14. Leng Y. G., Wang T. Y., Guo Y (2005) Stochastic resonance behaviors of bistable systems connected in series. *Acta Physica Sinica*, 54(3), p.p.1118-1124.

15. Zhang Y., Li S. M. (2011) Application of Generalized Stochastic Resonance to the Vibration Test. *Applied Mechanics & Materials*, 141, p.p.21-25.

An Optimal Feature Selection Method for the Classification of Ground Cover in Remote Sensing Images

Yu Yang

School of Automation Science and Electrical Engineering, Beihang University, Beijing, China

Hong Zheng

China Academy of Space Technology (CAST), Beijing, China

Chuanzhao Han

China Academy of Space Technology (CAST), Beijing, China

Abstract

Optimal feature selection is one of major problems associated with the design of classifiers for identifying ground cover in satellite remote sensing images. This paper proposes a novel method of optimization of feature selection, namely, Corrects and Errors Offset Each Other (CEOEO). The proposed method can resolve the close coupling that exists between the feature selection process and the classifier. First, our studies showed that the performance of feature combination mainly relies on the complementarity of the feature between correct and wrong recognition and does not relate to the performance of the classifier. Therefore, we built an optimal model of the feature subset

selection and proved that the model is optimal. Second, we used CEOEO to optimize five ground covers of the feature selection process: ocean, desert, mountain, vegetation, and city in satellite remote sensing images and obtained five optimal feature subsets. Third, we compared the five feature subsets with the feature subsets that were selected by Simulated Annealing (SA) and Sequential Forward Selection (SFS) methods. The advantages of the five feature subsets obtained by CEOEO that they have smaller feature dimensions, higher recognition rates, less computational complexities, and no relation to the performance of classifiers. The results verify the validity and superiority of the feature subsets selected using CEOEO.
Key words: OPTIMAL FEATURE SELECTION, CEOEO, REMOTE SENSING IMAGE

1. Introduction

With the rapid development of satellite remote sensing imaging technologies, remote sensing images are produced with very high resolution and a larger number of spectrums are more accurate in expressing and describing ground cover characteristics. However, selecting the appropriate ground cover features and extracting them using the large amount of available remote sensing data represent a challenge. For the selection of ground cover features in remote sensing images, we aim to find an optimized feature selection method that can reduce the number of features, retain major information on the ground cover, and provide good separability from other types of ground cover [1].Based on some particular separability criteria (classifier), available feature selection methods usually search the optimal feature subsets, and these methods are mainly divided into four categories. The first is the exhaustive search method [2–5], which selects the optimal feature subset by exhaustively searching the entire feature space and evaluating the separability of each possible candidate feature subset [6–8]. The second is the heuristic search method [9–13], such as sequential forward selection, sequential backward elimination which can obtain the classifier that depends on the feature subsets and classification parameters so that, for the subset of complicatedly interactive features, the feature subsets are not necessarily optimal. The third category includes random search methods [14], which set the maximum iteration numbers to limit the complexity of the algorithm through randomly generated feature subsets, thereby determining whether optimal solutions exist; an example is the simulated annealing method. By using the specific classifier as evaluation criterion, the selected results are closely related to the performance of the classifier. The fourth covers intelligent methods, such as neural network methods [15, 16], the SNR-based method [17], the fuzzy entropy method [18], and so forth [19–20], which focus on designing classifiers. These do not indicate whether the selected feature or the feature subsets truly describe the performance of the ground cover.

In summary, for aforementioned methods, the process of feature selection is closely integrated with the classifier accuracy, thus making the selection results mainly dependent on the performance of the classifier, as a result, it is very difficult to prove whether the optimality is because of the performance of the selected feature or the classifiers. Therefore, under acceptable computational complexity, the problem of selecting the optimal feature subsets is not yet completely solved. Targeting the problem of feature extraction of ground cover in remote sensing images, we analyze the object-background separability of each feature, construct a classifier-independent feature subset separability criterion that combines the complementarity and separability of features, uncover the internal principles of feature combination from the perspective of recognition, and finally obtain the optimal feature subset based on performance-complementarity. Therefore, we propose a feature subset selection method CEOEO and compare it with SA and SFS to show its effectiveness.

2. Feature selection issue

Feature selection studies the methods of the description and the separability of the object and background, including single feature selection and an optimal selection of the feature subset. The feature selected or the feature subset can describe the object and distinguish it from the background. We mainly investigate how to select the optimal feature subset.

Assumption: There is a feature set D including D features.

We select feature d of $D\,(d < D)$ to form a subset as small as possible. The subset does not significantly decrease the classification accuracy or influence the class distribution, and it is stable and adaptive as well [21, 22].

J_{ij} represents the separability rule of the i-th and the j-th feature, such as in formula (1):

$$J_{ij} = \begin{cases} 1, & i = j \\ [0,1), & i \neq j \end{cases} \tag{1}$$

J_{ij} meets the following conditions:

Monotonicity,

Additivity, when the features are independent of each other,

Symmetry: $J_{ij} = J_{ji}$.

