuo Wang
Luyuan Jiang

SPIE

Support vector machine-based decision tree for snow cover extraction in mountain areas using high spatial resolution remote sensing image

Liujun Zhu, Pengfeng Xiao,* Xuezhi Feng, Xueliang Zhang, Zuo Wang, and Luyuan Jiang

Nanjing University, Department of Geographic Information Science, Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, and Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China, 163 Xianlin Road, Nanjing 210023, China

Abstract. Snow cover extraction in mountain areas is a complex task, especially from high spatial resolution remote sensing (HSRRS) data. The influence of mountain shadows in HSRRS is severe and normalized difference snow index-based snow cover extraction methods are inaccessible. A decision tree building method for snow cover extraction (DTSE) integrated with an efficiency feature selection algorithm is proposed. The severe influence of terrain shadows is eliminated by extracting snow in sunlight and snow in shadow separately in different nodes. In the feature selection algorithm, deviation of fuzzy grade matrix is proposed as a class-specific criterion which improves the efficiency and robustness of the selected feature set, thus making the snow cover extraction accurate. Two experiments are carried out based on ZY-3 image of two regions (regions A and B) located in Tianshan Mountains, China. The experiment on region A achieves an adequate accuracy demonstrating the robustness of the DTSE building method. The experiment on region B shows that a general DTSE model achieves an unsatisfied accuracy for snow in shadow and DTSE rebuilding evidently improves the performance, thus providing an accurate and fast way to extract snow cover in mountain areas. [DOI: 10.1117/1.JRS.8.084698]

Keywords: decision tree for snow cover extraction; mountain areas; high spatial resolution remote sensing image; deviation of fuzzy grade matrix; feature selection; tree rebuilding.

Paper 13487SS received Nov. 28, 2013; revised manuscript received Feb. 10, 2014; accepted for publication Feb. 12, 2014; published online Apr. 2, 2014.

1 Introduction

Global snow and ice cover plays an important role in regulating surface energy budget, water cycle, sea level change, and surface gas exchange through a range of complex interactions and feedbacks.[1,2] Equally important, seasonal snow cover in the form of mountain snowpack, as temporal water storage, is the main water source of seasonal streamflow[3] and a valuable source of energy when a large volume of water is released during the melting process.[4] Much of the uncertainty and sensitivity in the regional hydrological cycle lies in these reservoirs of frozen water.[5] It is clear that hydrological researches and applications, e.g., snowmelt run-off prediction and water resource management, will benefit from accurate snow cover information.[6,7] Remote sensing, especially the high spatial resolution remote sensing (HSRRS), can provide well-distributed spatial and temporal snow cover information which evidently augments the hydrology researches and applications.

Over the last four decades, optical sensors on different satellites and corresponding algorithms have been developed to extract snow cover at different scales, e.g., SNOMAP,[8] S3.[9] These algorithms are mainly based on the spectral characteristics of snow, i.e., very high

*Address all correspondence to: Pengfeng Xiao, E-mail: xiaopf@nju.edu.cn

0091-3286/2014/$25.00

reflectance at visible wavelengths and low reflectance at shortwave-infrared wavelengths.[10,11] However, in rugged terrain, the reflectance of snow in shadow will be lower than that of soil or vegetation in sunlight thus the snow cover in shadows is hard to be distinguished from others. Some literatures have reported that principal components transform, band ratios, and supervised classifiers without topographic correction cannot effectively distinguish snow-free from snow.[12] Hence, topographic correction based on digital elevation model (DEM) becomes the essential preprocessing to extract snow cover in mountain areas.[13–15]

HSRRS has rapidly developed in the last decade. However, little attention is paid to high spatial resolution snow cover extraction, especially in mountain areas. Since the spectral resolution of most of the HSRRS sensors, e.g., IKONOS, QuickBird, and WorldView, is relativity low and the shortwave-infrared wavelengths are usually not covered, the normalized difference snow index (NDSI) cannot be obtained. Hence, some alternative snow cover extraction methods were proposed for HSRRS or digital camera, such as normalized difference snow and ice index (NDSII based on visible and near-infrared bands)[16] and RGBNDSI (visible bands are substituted for the shortwave-infrared band to compute NDSI).[17] Unsupervised classification algorithm, e.g., self-organizing data analysis technique (ISODATA) is also applied to snow cover extraction from HSRRS data.[18] However, the performance of extracting snow cover in terrain shadows and the generalization capability of these methods are not clear. As a result of relative high spatial resolution, the effect of shadows on HSRRS image is more severe than that in middle-resolution image. What is worse, high quality DEM that matches with HSRRS is normally inaccessible making topographic correction a tough task for HSRRS. In summary, the problems of HSRRS snow cover extraction in mountain areas are as follows: (1) NDSI cannot be obtained from HSRRS, thus the method based on NDSI cannot be directly used and (2) the influence of terrain shadows is more severe and topographic correction can hardly work. Therefore, a robust method for snow cover extraction using HSRRS is needed.

Fortunately, many new techniques, which combine the rich spatial features with spectral features, have been proposed for the classification of HSRRS images. These techniques have achieved satisfactory results in many complicated tasks, such as forest ecosystem mapping[19] and urban shaded area classifing,[20] and are thus suggested in the research of HSRRS snow cover extraction in mountain areas. The type of structure element (pixel or object), the classification model, and the extraction of texture features are treated as the key issues of a successful classification scheme.[21] Current research has shown that glacier and snow mapping based on object-based image analysis (OBIA) have at least 3% higher quality than pixel-based image analysis.[22]

For the classification model, support vector machine (SVM) is a valuable classifier widely used in the classification of HSRRS data for its ability to generalize well even with limited training samples.[23] Decision tree is another powerful classifier widely used in HSRRS applications, such as landscape mapping[24] and vegetation extracting in urban area.[25] Two kinds of techniques, i.e., top-down and bottom-up strategies, are usually used to determine the decision tree. Both techniques share the criterion that the interested classes or the classes having fewer overlays with other classes in the feature space are separated in top levels. Recently, the combination of SVM and decision tree is suggested in pattern recognition due to their complementary merits[26,27] and widely used in remote sensing applications.[21,28,29]

Multiscale texture features, such as gray-level co-occurrence matrix (GLCM),[30] pixel shape index (PSI),[31] morphological profiles (MP),[32,33] and wavelet-based texture[34] can evidently improve the accuracy of snow cover extraction. However, classification with more features does not naturally result in higher accuracy as classifiers are subject to the so-called Hughes effect.[35] What is more, a large feature set means more computation cost. Hence, an effective wrapper feature selection algorithm (refers to feature selection algorithms that tightly connect to the classifiers) is needed. There are many wrapper feature selection algorithms for classification, and some popular ones are listed in Ref. 36. The overall accuracy of the classifier on the selected feature subset is usually used to determine which features are selected. However, this criterion is a global indicator of performance that takes misclassification among nonsnow classes into consideration which we are not interested in, thus is not suitable for HSRRS snow cover extraction. What is more, feature selection based on overall accuracy may result in local optimum as a result of overfitting,[37] which would thereby reduce the generalization ability of

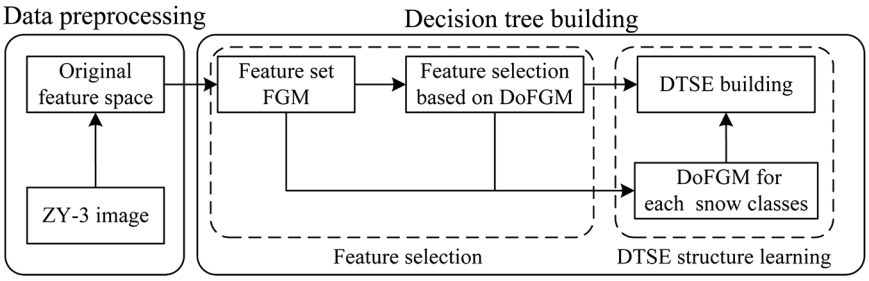

Fig. 1 Interconnect of design parts involved in this article.

selected features. Therefore, a class-specific criterion that can comprehensively reflect the performance of classification system is needed.

