Optik 124 (2013) 4697–4706

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Mean shift based clustering of neutrosophic domain for unsupervised constructions detection

Bo Yu [a,b], Zheng Niu [a,*], Li Wang [a]

[a] The State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China
[b] Graduate University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 6 September 2012
Accepted 21 January 2013

Key words:
Neutrosophic set
Mean shift
Constructions extraction
Image segmentation

ABSTRACT

Automation has been a hot issue in constructions extraction, but there has not yet been a universally accepted algorithm. Commonly, constructions are extracted by user-defined thresholds, and they have to be adjusted with the variation of images and types of constructions. To overcome the shortages, an unsupervised algorithm to extract constructions is proposed in this paper. It adopts mean shift clustering in neutrosophic set domain to segment images, which makes it possible to detect constructions with a stable threshold. The algorithm is compared with three welcomed and recently developed supervised techniques by six study images with two sorts of resolutions. Experiments show that among the four algorithms, the method proposed in this paper performs best in constructions detection. It not only maintains the original shape of buildings, but also generates extracted constructions as a neat whole. Furthermore, the new method has stronger robustness when faced with images with different resolutions and imaging qualities. As tests show that the new algorithm can reach a kappa coefficient of 0.7704 and an accuracy of 89.8054%, which are relatively high in constructions extraction, it can be a robust unsupervised technique to extract constructions.

1. Introduction

1.1. Constructions detection

Constructions detection is playing a significant role in urban planning and monitoring development of an area. Apart from that, detecting constructions contributes to exploring problems of scene segmentation, 3D recovery, and shape descriptions in a rich, realistic, and demanding environment [1].

For remotely sensed imaging, it is pretty difficult to extract constructions because it is made up of pixels which only describe simple topological adjacency rather than real-world objects [2]. Segmentation is a way to turn numerous pixels into various meaningful objects with more informative attributes, such as shapes, length, textures and contexual information [3]. A good deal of algorithms segmenting images have been proposed and they can generally be grouped into three categories, pixel-based category, edge-based category and region-based category. For pixel-based method [4], it is the conceptually simplest way to segment images [5]. Pixels are divided into different groups by thresholding. Always, the thresholds have to be adjusted every time to meet demands. As for edge-based approach [6,7], the most significant matter is edge detection. It segments images by edges detected, but the edges are fragmentized quite often. Therefore, edge linking also has to be discussed after detection. Region-based algorithm [8] concentrates on region growing and region merging. It takes more details of real-world objects into consideration, but the edges of segmentation results are discontinuous, and the regions merged are dispersive rather than a whole.

In this paper, we propose an unsupervised method synthesizing neutrosophic set and mean shift, and it is called NS-MS for short in this paper. It can not only detect constructions directly but also maintain the original shape of them. Mean shift clustering is done to the image which has been transformed to neutrosophic set domain. The segmented image can be used to extract constructions with spectral information unsupervisely rather than with textural and contour information supervisely like the one generated by region-based, edge-based or pixel-based technology.

1.2. Neutrosophic set

Neutrosophic set is a new concept in image segmentation. It is proposed by Smarandache [9] as extension of the fuzzy logic and has been used in philosophy, financial analysis [10,11] and semantic web services [12] widely. To our knowledge, neutrosophic set was first introduced to image processing by Guo and Cheng [21], and it

* Corresponding author.
E-mail addresses: yubo.zuzu@gmail.com, niuz@irsa.ac.cn (Z. Niu).

0030-4026/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.01.117

has been developing very fast in algorithms development in image processing, such as image segmentation [23], image classification [13] and image thresholding [34].

In neutrosophic set domain, there are three factors considered, indeterminate, true and false elements, rather than the two factors considered in fuzzy logic which include true and false ones only.