We assume that the quasi-complete feature set is $\boldsymbol{F}_D = \{f_1, f_2, \ldots, f_D\}$, the selected optimal feature subset is $\boldsymbol{F}_{md} = \{f_{m1}, f_{m2}, \ldots, f_{md}\}$, and the selection rule of the feature subset is (2):

$$J(f_{m1}, f_{m2}, \ldots, f_{md}) = \min\{f_1, f_2, \ldots, f_D\} \qquad (2)$$

Where i is the number of d-dimensional subsets F_d in the D-dimensional set F_D, m indicates the optimal feature subset, and f_{ij} indicates the feature subset only, including two features: the i-th and the j-th feature. $\min\{J(f_{i1}, f_{i2}, \ldots, f_{id})\}$ indicates the feature subset with the best separability in all $\boldsymbol{F}_d$ of $\boldsymbol{F}_D$.

3. Ceoeo feature selection

From the viewpoint of image comprehension, any object based on the image information can be described using specific combinations of features [23], which reflect the intrinsic characteristics of the image object. Among those features, every feature depicts the characteristics of the object in a distinct aspect, while the different combinations of different features express the overall property of the object. Through the method of offsetting the classification errors of different features, the feature set can acquire better classification performance compared to a single feature. In this section, we use set theories to research the method of CEOEO feature subset selection and thus prove that the selected subset is optimal.

3.1. Some Related Concepts

We define some concepts about features or feature subsets based on set theories.

Set of Image Features: The set composed of all features acquired from different physical and mathematical statistics that are able to distinguish the image object from background, $F_D = \{f_1, f_2, \ldots, f_D\}$.

Object-Background Pattern Class: (ω_1, ω_2) expresses the object-background array, in which ω_1 is the object class, and ω_2 is the background class.

Image Sample Set: The original image is divided into block samples with the total number N; the object blocks are N_1, and the background blocks are N_2.

Object block set: $O = \{o_1, o_2, \ldots, o_{N_1}\} \in \omega_1$, where o_i is the i-th block, and $i = 1, 2, \ldots, N_1$.

Background block set: $B = \{b_1, b_2, \ldots, b_{N_2}\} \in \omega_2$, where b_i is the i-th background block, and $i = 1, 2, \ldots, N_2$.

Description of Classification Abilities of Features: A feature f_i can divide one sample set into the following four subsets:

correctly classified object sample subset (correct), denoted as O_{ri};

wrongly classified object sample subset (missed alarm), denoted as O_{ei};

correctly classified background sample subset (correct), denoted as B_{ri}; and

wrongly classified background sample subset (false alarm), denoted as B_{ei};

in which the subscript r indicates the correctly classified samples, and e indicates the wrongly classified samples.

Let f_i be a single-threshold feature; the threshold t separates the object and background, and the object sample set is denoted as:

$$O_{ri} = \{o_{ki} \mid f_i \le t, k = 1, 2, \ldots, N_{ri}^1\}$$

$$O_{ei} = \{o_{ei} \mid f_i > t, k = 1, 2, \ldots, N_{ei}^1\}$$

where $N_{ri}^1 + N_{ei}^1 = N^1$

$$O_{ri} \cup O_{ei} = O$$

$$O_{ri} \cup O_{ei} = \varnothing$$

Similarly, the background sample set is denoted as:

$$B_{ri} = \{b_{ki} \mid f_i > t, k = 1, 2, \ldots, N_{ri}^2\}$$

$$B_{ei} = \{b_{ki} \mid f_i \le t, k = 1, 2, \ldots, N_{ei}^2\}$$

where $N_{ri}^2 + N_{ei}^2 = N^2$

$$B_{ri} \cup B_{ei} = B$$

$$B_{ri} \cap B_{ei} = \varnothing$$

The sample sets are described in Figure 1:

In Figure 1, the blank area represents the correctly classified samples, and the shaded area indicates the wrongly classified samples. Obviously, any feature that has separability to some extent will divide one sample set into four subsets with different meanings: two correct partitions O_{ri} and B_{ri} and two wrong partitions O_{ei} and B_{ei}.

(a) (b)

Figure 1. The sample sets partition of feature f_i. (a) Target sets partition (b) Background sets partition

Feature Subset Operation**:**

Because of the large and inconsistent dynamic ranges of different features, data pre-processing and

normalization are necessary. For all image samples with D features, calculate the values of every f_i to construct a sample vector X_i with N, with the number of training samples as its dimension:

$$X_i = [x_{1i}, x_{2i}, \ldots, x_{Ni}] = [f_i(o_1), \ldots, f_i(o_{N1}), f_i(b_1), \ldots, f_i(b_{N1})]$$

where x_{ji} is the value of f_i extracted from the j-th sample, and i=1,2,…,D, j=1,2,…,N.

Normalize every feature to make their ranges lie in [1, 0]. When the number of samples is finite, every feature value of the object samples and the background samples in vector X_i can be calculated to form the frequency distribution histogram. Then the class conditional probability density function $p(x_i \mid \omega_1)$ with respect to class ω_1 and the $p(x_i \mid \omega_2)$ with respect to ω_2 of feature f_i can be obtained by curve fitting. Let their crossing point t_i be the classification threshold of feature f_i.