In this article, an SVM-based decision tree building method for snow cover extraction (DTSE) integrated with a feature selection method is proposed. The difficulty that finds high resolution DEM for topographic correction is overcome by extracting snow in sunlight and snow in shadow separately and sequentially in different nodes. The outline of this method is shown in Fig. 1. First, fuzzy grade matrix (FGM) based on SVM outputs is introduced as a fuzzy evaluation of the performance of classifiers. Then, deviation of fuzzy grade matrix (DoFGM) for specific class is proposed as a simplified criterion of FGM, which is also a fuzzy criterion that merely takes the interested classes into consideration. Finally, DTSE structure learning algorithm that integrates sequential forward feature selection algorithm based on DoFGM is suggested.

The rest of this article is organized as follows. Section 2 introduces the study area and data preprocessing procedures, including image segmentation and feature extraction. Detailed description of the proposed method is presented in Sec. 3. Experimental results are provided in Sec. 4. Finally, concluding remarks are provided in Sec. 5.

2 Data and Preprocessing

2.1 *Study Area and Data*

Two representative regions A and B, located in Tianshan Mountains in northwestern China, are chosen in the study. Both are in alpine area with large topographic relief and complex terrain conditions (Fig. 2). The area of each region is 81 km^2. Elevation of regions A and B range from 3141 to 4655 m and 2582 to 4736 m, respectively. The ratio of area where slope is larger than 30 deg is 33.19% and 43.65% in regions A and B, respectively.

The test images are subsets of a ZY-3 CCD image acquired on December 22, 2012, in which snow widely covered the area. ZY-3 is the first Chinese high-resolution optical transmission stereosatellite, which was launched on January 9, 2012. The satellite is equipped with a CCD having four bands, i.e., band 1 0.45 to 0.52 μm, band 2 0.52 to 0.59 μm, band 3 0.63 to 0.69 μm, and band 4 0.77 to 0.89 μm, and the spatial resolution is 5.8 m.

2.2 *Segmentation and Feature Extraction*

The method proposed in this article is based on OBIA, thus segmentation with well performance is the key step for successful snow cover extraction. There are many segmentation algorithms for remote sensing image, such as eCognition,[38] SCRM,[39] and EWS.[40] For this study, boundary-constrained multiscale segmentation method is used to partition image into objects for its high segmentation accuracy and low computation cost.[41] A small scale factor of 30 is set for both images to avoid under-segmented objects.

In this article, three feature sets: spectral features, Haralick features (GLCM), and shape features are chosen as candidate features. All the features are extracted on the object level. Detailed information of candidate features is shown in Table 1.

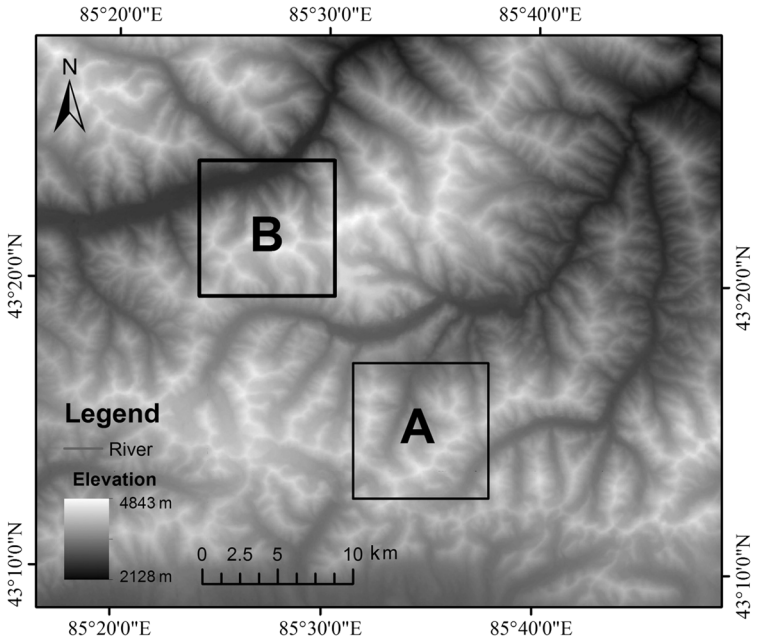

Fig. 2 Two study areas A and B in Tianshan Mountains, northwestern China.

1. Spectral features (32 features): These features include four statistics (average, variance, skewness, and range) of spectral and spectrally related data, i.e., four spectral bands of ZY-3 CCD, NDSII (red − infra red/red + infra red), the ratio vegetation index (RVI; infrared/red), shadow detection index for snow (red − infra red/blue + infra red), and the first principal component (PC1).
2. Haralick features (32 features): GLCM is built for each object for each spectral band first. Next, eight features based on GLCM are computed by setting the interpixel distance as 1 pixel. In order to obtain directional invariance, the extracted features are then averaged over four directions, i.e., 0, 45, 90, and 135 deg (same to the definition of azimuth).
3. Shape feature (1 feature): Although the shape of snow objects is not significantly different from other classes, the size of objects in shadows is evidently larger than that in sunlight, as a result of lower DN range in shadows than that in sunlight. Thus, the size of objects is chosen as a candidate feature in an attempt to augment the discrepancy between snow in shadow and others in sunlight.

3 Method

3.1 *Support Vector Machine*

Since DTSE is constructed using a portion of the training patterns to discriminate snow from others at different nodes, overfitting will occur if the classifier of the node is not robust with the noisy training patterns. The SVMs remarkable performance with regard to sparse and noisy data makes it suitable for DTSE. SVM is intrinsically a two-class classifier. For a two-class problem, the pattern set can be expressed as $[(x_i, y_i), i = 1, \ldots, N]$. Each pattern $p_i(x_i, y_i)$ is a vector in the d-dimensional feature space $x_i = [x_{i,1}, x_{i,2}, \ldots, x_{i,d}]^{\mathrm{T}} \in R^d$ with associated targets $y_i \in \{-1, 1\}$. The discriminant function in the kernel space can be defined as

$$f(x) = \sum_{i \in S} a_i y_i K(x_i, x) + b, \tag{1}$$

Table 1 Description of candidate features.

Feature family	Data description	Statistic	Abbreviation	Feature amount
Spectral features	Band 1, band 2, band 3, band 4	Four statistics (mean, range variance, skewness)	B_statistic	4×4
	NDSII		NDSII_statistic	4
	RVI		RVI_statistic	4
	Band 1 – band 4/band 1 + band 4		ND14_statistic	4
	PC1		PC1_statistic	4
Haralick features (GLCM)	Band 1, band 2, band 3, band 4	Mean	B_Ham	4
	Band 1, band 2, band 3, band 4	Variance	B_Hav	4
	Band 1, band 2, band 3, band 4	Homogeneity	B_Hah	4
	Band 1, band 2, band 3, band 4	Contrast	B_Hacon	4
	Band 1, band 2, band 3, band 4	Dissimilarity	B_Had	4
	Band 1, band 2, band 3, band 4	Entropy	B_Haent	4
	Band 1, band 2, band 3, band 4	Energy	B_Haene	4
	Band 1, band 2, band 3, band 4	Correlation	B_Hacor	4
Shape	Size of object	Sum	Size	1

where $S = \{i : 0 < a_i \leq C\}$. The samples associated with nonzero a_i are so-called support vectors (SVs). $f(x) = \pm 1$ are the negative and positive margins, and ambiguous patterns are among them. $K(x_i, x)$ is the kernel function, in this study, Gaussian RBF kernel was employed

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2), \tag{2}$$

where γ represents a parameter inversely proportional to the width of the Gaussian kernel. γ and C are determined via genetic algorithms (GAs)[42] where $\gamma \in (0, 2)$ and $C \in (0,100)$.