According to Smarandache, neutrosophic set is a domain which image can be transformed to. Neutrosophic set can be expressed as N, and it contains three sub-sets: T, representing true set which comprises all the true elements; I, standing for indeterminate set which includes all the indeterminate elements and F, the false set, consisting of every false element. In general, $N = \{(T, I, F): T, I, F\ [0,1]\}$, $T = \{t: t \in T\}$, $I = \{i: i \in I\}$, $F = \{f: f \in F\}$. Based on the concept of neutrosophic set, we can judge the degree of a sentence p being true or not with the formulation $v(p) = (t,i,f)$ [14] where t, i and f represent true degree, indeterminate degree and false degree respectively.

With introduction of neutrosophic set to image processing, many effective algorithms have adopted the concept to get applied. M. Zhang and L. Zhang [15] have proposed an approach with neutrosophic set based on watershed method. With their definitions of formulations mapped to neutrosophic domain, it gets stronger in resisting noise than the traditional pixel-based, edge-based, region-based and two watershed ones [15]. Cheng and Guo [16] improved a method to resist noise of an image with the neutrosophic set by a new filtering procedure to decrease the indeterminate degree which is expressed by entropy. Guo and Cheng [17] applied the theory into image segmentation with a clustering method. It has more stable and effective performance compared with the modified fuzzy C-means (MFCM) segmentation algorithm [18]. However, it can only deal with gray images and the parameters have to be defined manually, rather than automatically. To overcome that shortage, Sengur and Guo [19] applied the neutrosophic set theory into wavelet transformation theory, and it not only works automatically but also segments images into more intact details than new existing methods.

1.3. Mean shift

Mean shift is a nonparametric kernel density estimation technique, and it is based on Parzen window method to find the maximum of kernel density [20]. Recent achievements in mean shift have made it increasingly popular in image segmentation and computer vision. Park et al. have proposed an algorithm which combines adaptive mean shift with statistical theory [21]. By implying statistics into mean shift, it automates to detect optimal clustering number of mean shift, which frees mean shift clustering to be a 'one-step' algorithm. Dorin et al. have done much research in bandwidth selection and scale selection for mean shift [22,23]. Segmented results turn to be more continuous and real-world objects shaped than the images segmented by mean shift with fixed bandwidth and scale.

This paper demonstrates how our segmentation algorithm works and the performance of it. It is organized as follows: in the next two sections, we introduce neutrosophic set and mean shift respectively. As for the fourth section, our algorithm NS-MS is introduced. Experiments and discussion are in the fifth section. Conclusions are presented in section six.

2. Neutrosophic set

Neutrosophic set consists of three components: true set, indeterminate set and false set which are expressed by T, I and F separately. Moreover, T, I and F all belong to [0, 1]. The elements t, i and f are subsets of T, I and F respectively. In neutrosophic logic, we can describe a sentence with the formula $v(p) = (t,i,f)$, which means the sentence is t percent true, i percent uncertain t and f percent false. A pixel P (i,j) can be represented by $P_{NS}(t(i,j), i(i,j), f(i,j))$ after transformed from color space domain to neutrosophic domain, where $t(i,j)$, $i(i,j)$, $f(i,j)$ are the elements of T, I and F, respectively.

2.1. Transformation

According to neutrosophic theory that neutrosophic set is a combination of fuzzy logic and 'Indefinite' fuzzy logic [24,25], and the three factors are influenced by each other. We improve the part of transformation algorithm mentioned in Sengur and Guo [19], and they are defined as below:

$$t(i,j) = \frac{\overline{g(i,j)} - \overline{g_{\min}}}{\overline{g_{\max}} - \overline{g_{\min}}} \tag{1}$$

$$i(i,j) = \frac{\delta(i,j) = \delta_{\min}}{\delta_{\max} - \delta_{\min}} \tag{2}$$

$$f(i,j) = 1 - t(i,j) - i(i,j) \tag{3}$$

$$\overline{g(i,j)} = \frac{1}{w \times w} \sum_{m=i+(w/2)}^{m=i+(w/2)} \sum_{n=j+(w/2)}^{n=j+(w/2)} g(m,n) \tag{4}$$

$$\delta(i,j) = abs(g(i,j) - \overline{g(i,j)}) \tag{5}$$

$g(i,j)$ is the gray scale value of pixel $P(i,j)$, and $\overline{g(i,j)}$ is the local mean value of pixel P(i,j) when processed by a kernel with width of w.