Classes Sensitivity: The classes sensitivity of a feature is defined as its ability to describe the object or background classes, calculated by the Shannon entropy. For feature f_i, let H_{i1} be the sensitivity of the object and H_{i2} the sensitivity of the background; then both are given by (3)

$$H_{ij} = -\int p(w_j \mid x_i) \log_2 p(w_j \mid x_i) \qquad (3)$$

in which i=1, 2, …, D; j=1, 2;

$$p(w_j \mid x_i) = \frac{p(x_i \mid w_j) p(w_j)}{\sum_{j=1}^{2} p(x_i \mid w_j) p(w_j)};$$

The posterior probability of f_i is $p(x_i \mid w_j) = \frac{|O_{rj} \cup B_{rj}|}{N_j}$; and the prior probability of f_i is $p(w_j) = \frac{N_j}{N}$.

3.2. Feature Separability Rule

Class separability is defined as the ability of a feature or feature set to distinguish the object and the background. To describe the separability of the feature set, we build a set relationship description model between any feature f_i and f_j based on the sample consistency of their correctly and wrongly classified subsets. Because of its basis in recognition rates, this feature set separability criterion only utilizes recognition results and therefore is totally independent of the physical characteristics, geometric properties, and mathematical expression.

It has the following properties:

Inclusive: For the samples that can be completely recognized by f_i, if they can also be correctly recognized by f_j, i.e., $O_{ri} \subseteq O_{rj}$ and, $B_{ri} \subseteq B_{rj}$ then f_i is included in f_j (or f_j includes f_i), as shown in Figure 2.

Complementarity: The complementarity between two features or feature sets means that the wrongly classified samples of a feature or feature set can be replaced by the correct results of another feature (set), thus making the integrated classification results of the two features or feature sets combined more desirable than their respective results. Complementarity helps to make up for the disadvantages among the features and improves the separability of the combination of features.

Complementarity between two features: The samples that cannot be recognized by feature f_i but can be recognized by f_j or *vice versa* are denoted as the complementary set C_{ij} of features f_i and f_j.

$$C_{ij} = (O_{ei} \cap O_{rj}) \cup (B_{ei} \cap B_{rj}) \cup (O_{ej} \cap O_{ri}) \cup (B_{ej} \cap B_{ri})$$

The size of set C_{ij} implies the potential to improve the separability of f_i and f_j combined.

Complementarity of feature sets: The object and background classification result subsets of feature set $I = \{f_1, f_2, \ldots, f_{k-1}\}$ are as follows.

Wrongly classified target sample subset: $O_{eI} = \bigcap_{i=1}^{k-1} O_{ei}$,

Correctly classified target sample subset: $O_{rI} = O - O_{eI}$

Wrongly classified background sample subset: $B_{eI} = \bigcap_{i=1}^{k-1} B_{ei}$

Correctly classified background sample subset: $B_{rI} = B - B_{eI}$

Then the complementary set C_{Ik} between I and a new feature f_k is given by

$$C_{Ik} = (O_{eI} \cap O_{rk}) \cup (B_{eI} \cap B_{rk}) \cup (O_{ek} \cap O_{rI}) \cup (B_{ek} \cap B_{rI})$$

(a) (b) (c) (d) (e)

Figure 2. Five types of samples (a) Ocean sample (b) Desert sample (c) Mountain sample (d) Vegetation sample (e) City sample

Class Separability Rule of Features

Definition 1: Separability between two features.

For the given features f_i and f_j, the separability rule is defined in (4):

$$J_{ij} = \frac{\left|(O_{ei} \cap O_{ej}) \cup (B_{ei} \cap B_{ej})\right|}{\left|O \cup B\right|} \quad (4)$$

If $i \neq j$, J_{ij} corresponds to the number of wrongly recognized samples.

If all samples can be correctly classified, i.e., the number of wrongly recognized samples is zero, J_{ij}=0.

Definition 2: Separability between a feature set and a single feature.

For a given feature set $I = \{f_1, f_2, \ldots, f_{k-1}\}$ and a single feature f_k, the separability rule is defined in (5):

$$J_{Ik} = \frac{\left|\bigcap_{i=1}^{k} O_{ei} \cup \bigcap_{i=1}^{k} B_{ei}\right|}{\left|O \cup B\right|} \quad (5)$$

As separability rules of features, J_{ij} and J_{Ik} only utilize the classification ability of features (or feature sets) and their complementarities instead of requiring the designing of a classifier to provide classification results. Therefore, the separability rule is decoupled from the classifiers and directly reflects on the classification ability of features or features sets.