As previously mentioned, SVMs are intrinsically binary classifiers. For multiclass applications, different multiclass strategies can be adopted,[43] such as one against all (OAA), one against one (OAO), and directed acyclic graph. Since DoFGM is a criterion based on OAA strategy, more attention is paid to the commonly used OAA strategy in this article. Briefly, let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ be a set of M possible classes. The OAA strategy consists of an ensemble of M (parallel) binary SVM classifiers $\{f_1, \ldots, f_k, \ldots, f_M\}$, each classifier f_k solves a binary classification that discriminates class ω_k from the rest of classes $(\Omega - \omega_k)$. Then the "winner-takes-all" rule is used to decide the label of each pattern, which can be expressed as

$$\omega = \arg \max \left\{ f_k(x) = \sum_{i \in S} a_i y_i K(x_i, x) + b \right\}. \tag{3}$$

3.2 *Deviation of Fuzzy Grade Matrix*

In this section, a class-specific criterion DoFGM is proposed for feature selection. Let a pattern set $\Omega = \{\omega_1, \ldots, \omega_k, \ldots, \omega_M\}$, $\omega_k = \{(x_i, k), i = 1, 2, \ldots, N_i\}$, where (x_i, k) indicates a pattern p_i whose class label is k, the pattern number of Ω is $N = \sum_{i=1}^{M} N_i$. In our task, the pattern set is indeed the sample set used for DTSE building. The detail descriptions of DoFGM are as follows.

3.2.1 *Fuzzy grade matrix*

The FGM of a pattern set is a matrix that records the fuzzy membership grades of each pattern to multiple classes. The construction of FGM of this pattern set is concluded as follows:

Step 1: Adhering to OAA strategy, OAA class binarization method is used to transform the M-class problem into M binary problems. For each class ω_k, a binary problem $\langle \omega_k, \Omega_k \rangle$ where $\Omega_k = \bigcup_{\substack{j=1 \\ j \neq k}}^{M} \omega_j$ is created.

Step 2: An ensemble of M binary SVM classifiers $\{f_1, \ldots, f_k, \ldots, f_M\}$ is constructed. Each binary classifier f_k is designed to solve one binary problem created in step 1, and each pattern p_i will get an output $f_k(p_i)$ from f_k. Thereby, each pattern p_i will have M outputs $\{f_1(p_i), \ldots, f_k(p_i), \ldots, f_M(p_i)\}$, which can indicate the ambiguous of pattern p_i to each class.

Step 3: A sigmoid function is employed to transform the SVM outputs of all patterns to fuzzy grades (FGs), which is defined as

$$\mathrm{FG}_k(p_i) = \frac{1}{1 + e^{[\ln(\alpha) \cdot f_k(p_i)]}}, \tag{4}$$

where $\mathrm{FG}_k(p_i)$ is the fuzzy grade of pattern p_i to class k, α is a parameter that control the shape of function. Obviously, $\mathrm{FG}_k(p_i)$ is positively correlated to $f_k(p_i)$. That is to say, positive (negative) value of $f_k(p_i)$ leads to higher (lower) fuzzy grade to class k. In this paper, α is equal to 0.02, thereby $f_k(p_i) > 1$ and $f_k(p_i) < -1$ will result in a fuzzy grade extremely close to 1 and 0, respectively. After this step, there is a fuzzy grade vector $\mathrm{FGV}(p_i) = [\mathrm{FG}_1(p_i), \ldots, \mathrm{FG}_M(p_i)]$ for each pattern p_i. $\mathrm{FGV}(p_i)$ can reflect the performance of classification system including the classification model and the selected feature subsets on pattern p_i.

To illustrate the construction of FGVs, a case of three-class problem is shown in Fig. 3. Figure 3(a) is the result of class binarization, a total of three binary problems are created. Then, three SVMs are constructed in step 2 [see Fig. 3(b)], which result in three hyperplanes, i.e., f_1, f_2, and f_3 in two-dimension feature space. Taking pattern p_i as an example, it achieves three SVM outputs in this step, namely $f_1(p_i)$, $f_2(p_i)$, and $f_3(p_i)$. In Fig. 3(c), the SVM outputs of p_i are transformed to the FGV of p_i.

Step 4: The FGM of Ω is then defined as $\mathrm{FGM}(p_i) = [\mathrm{FGV}(p_1), \mathrm{FGV}(p_2), \ldots, \mathrm{FGV}(p_N)]$. Thus, each row of FGM is the FGV of a pattern, while each column is the fuzzy grades of all patterns to one class.

The FGM of a pattern set Ω can reflect the overall performance of classification system on Ω. However, FGM is so complex that hardly used as a criterion for feature selection directly. Thus, DoFGM is proposed as a class-specific criterion simplified from FGM in the next section, providing a more accurate description of the local and overall performance of feature sets.

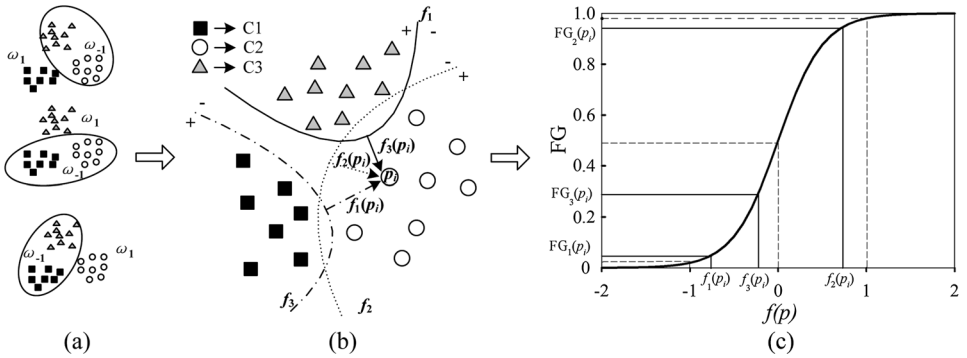

Fig. 3 Construction procedure of fuzzy grade vector. (a), (b) and (c) are corresponding to steps 1, 2, and 3 of the fuzzy grade matrix construction process, respectively.

3.2.2 *Deviation of fuzzy grade matrix*

Ideally, patterns of different classes are expected to be correctly separated by hyperplanes created by a classifier in the selected feature space and be away from hyperplanes as far as possible. With regard to a specific pattern, it definitely belongs to a specific class k with an FG of 1 to class k and FGs of 0 to other classes. The FGM of such pattern set derived from an ideal classification system is called ideal fuzzy grade matrix (IFGM) here. Naturally, the FGM is expected to be the same as IFGM. Thus, the difference between FGM and IFGM of the same pattern set can reflect the difference between the selected classification system and ideal classification system, and can be used to reflect the classification ability of selected feature subset. In this paper, this difference is defined as DoFGM, which is expressed as

$$\mathrm{DoFGM}(\Omega) = \frac{1}{MN}\sqrt{\sum_{i=1,j=1}^{M,N}(\mathrm{FGM}_{i,j} - \mathrm{IFGM}_{i,j})^2}, \tag{5}$$

where M and N are the number of columns and rows of FGM, respectively. Obviously, the $\mathrm{DoFGM} \in (0,1)$. A higher DoFGM means a larger difference between FGM and IFGM, which means lower classification ability.