2.2. Enhancement operation

When images are transformed to the neutrosophic set domain, it is divided into three sets, T, I and F. T set is what we need for further procession, but enhancement operation is necessary to enhance the differences among the values of elements in T. We adopt the idea put forward by Li et al. [27]

$$P'_{NS}(\beta) = P(t'(\beta), i'(\beta), f'(\beta)) \tag{6}$$

$$t'(\beta) = \begin{cases} t(i,j) & i(i,j) < \beta \\ t'_{\lambda}, & i(i,j) \geq \beta \end{cases} \tag{7}$$

$$t'_{\beta}(i,j) = \begin{cases} t^2(i,j)/\beta, & 0 \leq t(i,j) \leq \beta \\ 1 - (1 - t(i,j))^2/(1-\beta), & \beta \leq t(i,j) \leq 1 \end{cases} \tag{8}$$

$$f'(\beta) = \begin{cases} f(i,j) & i(i,j) < \beta \\ f'_{\lambda}, & i(i,j) \geq \beta \end{cases} \tag{9}$$

$$f'_{\beta}(i,j) = \begin{cases} f^2(i,j)/\beta, & 0 \leq f(i,j) \leq \beta \\ 1 - (1 - f(i,j))^2/(1-\beta), & \beta \leq f(i,j) \leq 1 \end{cases} \tag{10}$$

$$i'_{\beta}(i,j) = 1 - t'_{\beta}(i,j) - f'_{\beta}(i,j) \tag{11}$$

β is the parameter self-determinated by entropy of the image. Since entropy is used to evaluate the distribution of pixels in the image, it is defined as below:

$$EnI = -\sum_{i=1}^{h} \sum_{j=1}^{w} i(i,j) \log_2 i(i,j) \tag{12}$$

$$EnT = -\sum_{i=1}^{h}\sum_{j=1}^{w} t(i,j)\log_2 t(i,j) \tag{13}$$

$$EnF = -\sum_{i=1}^{h}\sum_{j=1}^{w} f(i,j)\log_2 f(i,j) \tag{14}$$

$$En = EnI + EnT + EnF \tag{15}$$

En represents entropy of the image in the neutrosophic set domain, and it is the summary of entropies of *T*, *I* and *F*. *EnI*, *EnT* and *EnF* are the entropies of subset *I*, *T* and *F* respectively, and

$$\beta = 0.99 - 0.99 \times \frac{EnI - En_{\min}}{En_{\max} - En_{\min}} \tag{16}$$

$$En_{\max} = -\log_2 \frac{1}{hw} \tag{17}$$

h and *w* represents height and width of the image. Enhancement operation will be kept doing until *EnI* changes little.

3. Mean shift algorithm

Given n random points X_i, $i = 1, 2, \ldots, n$ which are of d-dimensional Euclidean space, the multivariate kernel density estimate at point *X* can be defined as follows [17]:

$$f(\hat{x}) = \frac{1}{n}\sum_{r=1}^{n} K_H(X - X_t) \tag{18}$$

where

$$K_H(X) = |H|^{-1/2} K(H^{-1/2}X) \tag{19}$$

and *H* is a symmetric positive-definite $d \times d$ matrix. The definition of *H* has been discussed by Dorin [22]. *K*(*X*) is a symmetric kernel and it satisfies

$$K(X) = c_{k,d}\, k(\left\|X\right\|^2) \tag{20}$$

The normalized constant $c_{k,d}$ is strictly positive, which makes *K*(*X*) integrate to one, while profile of kernel *k*(*x*) can be tenable on condition that $X \geq 0$. To simplize the algorithm, H has been defined as $H = h^2 I$. *I* is an identity matrix and h is bandwidth, meanwhile, formula (18) can be rewritten as

$$f_{h,K}(\hat{X}) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n} k\left[\left\|\frac{X - X_i}{h}\right\|^2\right] \tag{21}$$