3.3. CEOEO Algorithm

The CEOEO selection can be seen as a process that compensates for the classification error of each feature utilizing the complementarities between them. The classification error gradually decreases as the number of features increases. The next selection of features relies on the previous result to a large extent, forming a group of series-dependent actions considering the complementarity between the selected features (or feature set) and the candidate feature, which can only be added when it takes the complementary effect. Then the separability of the new feature subset is measured with the feature separability rule $J(n)$, where n refers to the times of selection. Take complementarity as the standard to select the next feature to combine such that it can further "offset" the existing recognition error, guaranteeing that the newly formed feature subset has better object-background separability.

The steps in CEOEO are as follows:

Step 1: Sample pre-processing. D' features form a quasi-complete space, $F' = \{f_1, f_2, \ldots, f_{D'}\}$.

F is the minimal quasi-complete space, in which the redundant features (that are included in other features) are removed, i.e., $F = F' - \{f_i \mid \exists j, s.t. f_i \subseteq f_j\}$

Training samples: The object samples number N_1 and the background samples number N_2.

The sample vector calculated from feature f_i is:

$$X_i = [x_{1i}, x_{2i}, \ldots, x_{Ni}] = [f_i(o_1), \ldots, f_i(o_{N_1}), f_i(b_1), \ldots, f_i(b_{N_2})]$$

where x_{ji} is the j-th sample's feature value of f_i; i=1, 2, …, D; and j=1, 2 , …, N.

In the case of finite samples, calculate the feature value of the object samples and the background samples in each vector X_i to obtain its frequency distribution histogram statistically. Then acquire the class conditional probability density function $p(x_i|\omega_1)$ with respect to class w_1 and the $p(x_i|\omega_2)$ with respect to ω_2 of feature f_i by curve fitting. Set their crossing point t as the classification threshold of feature f_i.

$$\forall f_i,\ O_{ri} = \{o_{ik} \mid f_i(o_k) \le t_i\},\ O_{ei} = \{o_{ik} \mid f_i(o_k) > t_i\},$$
$$k = 1,2,\ldots,N_1$$

$$B_{ri} = \{b_{ik} \mid f_i(b_k) > t_i\},\ B_{ei} = \{o_{ik} \mid f_i(o_k) \le t_i\},$$
$$k = 1,2,\ldots,N_2$$

$$J_i = \frac{\left|O_{ei} \cup B_{ei}\right|}{\left|O \cup B\right|} \quad i=1,2,\ldots,D$$

Let $J(1)=\min\{J_i\}$ and $i=1,2,\ldots,D$.

$J(1)$ corresponds to the first selected feature with maximal separability f_{m1}; reorder $\{f_i\}$ into $\{f_{mi}\}$ such that $J_{m1} < J_{m2} < \ldots < J_{mD}$. j=1, and choose the first feature f_{m1}, $F_{D-1} = F - f_{m1}$.

Step 2: Given j>1, without loss of generality, the selected feature set is $I_j = \{f_1, f_2, \ldots, f_j\}$, leaving the unselected features to form the set $F_{D-j} = F - I_j = \{f_{j+1}, f_{j+2}, \ldots, f_D\}$, with the dimensions of D-j. According to the property of the complete partition of features, the lower boundary of C_{1_j} is calculated as:

$$\delta_j = \frac{N}{(D-j) \times 2^{(D-j)}}.$$

Calculate the partition of the sample set caused by I_j:

$$O = O_{rI_j} \cup O_{eI_j}, B = B_{rI_j} \cup B_{eI_j}.$$

If $O_{eI_j} \cup B_{eI_j} = \varnothing$, stop; otherwise, calculate δ_j.

Step 3: Among all selectable features in F_{D-j}, choose the feature that has the best complementarity to I_j as the $(j+1)$-th feature.

If C_{mI}=max$\{C_{iI}\}$ and $i=j+1,j+2,\ldots,D$, then choose C_{mI} as the corresponding f_{mI}.

The concrete operations are as follows:

(a) $j > 1$, $\forall f_{jk} \in F_{D-j}$, and $k = j+1, j+2, \ldots, D$.

$$C_{Ik} = (O_{eI} \cap O_{rk}) \cup (B_{eI} \cap B_{rk}) \cup (O_{ek} \cap O_{rI}) \cup (B_{ek} \cap B_{rI})$$

(b) Delete the features that have no classification ability in F_{D-j}:

$$F'_{D-j-1} = F_{D-j} - \{f_{jk} \mid C_{Ik} < \delta_j, k = j+1, j+2, \ldots, D\}$$

(c) Count the number of features that have the largest complementary sets to I_j, denoted as k:

If $k \geq 2$, go to (d); otherwise,

choose $f_{m(j+1)}$ as the $(j+1)$-th feature, and go to (e).

(d) Calculate the class sensitivities H_{m1} and H_{m2} of the k features that have the same complementary set to I_j, where $m = 1, 2, \ldots, k$.

If $|O_{rI}| > |B_{rI}|$, then choose the feature that has the biggest target sensitivity max(H_{m1}) as the $(j+1)$-th feature $f_{m(j+1)}$; otherwise, choose the feature that has the biggest background sensitivity max(H_{m2}) as the $(j+1)$-th feature $f_{m(j+1)}$.