DoFGM is an overall measurement as it takes all elements of FGM into consideration. It reflects the capability of solving whole classification task including the classes we are not interested in. Hence, a measurement that can reflect the performance of classification system on specific class is needed.

Two indicators, i.e., omission error and commission error, are considered to evaluate the classification capability on a specific class. For patterns of class ω_k, high FGs of these patterns to other classes and low FGs to class ω_k can result in a high omission error. The commission error may occur when patterns of other classes have high FGs to class ω_k. In other words, omission error is related to the difference between corresponding rows (the class label of these rows is ω_k) of FGM and IFGM, which can be expressed as

$$\mathrm{DoFGM}_{\mathrm{o}}(\omega_k) = \frac{1}{{}^{\#}I_k M}\sqrt{\sum_{i=1,j=1}^{{}^{\#}I_k,M}(\mathrm{FGM}_{i,j} - \mathrm{IFGM}_{i,j})^2}, \tag{6}$$

where I_k is the set of row number of patterns that belongs to ω_k, and ${}^{\#}I_k$ is the element amount of I_k. The commission error is related to the difference between corresponding columns (the fuzzy grade of all patterns to class ω_k) of FGM and IFGM, which can be expressed as

$$\mathrm{DoFGM}_{\mathrm{c}}(\omega_k) = \frac{1}{N - {}^{\#}I_k}\sqrt{\sum_{i=1,i\notin I_k}^{N}(\mathrm{FGM}_{i,\omega_k} - \mathrm{IFGM}_{i,\omega_k})^2}, \tag{7}$$

where $\mathrm{FGM}_{i,\omega_k}$ represent the fuzzy grade of pattern p_i to class ω_k. DoFGM considering both errors for specific class can be computed as

$$\mathrm{DoFGM}(\omega_k) = \alpha\mathrm{DoFGM}_{\mathrm{o}}(\omega_k) + (1-\alpha)\mathrm{DoFGM}_{\mathrm{c}}(\omega_k), \tag{8}$$

where α (user-defined) reflects the importance attached to omission error, in this paper, α is equal to 0.5.

To illustrate the DoFGM for specific class, a case is shown in Fig. 4. There are four patterns belong to four different classes, namely snow in sunlight, snow in shadow, others in sunlight, and others in shadow, respectively, each of which corresponds to a row of FGM. $\mathrm{DoFGM}_{\mathrm{o}}$ and $\mathrm{DoFGM}_{\mathrm{c}}$ for snow in shadow are computed based on elements with same color using Eqs. (6) and (7), respectively. Then the DoFGM for specific class is obtained by Eq. (8).

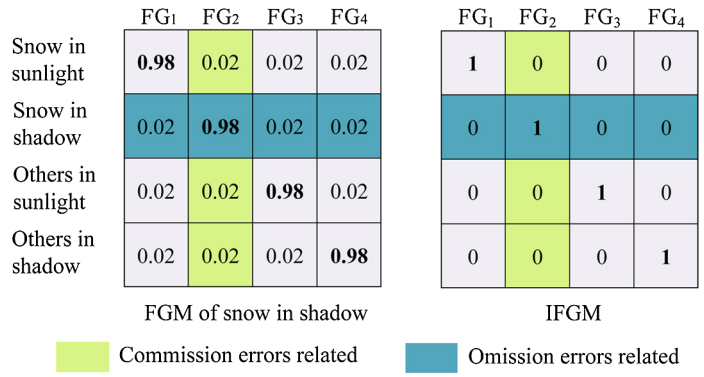

Fig. 4 Concept of deviation of fuzzy grade matrix for specific class.

3.3 *Feature Selection Algorithm Based on SVM-DoFGM*

The sequential forward feature selection strategy combined with DoFGM is suggested in this study with regard to its relatively low computation cost. This method selects the feature with minimum DoFGM as the first feature, and then adds the feature that has maximum contribution with respect to reducing DoFGM to optimal feature subset.

Consider a set of initial features $F = \{f_1, \ldots, f_k, \ldots, f_T\}$, where T denotes the total number of features. $F^*(t)$ is the optimum feature subset at iteration t. DoFGM(t) indicates the DoFGM based on the features selected at iteration t. AC(t, f_k) denotes the additional contribution of feature f_k with respect to $F^*(t-1)$. Feature selection procedure stops when the improvement percentage of selected feature with respect to the current optimal feature set is less than e_z (a user-defined threshold, i.e., 1%) or the search completes.

The detailed procedure of feature selection for specific class according to SVM-DoFGM is shown in Table 2. At step 1, some parameters are predefined to ensure that the output F^* is not an $\varnothing$. Step 2 is an iteration controlled by the stop criterion. At step 3, a powerful and nonredundant feature subset is output.

Table 2 Feature selection algorithm based on SVM-DoFGM.

Step 1 Initialization

Set $F^* = \varnothing$, $t = 1$, DoFGM$(0) = \varepsilon$ (a very small float).

Step 2 Feature selection based on DoFGM

Do begin

2.1 Add each remaining feature f_j to F^* temporally and compute the additional contribution (AC) of this feature

$$\mathrm{AC}(t, f_j) = \mathrm{DoFGM}(t) - \mathrm{DoFGM}(t-1)\ j = 1, \ldots, n,\ j \neq l_1, \ldots, l_{t-1}$$

2.2 Find $f_{l_t} \in F$

$$l_t = \underset{\substack{j=1,\ldots,n \\ j \neq l_1, \ldots, l_{(t-1)}}}{\arg\max} \{\mathrm{AC}(t, f_j)\}$$

2.3 Calculate the improvement percentage of f_{l_t} with respect to DoFGM$(p-1)$

$$l_{l_t} = \{[(|\mathrm{AC}(t, f_{l_t})|)/(|\mathrm{DoFGM}(t-1)|)]\} \times 100\%$$

2.4 $F^*(t) \leftarrow F^*(t-1) + \{f_{l_t}\}$, $t \leftarrow t+1$, $t = \mathrm{t}+1$

End until ($l_{l_t} \leq e_z$ or $t \leq T$)

Step 3 Output: The set $F^* = \{f_{l_1}, \ldots, f_{l_{t-1}}\}$ of the $t-1$ finally selected features.

3.4 *DTSE Structure Learning*

As previously mentioned, DTSE is a decision tree that extracts snow cover in sunlight and snow cover in shadows separately and sequentially at different nodes using binary SVMs. For a minimum accumulation error, the two snow classes will be extracted at the first and second nodes and the sequence is decided by the minimum DoFGM of each snow class. A top-down procedure combined with feature selection is suggested for DTSE structure learning, which is expressed in Table 3.

Figure 5 is the flow chart of the DTSE structure learning process. The snow cover (labeled as 1) having the minimum DoFGM is selected at level 1, and separated from remaining classes. Then feature selection is applied for the remaining snow cover (labeled as 2) and separated from the others at level 3. Finally, DTSE structure learning finished with three levels.

3.5 *Validation Methods*

In this paper, we use a set of four binary metrics that rely on the common classification of possible outcomes in identifying a polygon as snow or snow free, i.e., true positive (TP), true negative (TN), false positive (FP), and false negative (FN).[44] Three measures based on these metrics are

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{9}$$

Table 3 Procedure of DTSE structure learning.