The gradient of the kernel density estimate can be

$$\nabla\hat{f}_{h,K}(X) = \frac{2c_{k,d}}{nh^{d+2}}\sum_{i=1}^{n}(X - X_i)k'\left[\left\|\frac{X - X_i}{h}\right\|^2\right] \tag{22}$$

To simplify the expressions of (21), two new functions are defined:

$$g(X) = -K'(X) \tag{23}$$

$$G(X) = c_{g,d}\, g(\left\|X\right\|^2) \tag{24}$$

For formula (22), $k'(X)$ exists in most cases when $X \geq 0$ [21]. As for (6), $c_{g,d}$ is a normalized constant, then the gradient of the kernel density expressed by (21) can be rewritten as

$$\begin{aligned}\hat{\nabla} f_{h,K(X)} &= \frac{2c_{k,d}}{nh^{d+2}}\sum_{i=1}^{n}(x_i - x)g\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \\ &= \frac{2c_{k,d}}{nh^{d+2}}\left[\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)\right]\left[\frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x\right]\end{aligned} \tag{25}$$

Here, mean shift vector is defined as the second term

$$m_{h,G} = \frac{\sum_{i=1}^{n} X_i g\left(\left\|\frac{X-X_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{X-X_i}{h}\right\|^2\right)} - X \tag{26}$$

Synthesize the six formulas from (21) to (26), mean shift vector can be reexpressed as

$$m_{h,G}(X) = \frac{h^2}{2C}\frac{\nabla\hat{f}_K(X)}{\hat{f}_G(X)} \tag{27}$$

It can be seen from (27) that mean shift vector of *X* obtained with kernel *G* is proportional to the normalized gradient of the kernel density yielded with kernel *K*. $m_{h,G}(X)$ goes toward the trend of maximum increase of the density. The point with maximum increase is the one where $\nabla\hat{f}_{h,K}(X) = 0$. Therefore, we get

$$y_{j+1} = \frac{\sum_{i=1}^{n} X_i g\left(\left\|\frac{y_j - X_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{y_j - X_i}{h}\right\|^2\right)}, \quad j = 1, 2, \ldots \tag{28}$$

It is expressed as weighted average of y_j, computed with kernel *G*. y_1 is the original position of kernel *G*. Based on (26) and (28), we can safely come to an iteration

$$m_{h,G}(y_j) = y_{j+1} - y_j \tag{29}$$

until

$$m_{h,G}(y_c) = y_c - y_c = 0 \tag{30}$$

and that is when the gradient of kernel density equals to zero, y_c is the final result of mean shift procession.

4. NS-MS algorithm

Based on research process of Sengur and Guo [19] and the characteristics about $L^*u^*v^*$ (L^* means light intensity, $u^*$and v^* represent aberration separately) that it is better in separating different pixels in accordance with spectral character. Secondly, decomposite image to three channels. Thirdly, the three channels are transformed to neutrosophic set domain respectively. Fourthly, obtain parameter by computing entropies of the image and enhance image in neutrosophic set domain. Later, merge true subsets of three channels into one set. Then, mean shift operation is done on the merged true set and during the process of Mean Shift, kernel function is chosen to be

$$K_{h_s,h_r}(X) = \frac{C}{h_s^2 h_r^p} k\left[\left\|\frac{X^s}{h_s}\right\|^2\right] k\left[\left\|\frac{X^r}{h_r}\right\|^2\right] \tag{31}$$

where *P* is color dimension of the image and spacial dimension is two. X^s is a vector of spacial characteristic, and X^r is color characteristic vector. h_s and h_r are the spatial bandwidth and color