(e) Let $I'_{j+1} = \{f_{m1}, f_{m2}, \ldots, f_{m(j+1)}\}$; thus $F_{D-(j+1)} = \{f_{j+2}, f_{j+3}, \ldots, f_D\}$.

(f) Choose I''_{j+1}, which satisfies:

$$J_{I''_{j+1}} = \min\left(J_{I'''_{j+1}}\right)$$

$$I'''_{j+1} = \left\{\begin{array}{l}\{I'_{j+1} - f_{ma}\} \cup f_{mb} \mid \forall f_a \in I_j, \forall f_b \in F_{D-(j+1)} \\ \text{and } f_b \neq f_{m(j+1)}\end{array}\right\}$$

Compare the separabilities of I'_{j+1} and I''_{j+1}.

If $J_{I'_{j+1}} < J_{I''_{j+1}}$, let $I_{j+1} = I'_{j+1}$;

otherwise, let $I_{j+1} = I''_{j+1}$.

$F_{D-(j+1)} = F - I_{j+1}$.

(g) Reorder $\{f_i\}$ in I_{j+1} into $\{f_{mi}\}$ such that $J_{m1} < J_{m2} < J_{m3} < \ldots < J_{m(j+1)}$.

If $F_{D-j-1} = \varnothing$, go to Step 4; otherwise,

$j = j+1$.

If $j = D$, go to Step 4; otherwise, go to Step 2.

Step 4: Let $d = j$ and stop.

The selected feature subset is:

$I_d = \{f_{m1}, f_{m2}, \ldots, f_{md}\}$.

4. Ceoeo set of Five Ground Covers in Remote Sensing Images

This section computes the optimal ground cover subsets selected with CEOEO, SA, and SFS and demonstrates the validity of CEOEO by comparing the size, object-background separability, and recognition rate of the selected feature subsets.

4.1. Feature Set of Remote Sensing Images

The information on the ground cover in remote sensing images can be expressed in their shape, hue, texture, location, and so forth. We take the 19 features in Table1to describe the image information, which form the candidate feature set $F = \{f_1, f_2, \ldots\ldots, f_{19}\}$.

4.2. Samples of Remote Sensing Images

We chose five typical ground cover categories detected in EO-1: ocean, desert, mountain, vegetation, and city. Then we selected the feature subsets of the different types of ground cover.

The images of five types of ground cover were divided into blocks with the processing scale of 64×64 pixel size as the training samples. Each sample was manually marked as object (positive sample) or background (negative sample). Every ground cover category contains 400 samples, including 200 positive samples and 200 negative samples. When training a type of ground cover, other types were treated as negative samples. For example, when testing ocean samples, desert\vegetation\city\mountain types were treated as negative samples. Training and testing samples, respectively, accounted for 50% of all the samples. Figure 2. (a)–(e) illustrates some examples of the five types of samples.

4.3. CEOEO Selection Results

With all the features defined in Table1, each feature value of the five types of samples was calculated to obtain 5×19 groups of 400-dimensional sample arrays. Then each feature space formed by these sample data was normalized. The optimal feature subsets of

Table 1. Candidate feature set of remote sensing images

Index	Feature	Index	Feature
1	f_1: first momentum	11	f_{11}: first momentum of hue
2	f_2: second momentum	12	f_{12}: second momentum of hue
3	f_3: third momentum	13	f_{13}: third momentum of hue
4	f_4: angular second momentum	14	f_{14}: average length of chain code
5	f_5: fractal dimension	15	f_{15}: maximum length of chain code
6	f_6: entropy	16	f_{16}: total length of chain code
7	f_7: maximum of FFT	17	f_{17}: total number of chain code
8	f_8: Euler number	18	f_{18}: sum of absolute value of curvature
9	f_9: average of skeleton	19	f_{19}: average of absolute value of curvature
10	f_{10}: maximum of hue histogram		

the five types of ground cover and their recognition training results were obtained by using the CEOEO algorithm.

4.4. Comparison with Other Methods

With the same training samples, we utilized the SA and SFS to obtain the selected feature subset ISA and ISFS, respectively, as listed in Tables 2 and 3.

Both the SA and SFS methods require setting the dimensions of the selected feature sets manually, while the CEOEO method automatically determines the dimensions of the feature sets. In the SA and SFS methods, the dimensions of the selected feature sets are calculated before setting the dimension d=1, 2, 3.

From Table 2, it can be observed that SA has different recognition rates for each type at different *d*, of which the results with the best performance are:

$I_{SA\_ocean}=\{10, 4\}$
$I_{SA\_desert}=\{6, 13, 11\}$
$I_{SA\_mountain}=\{4, 13\}$
$I_{SA\_vegetation}=\{4, 1, 7\}$
$I_{SA\_city}=\{11\}$

We chose these as the final results for comparison with I_{CEOEO}.