Step 1 Apply feature selection introduced in Sec. 3.3 and select the optimum feature subset for each snow class

Step 2 Find the snow class (leaf node candidate) having the minimum DoFGM:

$$C_1 = \arg\min_{i=1,2}\{\text{DoFGM}_i\}$$

Class C_1 has the lowest DoFGM indicating that separate the class C_1 first can improve the final performance.

Step 3 Train a binary SVM on the feature subset $F^*_{c_1}$ to discriminate class C_1 and others

Step 4 Apply feature selection introduced in Sec. 3.3 for class C_2 and train a binary SVM on the feature $F^*_{c_2}$ to discriminate class C_2 and others.

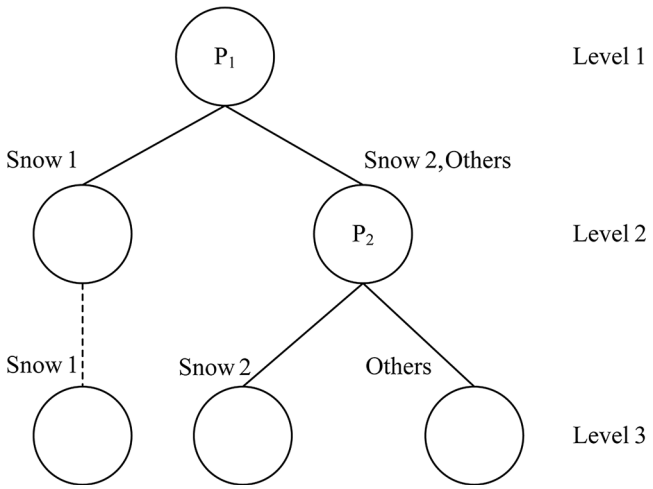

Fig. 5 Flow chart of the decision tree building method for snow cover extraction (DTSE) structure learning process.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{10}$$

$$F = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \tag{11}$$

Precision is the probability that a polygon identified as snow in deed has snow. Recall is the probability of detection of a snow-covered polygon. The F score is recommended as it balances precision and recall in Ref. 44. For comprehensive evaluations of proposed algorithm, the F score of the whole task and the F score of a specific snow class are used simultaneously. When calculating F of whole task, snow in sunlight and snow in shadow are merged into one class. When calculating F of a specific snow class, the misclassification between snow classes is not considered. Taking F of snow in sunlight as an example, TP is the number of snow in sunlight polygons which are correctly identified; FP is the number of snow in sunlight polygons which are identified as snow free; FN is the number of snow free polygons that identified as snow in sunlight and the misclassification between snow in sunlight and snow in shadow is omitted.

4 Experiments

In this section, the DTSE building method and the DTSE models are comprehensively tested with two experiments and three sample sets.

4.1 *Samples for Experiments*

The test areas are human-unreachable during the snow season. Thus, 800 samples (polygons) in region A and 100 samples in region B were selected randomly from the image and identified using photointerpretation to form three data sets. The first data set (DS-1) including 500 samples over the region A was used to select optimal feature subsets and to train the DTSE via cross-validation. Then the decision trees were tested on the second data set (DS-2) comprising 300 samples over the region A in an attempt to validate their generalization capabilities in the same region. The third data set (DS-3) including 100 samples over the region B was used to rebuild the decision trees, which would improve the performance of decision trees on region B. The considered classes and the information of samples are shown in Table 4.

4.2 *DTSE Building*

An experimental scheme which is similar to Ref. 21 was used here. A fivefold cross-validation was used to investigate the robustness of DTSE building method. DS-1 was randomly divided into five equal subsets, from which four subsets were used for DTSE building and the remaining subsets were used for testing. After five repetitions, five different DTSEs were obtained. The SVM parameters of different models were determined by GA,[42] where the maximum

Table 4 Information of classes and three data sets.

	DS-1	DS-2	DS-3
Snow in sunlight (ω_1)	125	75	25
Snow in shadow (ω_2)	125	75	25
Others in sunlight (ω_3)	125	75	25
Others in shadow (ω_4)	125	75	25
Total	500	300	100
Source of samples	Image of region A	Image of region A	Image of region B

Table 5 Structure and performance of five DTSEs.

	Node 1				Node 2				
	Class	C	γ	Feature	Class	C	γ	Feature	DS-2 F
DTSE1	Snow in sunlight	0.17	0.920	B1_m B1_Hacon	Snow in shadow	0.51	0.10	PC1 B1_Ham	0.920
DTSE2		1.90	0.937	B1_m B4_Hacon		1.76	0.03	PC1 B2_Hav	0.937
DTSE3		1.29	0.923	B1_m B4_Hacon B1_Hacon		1.29	0.02	B1_m B4_Hacon B1_Hacon	0.923
DTSE4		0.68	0.935	B1_m B4_Hacon		0.93	0.02	PC1 B1_Ham	0.935
DTSE5		0.38	0.944	B1_m B3_Hacon		1.16	0.01	PC1 b3_Ham b3_Hacon	0.944

generations, size of population, crossover probability, and mutation probability are 100, 20, 0.4, and 0.01, respectively.

The structures of five DTSEs are shown in Table 5, and the features are sorted as the selecting sequence. This means that the most powerful feature is in the front of the feature set. Snow cover in sunlight was extracted first and snow cover in shadows was extracted later in all five DTSEs, which is a natural result as snow cover in sunlight has a very high reflectance in visible bands and can be discriminated from other classes easier than snow cover in shadows. With regard to features selected for each snow class, some comments are as follows according to Table 5. (1) The optimal subsets are composed of at least one spectral feature and one textural feature for all DTSEs, which is coincident with the widely accepted knowledge that the combination of spectral and textural features can evidently improve the classification performance. Furthermore, spectral features are more powerful than texture features for snow extraction in view of their sequences in feature set. (2) For snow in sunlight, B1_m is the most powerful feature. In addition, due to the relatively large DN ranges of all bands of snow in sunlight, the variation of local pixel-pairs is large. Hence, the GLCM contrast is also quite robust. (3) For snow in shadow, PC1 is a robust and powerful feature, while the classification ability of other features varies with different training sets. The performance of five DTSEs is also shown in Table 5. The F of five DTSEs on DS-2 is over 0.92, which means that the proposed method performs well.

The best model DTSE5 was then applied to the entire region A (Fig. 6) as a more general test. An accuracy assessment based on 1000 random points whose labels identified by human interpretation was carried out. The results show that both snow classes achieve adequately high

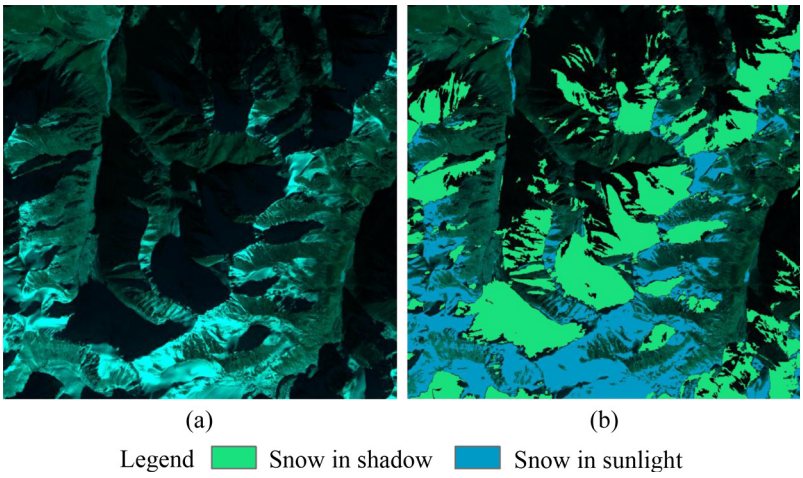

Fig. 6 Snow extraction result in region A. (a) is the image of region A with band combination of 4, 3, and 2 and (b) is the extraction result of DTSE5.