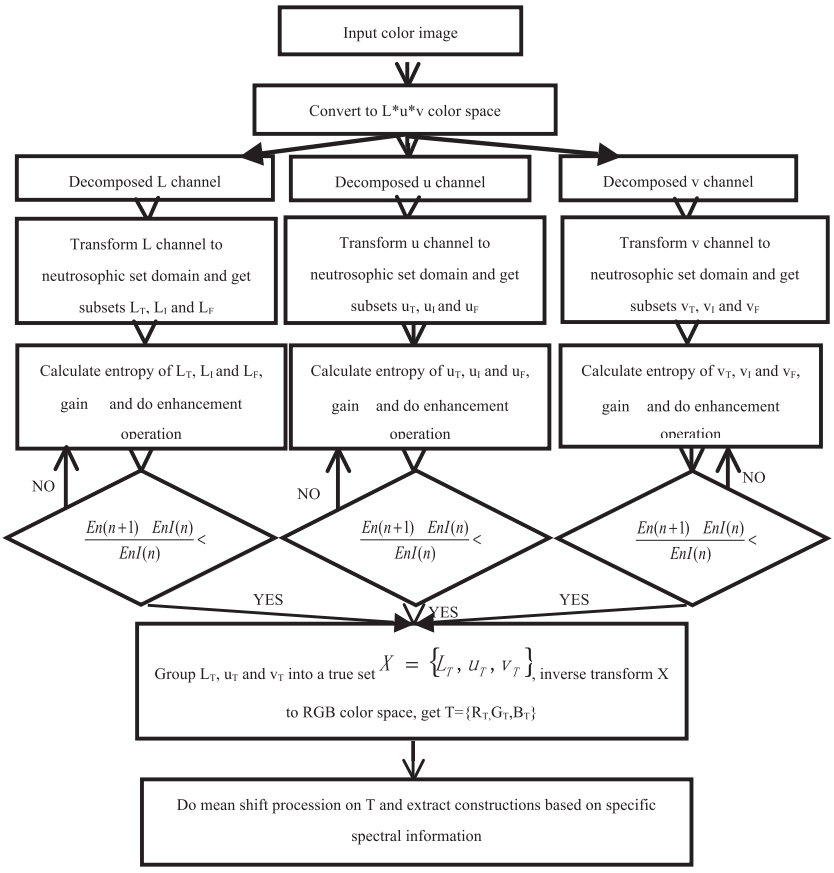

Fig. 1. Procedure of NS-MS algorithm.

bandwidth respectively which need to be determined by user. Here, h_s is assigned to be 20 while h_r is 16. C is the normalized constant [26].

Finally, turn the image back to RGB color space. The whole process can be summarized in the following chart

5. Experiments and discussion

To the best of our knowledge, among numerous methods proposed [27–30] evaluating performance of image segmentation, there is still not an universally accepted algorithm. Most ideas presented in image segmentation are assessed by comparing their performance in classification with some well-known methods or latest development in this area.

In order to see how well NS-MS algorithm can work in constructions detection, we contrast its performance not only with traditional mean shift method but also with two latest and most welcomed software packages in image segmentation. They are Berkeley ImageSeg (BIS: http://www.imageseg.com) and Environment for Visualizing Images Feature Extraction (ENVI-EX: http://www.exelisvis.com/). Both of them are object-based methods to segment images into real-world segments. In object-based theory, every pixel is considered as a object and two operations are needed to segment image [31].

The first one is calculating difference between continuous objects, obtained by difference in spectral heterogeneity h_p and difference in shape heterogeneity h_t.

$$h_p = \sum_I w_i(n_{ab}\sigma_{i,ab} - (n_a\sigma_{i,a} + n_b\sigma_{i,b})) \tag{32}$$

where $0 < w_i \le 1$, $\sum_P W_i = 1$, and w_i is the weight of band i, P is the number of bands of image. n represents the area of an object and σ_i is the standard deviation of an object in band i.

$$h_t = w_c h_c + w_s h_s \tag{33}$$

$$h_c = \frac{n_{ab}l_{ab}}{\sqrt{n_{ab}}} - \frac{n_a l_a}{\sqrt{n_a}} - \frac{n_b l_b}{\sqrt{n_b}} \tag{34}$$

$$h_s = \frac{n_{ab}l_{ab}}{b_{ab}} - \frac{n_a l_a}{b_a} - \frac{n_b l_b}{b_b} \tag{35}$$

and l is the perimeter of an object. b is the perimeter of an object's minimum enclosing rectangle. $w_c + w_s = 1$, $0 < w_c, w_s \le 1$, and they are user-defined.