From Table 3, it can be observed that SFS has different recognition rates for each type at different *d*, of which the results with the best performance are:

$I_{SFS\_ocean}=\{10,4,17\}$
$I_{SFS\_desert}=\{12,16,11\}$
$I_{SFS\_mountain}=\{4,13\}$
$I_{SFS\_vegetation}=\{7,1,4\}$
$I_{SFS\_city}=\{11\}$

We chose these as the final results for comparison with I_{CEOEO}. The results of the three methods are compared in Table 4.

Table 4 shows that CEOEO has the best separability, which completely distinguishes the 5 types of

Table 2. Feature set selected by I_{SA} (iterated 20000 times at each temperature)

Category	Feature Subset	d	Missed Alarm Rate	False Alarm Rate	Recognition Rate
Ocean	10	1	0.00%	37.00%	63.00%
Desert	12		1.00%	45.75%	53.25%
Mountain	4		23.00%	24.50%	52.5%
Vegetation	7		0.25%	4.50%	95.25%
City	11		0.00%	0.00%	100.00%
Ocean	10, 4	2	0.00%	6.75%	93.25%
Desert	13, 16		0.00%	16.00%	84.00%
Mountain	4, 13		0.00%	24.50%	75.50%
Vegetation	7, 1		0.25%	4.50%	95.25%
City	11, 6		0.00%	0.00%	100.00%
Ocean	4, 3, 13	3	2.00%	18.50%	79.50%
Desert	6, 13, 11		0.00%	2.75%	97.25%
Mountain	4, 3, 13		0.00%	24.50%	75.50%
Vegetation	4, 1, 7		0.00%	4.50%	95.50%
City	13, 11, 6		0.00%	0.00%	100.00%

Table 3. Feature set selected by I_{SFS}

Category	Feature Subset	Missed Alarm Rate%	False Alarm Rate%	Recognition Rate%
Ocean	10	0.00%	37.00%	63.00%
Desert	12	1.00%	45.75%	53.25%
Mountain	4	23.00%	24.50%	52.5%
Vegetation	7	0.25%	4.50%	95.25%
City	11	0.00%	0.00%	100.00%
Ocean	10, 4	0.00%	6.75%	93.25%
Desert	12, 16	0.50%	14.25%	85.75%
Mountain	4, 13	0.00%	24.50%	75.50%
Vegetation	7, 1	0.25%	4.50%	95.25%
City	11, 6	0.00%	0.00%	100.00%
Ocean	10, 4, 17	0.00%	0.00%	100.00%
Desert	12, 16, 11	0.50%	1.50%	98.00%
Mountain	4, 13, 3	0.00%	24.50%	75.50%
Vegetation	7, 1, 4	0.00%	4.50%	95.50%
City	11, 6, 13	0.00%	0.00%	100.00%

Table 4. Comparison of the performance of the feature subsets selected by CEOEO, SA, and SFS

Category	Feature Subset	Feature Number	Missed Alarm Rate	False Alarm Rate	Recognition Rate	Time (ms)
Ocean	6,11	2	0.00%	0.00%	100.00%	0.16
	4,10	2	0.00%	6.75%	93.25%	105.34
	4,10,17	3	0.00%	0.00%	100.00%	0.02
Desert	4,11	2	0.00%	0.00%	100.00%	0.42
	6,11,13	3	0.00%	2.75%	97.25%	127.20
	11,12,16	3	0.50%	1.50%	98.00%	0.02
Mountain	1,4,6	3	0.00%	0.00%	100.00%	0.17
	4,13	2	0.00%	24.50%	75.50%	105.34
	4,13	2	0.00%	24.50%	75.50%	0.02
Vegetation	1,19	2	0.00%	0.00%	100.00%	0.13
	1,4,7	3	0.00%	4.50%	95.50%	127.20
	11,6,13	3	0.00%	4.50%	95.50%	0.02
City	11	1	0.00%	0.00%	100.00%	0.19
	11	1	0.00%	0.00%	100.00%	120.11
	11	1	0.00%	0.00%	100.00%	0.02

ground cover from their backgrounds with a recognition rate of 100%. Except for the city type, SA cannot completely distinguish the other four earth surfaces from their backgrounds, with the desert type having the highest recognition rate of 97.25%. Except for the city and ocean types, SFS cannot completely distinguish the other three earth surfaces from their backgrounds. Considering the feature subset dimensions, although SFS has complete separability for the ocean type, its subset dimensions are twice that of the feature subset selected with CEOEO. In addition, SA and SFS both require setting *d*, the value of the dimensions of the selected feature sets, beforehand so as to diminish the degree of automation of the algorithm. Therefore, considering the optimality of the selected feature set and algorithm automation, CEOEO is obviously advantageous.

Considering the computing time, SA is the slowest, with the computation complexity $O(d^2t(D))$; SFS has the computation complexity $O(D\log_2 D)$, and CEOEO has $O(t(D))$, where $t(D)$ is a polynomial of D.

Because of the complexity of the background of remote sensing images and the drastic variance of illumination, it is difficult to precisely describe the ground cover with features selected by conventional methods. The test results show that CEOEO's computation complexity is identical to that of sequential methods and is lesser than that of the other two methods; additionally, it achieves the best classification results and highest recognition rates with the least dimensions of feature subsets.