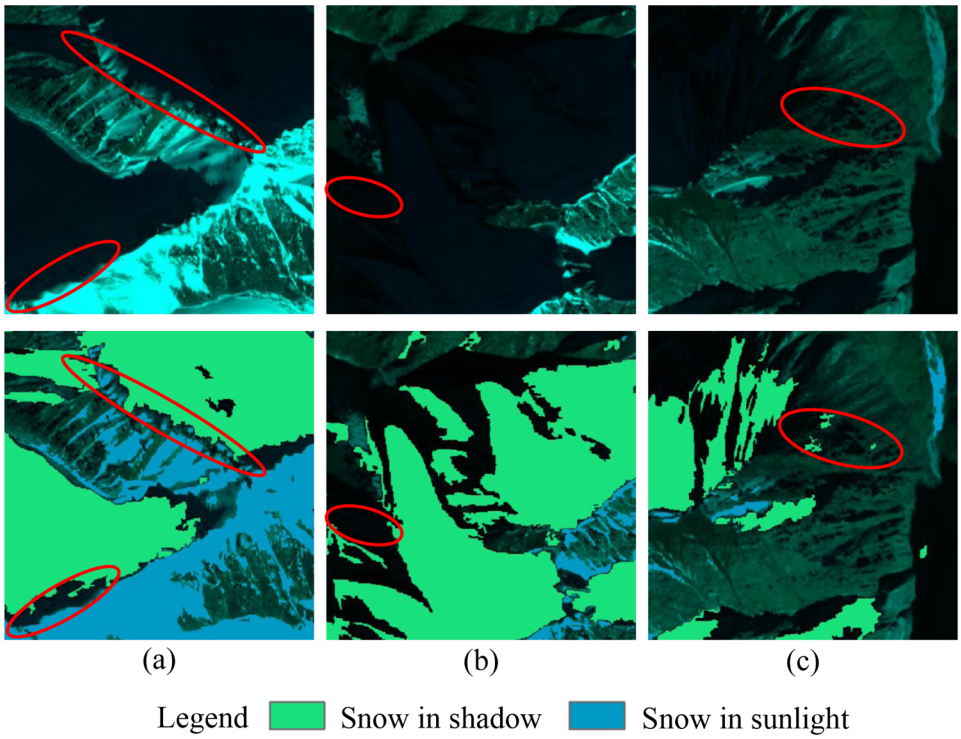

Fig. 7 Local comparisons of original images and extraction results. (a) The discontinuity of snow cover extraction nearby the boundary of sunlight and shadows, (b) the omission error of snow in shadow, and (c) the commission error caused by fraction polygons.

accuracy ($F = 0.905$). The F of snow in sunlight and snow in shadow is as high as 0.924 and 0.897, respectively.

Furthermore, the original image and classification results are zoomed in for visual analysis, as is shown in Fig. 7. Generally, both snow classes have satisfactory results in all local areas. However, some problems cannot be omitted yet. (1) The red ellipses in Fig. 7(a) indicate that the snow cover nearby the boundary of sunlight and shadows was not extracted; (2) though the method is quite excellent in dealing with snow in shadow, omission error is not absolutely avoided yet, as is shown in the red ellipses in Fig. 7(b); (3) some fraction polygons under complex radiation conditions, such as the red ellipses in Fig. 7(c), can easily cause commission error. Accordingly, some improvements are needed in future. Problems (1) and (3) are inherently related to OBIA, thus postprocessing algorithms are needed to settle these problems. With regard to problem (2), a more representing pattern set for DTSE training is needed, which is time consuming.

4.3 *Generalization Capability of DTSE*

The experiment in Sec. 4.2 shows that the proposed method can achieve adequate accuracy. However, it is somewhat complex and needs the support of a relatively large set of samples, thus cannot be applied to large area quickly. Therefore, the generalization capability of a general DTSE is tested in an attempt to explore a faster method for snow cover extraction in mountain areas.

The DTSE5 obtained in Sec. 4.2 was applied directly to region B following accuracy assessment based on random points. In other words, an extraction procedure without support of local patterns was carried out. The experimental result is shown in Fig. 8(b). Despite the different spectral and terrain conditions of region B, DTSE5 is still robust for the extraction of snow in sunlight, achieving a high F score (0.906). However, the performance of snow in shadow declines obviously, having a quite low F (0.496). It shows that the support of local training patterns is necessary to the successful extraction of snow in shadow.

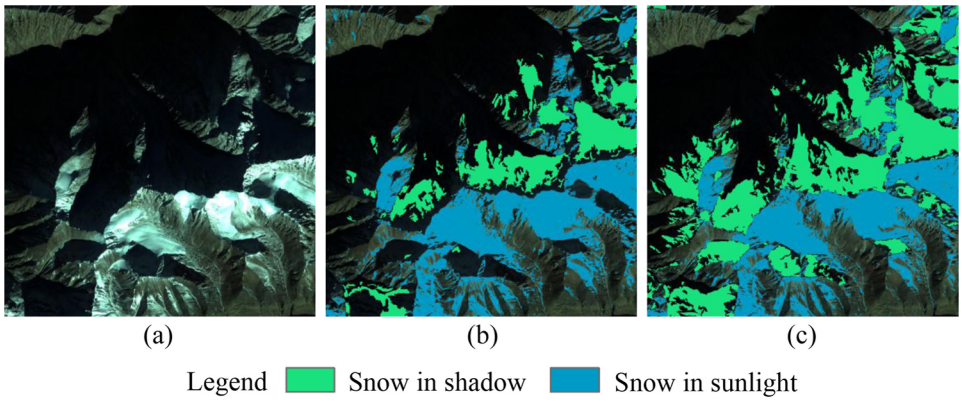

Fig. 8 Results of experiments on region B. (a) is the CCD image of region B with band combination of 4, 3, and 2, (b) is the extraction result of region B based on DTSE5, and (c) is the extraction result of region B based on DTSE5*.

Accordingly, a small pattern set DS-3 together with the SV sets of DTSE5 was employed to rebuild the DTSE5 using the DTSE structure learning algorithm. The new extraction model DTSE5* was then applied to region B following accuracy assessment. The extraction result is shown in Fig. 8(c). There is no significant difference between two experiments for snow extraction in sunlight, while the extraction of show in shadow evidently benefits from the DT-3 achieving an adequate F (0.890). Thus, the DTSE rebuilding strategy which is based on a small extra number of patterns and a general DTSE (such as DTSE5) provides a simple way to extract snow cover in mountain areas as SNOMAP does in plain areas.

5 Conclusion

An SVM-based decision tree for snow cover extraction from HSRRS image is presented, providing high accuracy and interpretable structure. This method extracts snow in sunlight and snow in shadow separately and sequentially, overcoming the difficulty of high resolution topographic correction. The feature selection algorithm based on DoFGM makes full use of rich spatial and spectral information of HSRRS and reduces the Hughes effect simultaneously, thus enhances the extraction and generalization capabilities of decision trees. The DTSE structure learning algorithm simplifies the extraction procedure, providing a fast method for snow cover extraction.

The experiments on region A demonstrate the effectiveness of proposed method. The snow in sunlight and snow in shadow achieve high F scores of 0.918 and 0.890, respectively. Though the proposed method is dependent on local training patterns, and DTSE that built for specific region cannot be directly applied to other regions, DTSE rebuilding is proposed as a supplement for a fast snow cover extraction in mountain areas based on a general DTSE. The efficiency of this supplement has been demonstrated in experiments on region B.