The other one is merging. Considering the difference between regions calculated above, a synthetical criterion of differences can be generated as

$$f = w \cdot h_p + (1 - w) \cdot h_t \tag{36}$$

w is also assigned by user. If f is smaller than determinated merge-scale, object a and b can be merged into one object (Fig. 1).

5.1. Data

We use part of Xinjiang Province and Beijing as study area. For Xinjiang Province, four images of GeoEye are adopted to extract constructions (from Figs. 2–5). They were collected on May 24th, 2011. Constructions and roads mainly filled the images, and there is relatively large confusion among buildings, space

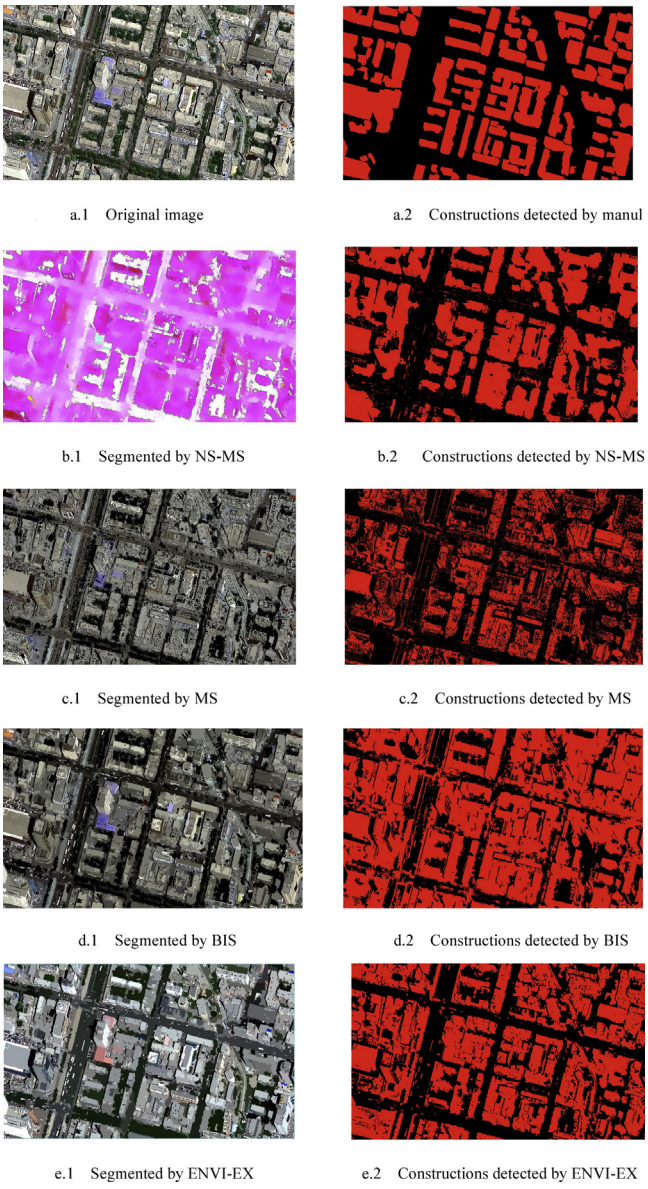

Fig. 2. Comparison of segmented results and extracted constructions.