The feature subsets of each type of ground cover selected by CEOEO, SA, and SFS are respectively illustrated in Figure 3. In the figures, P represents the target sample with "*", and N represents the background samples with "Δ".

Figure 3. (e-l) show that the SA and SFS feature selection applies the Fisher classification rule, which minimizes the intra-class distance. Therefore, the selected feature will be affected by the performance of the classifier, as shown in Figure 3, in which the object samples are highly concentrated. This method, however, ensures the optimal classification performance of the feature subsets. In contrast, CEOEO takes the complementarities and the minimal number of wrongly classified samples of the features as the classification standard, independent of any specific classification rules associated with the classifier, such that the classifier-independent optimal feature subset is selected within a relatively short computing period. This avoids cases in which the feature selection results are closely related to the performance of the classifier, which means that the feature subsets are not "really" optimal. A typical example is the fourth kind of feature selection algorithm mentioned in the Introduction, which utilizes complicated classifiers to select features and possibly causes the selected features to reflect the performance of the classifiers, whereas the classification ability of the feature subset *per se* is not always optimal.

5. Conclusion

This study investigates the feature selection of five types of ground cover in remote sensing images and objects the problem of the current selection methods, which are too closely coupled with the classifiers in selecting an effective feature subset. A new method called CEOEO is proposed as an alternative. CEOEO uses the complementarity of errors and the minimization of classification errors between features. The

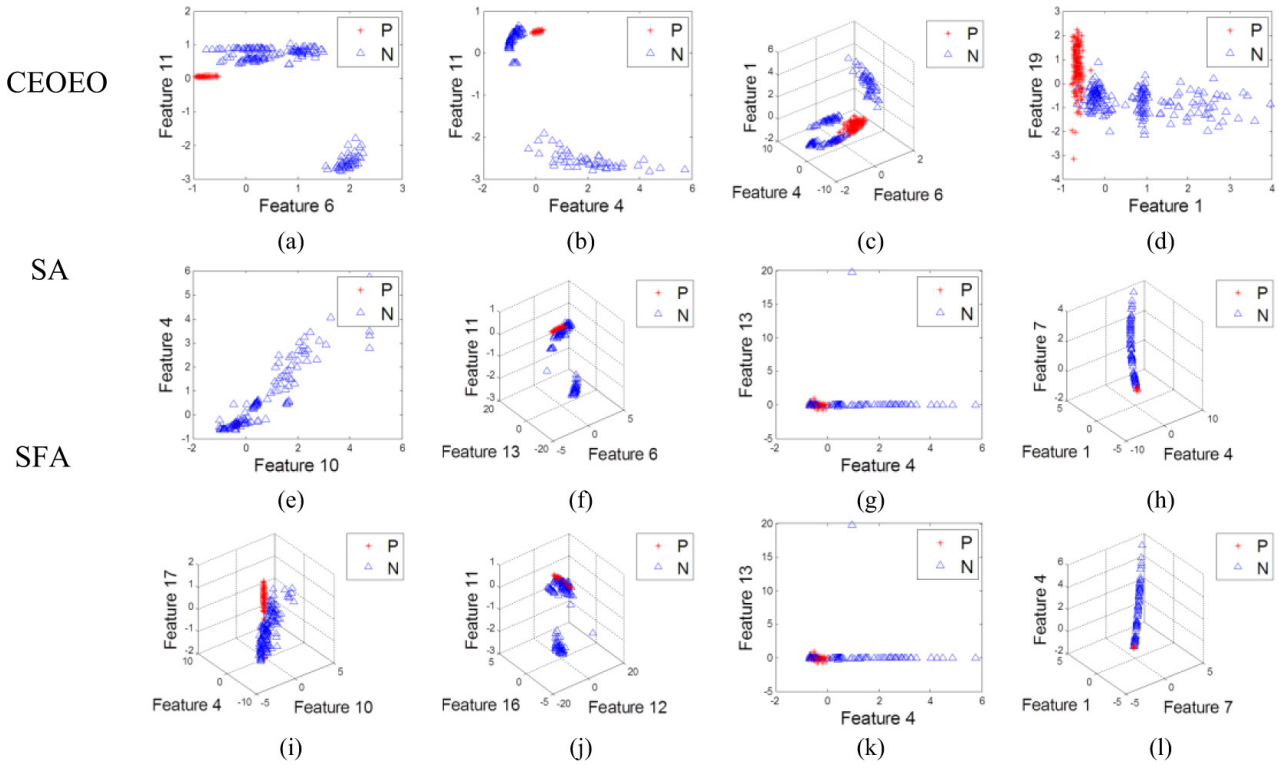

Figure 3. Feature subsets selected by CEOEO,SA and SFA

inclusion, complementarity, and separability are defined by using set theories, and the optimality of the feature subset selected by this separability criterion is proven.