Generally, the DTSE is an object-based classification framework with some special consideration for snow cover extraction in rugged terrain using HSRRS. Thus, other HSRRS data similar to ZY-3 CCD image, e.g., IKONOS, QuickBird, and WorldView, can be directly used to extract snow cover with the proposed method. Regard to relatively moderate resolution data such as Landsat and ASTER with the aid of well-matched DEM, current researches have shown that topographic correction can rebuilt the shadow areas effectively and make the extraction of snow cover an easy task.[13–15] The proposed method can be directly used to extract snow cover after topographic correction as it did in extracting snow cover in sunlight. What is more, terrain shadows can be detected easily based on high quality DEM. Hence, we can escape from the dilemma caused by the overlay in feature space between snow in shadow and others in sunlight. Therefore, DTSE will perform much better with the assistance of high quality DEM. If high quality DEM is unavailable, two problems should be considered carefully to extract snow cover using moderate resolution data based on DTSE: (1) whether OBIA is necessary with the consideration of easy access; (2) which feature family should be added into or removed from the

candidate feature set for feature selection. Thus, further researches are necessary and some adjustments are needed.

More generally, the proposed method can work as a common classification framework subject to some adjustments. They are: (1) an iterative procedure is necessary to the decision tree structure learning algorithm, thus, the proposed method can be applied to multiclass classification; (2) DoFGM for specific and total DoFGM could be used together for feature selection; (3) other feature families such as PSI, MP, and wavelet-based texture can be added to candidate feature set for more applications. What is more, the feature selection method based on DoFGM can be independently applied to dimensionality reduction of HSRRS image classification. Furthermore, total DoFGM and DoFGM for specific classes can be used as evaluation criterions independently with other search strategies, such as complete search and random search.[36]

Similar to most of the supervised classification methods, the generalization ability and the dependence on training samples are indeed the essential shortages of the proposed method. An attempt to decline the dependency on local training patterns and candidate features limits was carried out in this paper, which is so-called DTSE-rebuilding. In fact, we just substitute a small training pattern set and the prior knowledge for larger amount of local patterns, which indeed cannot be treated as a complete solution to decline the heavy dependency on training patterns. Two promising improvements are intended to make the proposed method more robust. One is similar to DTSE-rebuilding strategy, a time series images will be used, and the snow cover extraction of former image will be used as prior knowledge for the next image. The other will take the semisupervised techniques to DTSE, reducing the number of training patterns.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (Grant No. 41271353), the Open Foundation of State Key Laboratory of Cryospheric Sciences, Cold and Arid Regions Environment and Engineering Research Institute, Chinese Academy Sciences (Grant No. SKLCS-OP-2013-07), and a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). Sincere thanks are given to Mr. Zhou Chen of Nanjing University for insightful suggestions. The authors gratefully acknowledge the insightful suggestions, comments, and criticisms from the anonymous reviewers.

References

1. P. Lemke et al., "Observations: changes in snow, ice and frozen ground," in *Climate Change 2007: The Physical Science Basis, Contribution of Working Group 1 to the Fourth Assessment Report of the IPCC*, S. Solomon et al., Eds., pp. 337–383, Cambridge University, Cambridge (2007).
2. R. G. Barry, "The role of snow and ice in the global climate system: a review," *Polar Geogr.* **26**(3), 235–246 (2002), http://dx.doi.org/10.1080/789610195.
3. T. P. Barnett, J. C. Adam, and D. P. Lettenmaier, "Potential impacts of a warming climate on water availability in snow-dominated regions," *Nature* **438**(7066), 303–309 (2005), http://dx.doi.org/10.1038/nature04141.
4. S. J. Metsämäki et al., "A feasible method for fractional snow cover mapping in boreal zone based on a reflectance model," *Remote Sens. Environ.* **95**(1), 77–95 (2005), http://dx.doi.org/10.1016/j.rse.2004.11.013.
5. J. Dozier, "Spectral signature of alpine snow cover from the Landsat Thematic Mapper," *Remote Sens. Environ.* **28**, 9–22 (1989), http://dx.doi.org/10.1016/0034-4257(89)90101-6.
6. C. Steele et al., "Sensitivity of the snowmelt runoff model to underestimates of remotely sensed snow covered area," in *in Proc. IEEE Int. Geoscience and Remote Sensing Symposium*, pp. 25–30, IEEE, Honolulu (2010), http://dx.doi.org/10.1109/igarss.2010.5653240.

7. S. W. Lee, A. G. Klein, and T. M. Over, "A comparison of MODIS and NOHRSC snow-cover products for simulating streamflow using the Snowmelt Runoff Model," *Hydrol. Processes* **19**(15), 2951–2972 (2005), http://dx.doi.org/10.1002/(ISSN)1099-1085.
8. D. K. Hall, G. A. Riggs, and V. V. Salomonson, "Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data," *Remote Sens. Environ.* **54**(2), 127–140 (1995), http://dx.doi.org/10.1016/0034-4257(95)00137-P.
9. Y. Shimamura, T. Izumi, and H. Matsuyama, "Evaluation of a useful method to identify snow-covered areas under vegetation-comparisons among a newly proposed snow index, normalized difference snow index, and visible reflectance," *Int. J. Remote Sens.* **27**(21), 4867–4884 (2006), http://dx.doi.org/10.1080/01431160600639693.
10. S. G. Warren, "Optical-properties of snow," *Rev. Geophys.* **20**(1), 67–89 (1982), http://dx.doi.org/10.1029/RG020i001p00067.
11. W. J. Wiscombe and S. G. Warren, "A model for the spectral albedo of snow. I: pure snow," *J. Atmos. Sci.* **37**(12), 2712–2733 (1980), http://dx.doi.org/10.1175/1520-0469(1980)037<2712:AMFTSA>2.0.CO;2.
12. W. Rosenthal and J. Dozier, "Automated mapping of montane snow cover at subpixel resolution from the Landsat Thematic Mapper," *Water Resour. Res.* **32**(1), 115–130 (1996), http://dx.doi.org/10.1029/95WR02718.
13. H. S. Negi, A. V. Kulkarni, and B. S. Semwal, "Estimation of snow cover distribution in Beas basin, Indian Himalaya using satellite data and ground measurements," *J. Earth Syst. Sci.* **118**(5), 525–538 (2009), http://dx.doi.org/10.1007/s12040-009-0039-0.
14. P. Sirguey, R. Mathieu, and Y. Arnaud, "Subpixel monitoring of the seasonal snow cover with MODIS at 250 m spatial resolution in the Southern Alps of New Zealand: methodology and accuracy assessment," *Remote Sens. Environ.* **113**(1), 160–181 (2009), http://dx.doi.org/10.1016/j.rse.2008.09.008.
15. C. J. Crawford et al., "Multitemporal snow cover mapping in mountainous terrain for Landsat climate data record development," *Remote Sens. Environ.* **135**, 224–233 (2013), http://dx.doi.org/10.1016/j.rse.2013.04.004.
16. J. Hinkler et al., "Automatic snow cover monitoring at high temporal and spatial resolution, using images taken by a standard digital camera," *Int. J. Remote Sens.* **23**(21), 4669–4682 (2002), http://dx.doi.org/10.1080/01431160110113881.
17. J. Hinkler, J. B. Orbaek, and B. U. Hansen, "Detection of spatial, temporal, and spectral surface changes in the Ny-Alesund area 79 degrees N, Svalbard, using a low cost multispectral camera in combination with spectroradiometer measurements," *Phys. Chem. Earth* **28**(28–32), 1229–1239 (2003), http://dx.doi.org/10.1016/j.pce.2003.08.059.
18. S.-H. Kim and C.-H. Hong, "Antarctic land-cover classification using IKONOS and Hyperion data at Terra Nova Bay," *Int. J. Remote Sens.* **33**(22), 7151–7164 (2012), http://dx.doi.org/10.1080/01431161.2012.700136.
19. K. Johansen et al., "Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification," *Remote Sens. Environ.* **110**(1), 29–44 (2007), http://dx.doi.org/10.1016/j.rse.2007.02.014.
20. W. Zhou et al., "Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: a comparison study," *Remote Sens. Environ.* **113**(8), 1769–1777 (2009), http://dx.doi.org/10.1016/j.rse.2009.04.007.
21. S. Moustakidis et al., "SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images," *IEEE Trans. Geosci. Remote* **50**(1), 149–169 (2012), http://dx.doi.org/10.1109/TGRS.2011.2159726.
22. P. Rastner et al., "A comparison of pixel-and object-based glacier classification with optical satellite images," *IEEE J. Sel. Top. App. Earth Obs. Remote Sens.* **6**(4), 1–10 (2013), http://dx.doi.org/10.1109/JSTARS.2013.2274668.
23. G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote* **43**(6), 1351–1362 (2005), http://dx.doi.org/10.1109/TGRS.2005.846154.

24. P. ManojKumar, R. Sugumaran, and D. Zerr, "A rule-based classifier using Classification and Regression Tree (CART) approach for urban landscape dynamics," in *Proc. IEEE Int. Geoscience and Remote Sensing Symposium*, E. LeDrew, Ed., pp. 1193–1194, IEEE, Toronto (2002), http://dx.doi.org/10.1109/IGARSS.2002.1025880.
25. T. R. Tooke et al., "Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications," *Remote Sens. Environ.* **113**(2), 398–407 (2009), http://dx.doi.org/10.1016/j.rse.2008.10.005.
26. K. P. Bennett and J. A. Blue, "A support vector machine approach to decision trees," in *Proc. of IEEE Int. Joint Conference on Neural Networks*, Vol. 3, pp. 2396–2401, IEEE, Anchorage (1998), http://dx.doi.org/10.1109/IJCNN.1998.687237.
27. L. Zhang et al., "Decision tree support vector machine," *Int. J. Artif. Intell. Trans.* **16**(01), 1–15 (2007), http://dx.doi.org/10.1142/S0218213007003163.
28. Y. Tarabalka et al., "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens.* **7**(4), 736–740 (2010), http://dx.doi.org/10.1109/LGRS.2010.2047711.
29. J. Chen, C. Wang, and R. Wang, "Combining support vector machines with a pairwise decision tree," *IEEE Geosci. Remote Sens.* **5**(3), 409–413 (2008), http://dx.doi.org/10.1109/LGRS.2008.916834.
30. R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **3**(6), 610–621 (1973), http://dx.doi.org/10.1109/TSMC.1973.4309314.
31. L. Zhang et al., "A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote* **44**(10), 2950–2961 (2006), http://dx.doi.org/10.1109/TGRS.2006.876704.
32. J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote* **41**(9), 1940–1949 (2003), http://dx.doi.org/10.1109/TGRS.2003.814625.
33. M. Fauvel et al., "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote* **46**(11), 3804–3814 (2008), http://dx.doi.org/10.1109/TGRS.2008.922034.
34. X. Huang, L. Zhang, and P. Li, "A multiscale feature fusion approach for classification of very high resolution satellite imagery based on wavelet transform," *Int. J. Remote Sens.* **29**(20), 5923–5941 (2008), http://dx.doi.org/10.1080/01431160802139922.
35. G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory* **14**(1), 55–63 (1968), http://dx.doi.org/10.1109/TIT.1968.1054102.
36. H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.* **17**(4), 491–502 (2005), http://dx.doi.org/10.1109/TKDE.2005.66.
37. R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.* **97**(1), 273–324 (1997), http://dx.doi.org/10.1016/S0004-3702(97)00043-X.
38. M. Baatz and A. Schäpe, "Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation," in *Angewandte geographische informationsverarbeitung XII*, J. Strobl, T. Blaschke, and G. Griesebner, Eds., pp. 12–23, Wichmann Verlag, Heidelberg (2000).
39. G. Castilla, G. J. Hay, and J. R. Ruiz-Gallardo, "Size-constrained region merging (SCRM): an automated delineation tool for assisted photointerpretation," *Photogramm. Eng. Remote Sens.* **74**(4), 409–419 (2008), http://dx.doi.org/10.14358/PERS.74.4.409.
40. D. Li et al., "An edge embedded marker-based watershed algorithm for high spatial resolution remote sensing image segmentation," *IEEE Trans. Image Process.* **19**(10), 2781–2787 (2010), http://dx.doi.org/10.1109/TIP.2010.2049528.
41. X. L. Zhang et al., "Boundary-constrained multi-scale segmentation method for remote sensing images," *ISPRS J. Photogramm.* **78**, 15–25 (2013), http://dx.doi.org/10.1016/j.isprsjprs.2013.01.002.
42. Y. Bazi and F. Melgani, "Toward an optimal SVM classification system for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote* **44**(11), 3374–3385 (2006), http://dx.doi.org/10.1109/TGRS.2006.880628.

43. C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Networks* **13**(2), 415–425 (2002), http://dx.doi.org/10.1109/72.991427.
44. K. Rittger, T. H. Painter, and J. Dozier, "Assessment of methods for mapping snow cover from MODIS," *Adv. Water Resour.* **51**, 367–380 (2013), http://dx.doi.org/10.1016/j.advwatres.2012.03.002.

Liujun Zhu received the BS degree in geography science from Zhejiang Normal University, Jinhua, China, in 2012. From 2012 to now, he is a graduate student majored in cartography and geographical information system at Nanjing University, Nanjing, China. His research interests include remote sensing image processing and high-resolution remote sensing of snow cover.

Pengfeng Xiao received the BS degree in land resource management from Hunan Normal University, Changsha, China, in 2002, and the PhD degree in cartography and geographical information system from Nanjing University, Nanjing, China, in 2007. From 2007 to 2009, he was a lecturer with the Nanjing University. Since 2009, he has been an associate professor with the Nanjing University. His research interests include remote sensing image processing and remote sensing of snow cover.

Xuezhi Feng received the BS degree in cartography from Nanjing University, Nanjing, China, in 1979. From 1980 to 1981, he was a lecturer with the Nanjing University. He was then employed by the Chinese Academy of Sciences, Lanzhou, China. Since 1997, he has been a professor with the Nanjing University. His research interest is remote sensing of snow cover.

Xueliang Zhang received the BS degree in geographical information system from Nanjing University, Nanjing, China, in 2010. Currently, he is working toward the PhD degree in cartography and geographical information system at Nanjing University. His research interests include remote sensing image processing and high-resolution remote sensing image segmentation.

Zuo Wang received the BS degree in geographical information system from Anhui Normal University, Wuhu, China, in 2008, and the MS degree in cartography and geographical information system from Fujian Normal University, Fuzhou, China, in 2011. Currently, he is studying for the PhD degree in cartography and geographical information system at Nanjing University. His research interest is remote sensing of snow cover.

Luyuan Jiang received the BS degree in geographical information system from Northwest University, Xi'an, China, in 2012. Currently, she is a graduate student majored in cartography and geographical information system at Nanjing University. Her research interests include land use and land cover change and high-resolution remote sensing of snow cover.