Compared with the SA and SFS feature selection methods, CEOEO can automatically find the dimension d of the optimal feature subset by increasing the rationality of the feature selection. The relation of the CEOEO separability criterion to the selection sequence, which not only considers the relation between features but also the complementarity and inclusion of feature subsets, is more effective at simplifying the computation and eliminating feature redundancy. This avoids the disadvantage of the current methods, which do not consider the influences of the added (or deleted) features and require a "black box" calculation from the beginning. More importantly, the current methods process feature selection through complex classifiers, whose performance greatly affects the selection results, making the selected feature subsets highly dependent on the performance of the classifiers. In addition, the advantage of the performance of the classifiers may overshadow the selected feature subsets, whose optimality in object-background separability is therefore not ensured. Consequently, when applied with other classifiers, these features may have unexpected results. Meanwhile, compared with the test results for SA and SFS, CEOEO shows a better recognition rate and lower computation complexity.

It is worth mentioning that, being independent of recognizing concrete contents, the CEOEO algorithm is applicable not only in the feature selection of ground cover in remote sensing images but also in the recognition of other objects in remote sensing images and ordinary images.

Acknowledgements

The authors would like to thank Prof. Zheng for providing high-quality aerial photographs and her feedback and advice.

References

1. Cao Qiong (2008) *Research on Intelligent Evaluation Method of Availability of Satellite Remote Sensing Images*. Doctoral dissertation, Beihang University.
2. Siedleeki W., Sklansky J. (1988) On automatic feature selection. *International Journal of Pattern Recognition*, 2(2), p.p. 197–220.
3. Doak J. (1992) *An evaluation of feature selection methods and their application to computer security*. Technical Report CSE-92-18, Department of Computer Science, University of California, Davis, CA.
4. Dash M., Liu H. (1997) Feature selection for classification. *Intelligent Data Analysis*, 1(S1-4), p.p. 131-156.

5. Almuallim H., Dietterich T. (1994) Learning Boolean concepts in the Presence of many irrelevant features, *Artificial Intelligence*, 69(S1-2), p.p. 279-305.
6. Sheinvald J., Dom B., Niblae K. W. (1990) A modeling approach to feature selection. *Proceedings of the Tenth International Conference on Pattern Recognition*, pp. 535–539.
7. Rissanen J. (1978) Modeling by the shortest data description. *Automatica*, 14(5), pp. 465-471.
8. Cormen T., Leiserson C., Rivest R. (1990) *Introduction to Algorithms*. MIT Press, Cambridge, MA.
9. Russell S., Norvig P. (1995) *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs, NJ.
10. Motvani R., Raghavan P. (1996) Randomized algorithms, *ACM Computing Surveys*, vol. 28(l), pp. 33–37.
11. Milan Sonka, Vaclav Hlavac, Roger Boyle (2014) *Image Processing, Analysis, and Machine Vision*. CL Engineering.
12. Bishop C.M. (2012) *Pattern Recognition and Machine*, New York: Verlag New York Inc.
13. Setiono R., Liu H. (1997) Neural to network feature selector. *IEEE Trans. Neural Networks*, 8(3), p.p. 654-662.
14. Bauer K. W., Alsing S.G., Greene K.A. (2000) Feature screening using signal-to-noise ratios. *Neural computing*, vol. 31(1), p.p. 29-44.
15. Motvani R., Raghavan P. (1996) Randomized algorithms. *ACM Computing Surveys*, 28(l), p.p. 33-37.
16. Setiono R., Liu H. (1997) Neural to network feature selector. *IEEE Trans. Neural Networks*, 8(3), p.p. 654-662.
17. De R. K., Palnr N. R., Pal S. K. (1997) Feature analysis: neural network and fuzzy set theoretic approaches, *Pattern Recognition*, 30(10), p.p. 1579-1590.
18. Qiang Cheng, Hongbo Zhou, Jie Cheng (2011) The Fisher-Markov Selector: Fast Selecting Maximally Separable Feature Subset for Multiclass Classification with Applications to High-Dimensional Data, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 33(6), p.p. 1217-1233.
19. Pal S. K., De R. K., Basak J. (2000) Unsupervised feature evaluation: a neuron-fuzzy approach, *IEEE Trans. Neural Networks*, 11(2), p.p. 366-376.
20. Pradipta Maji, Sankar K. Pal. (2010) Feature Selection Using f-Information Measures in Fuzzy Approximation Spaces. *IEEE Trans. Knowledge and Data Engineering*, 22(6), p.p. 854-867.
21. Batiti R. (1994) Using mutual information for selecting features in supervised neural network learning. *IEEE Trans. Neural Networks*, 5(4), p.p. 537-550.
22. Brunzell H., Erikcson J. (2000) Feature reduction for classification of multidimensional data. *Pattern Recognition*, 33(10), p.p. 1741-1748.
23. Young T.Y., Fu K. S. (1986) *Handbook of Pattern Recognitions and Image Processing*. College of Engineering, University of Miami, Coral Gables, Florida.
24. Sonak M., Hlavac V. (1999) Image Processing, Analysis and Machine Vision. *Thomson Learning*, 14(82), p.p.685–686.