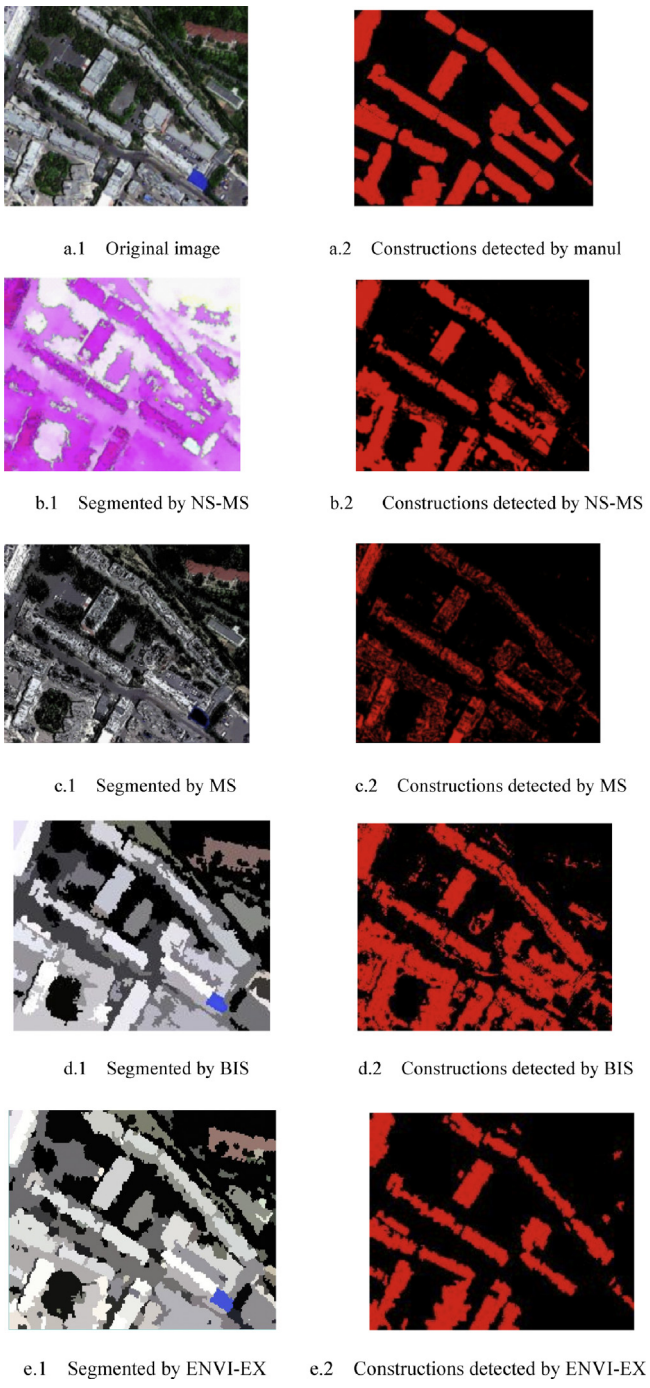

Fig. 3. Comparison of segmented results and extracted constructions.

and roads in Fig. 2. With such obstacles, robustness of the compared algorithms can be obviously extracted. GeoEye now offers the highest-resolution and the most accurate unclassified Earth imagery for clear insight. The proposed method, Mean Shift (MS for short), Berkeley ImageSeg (BIS for short) and ENVI-EX are used to segment and extract constructions from the four study images one by one, and the results of one image by the four methods are listed in one figure. Segmented results by the four methods test are shown respectively in the left line and the corresponding extracted results are displayed in the right (see Figs. 2–7).

With an intention to test whether the technique proposed in this paper can still work well in extracting constructions when faced with images of lower resolution and relatively poor quality, we use images of part of Beijing, collected from Resources Satellite number one 02C which was launched by China, to learn the pros and cons of the algorithm proposed in this paper. The images were recorded on March 8th, 2012 with a 2.36 m' spatial resolution and mainly focus on constructions (Figs. 6 and 7).

5.2. Results and discussion

5.2.1. Parameters determination

Since green vegetation has specific spectral characteristics, we extract plants out of the six study images to simplify features of images. All the four techniques need parameters to run, and some of the parameters need to be adjusted to the variation of images, while some others are defined 'one-off'. All the parameters can be grouped into two parts, one is segmentation part, the other is part of extraction.

In segmentation part, to objectively assess the robustness of NS-MS, MS, BIS and ENVI-EX, the same parameters are used for one method to segment six images. As for NS-MS and MS, bandwidth in

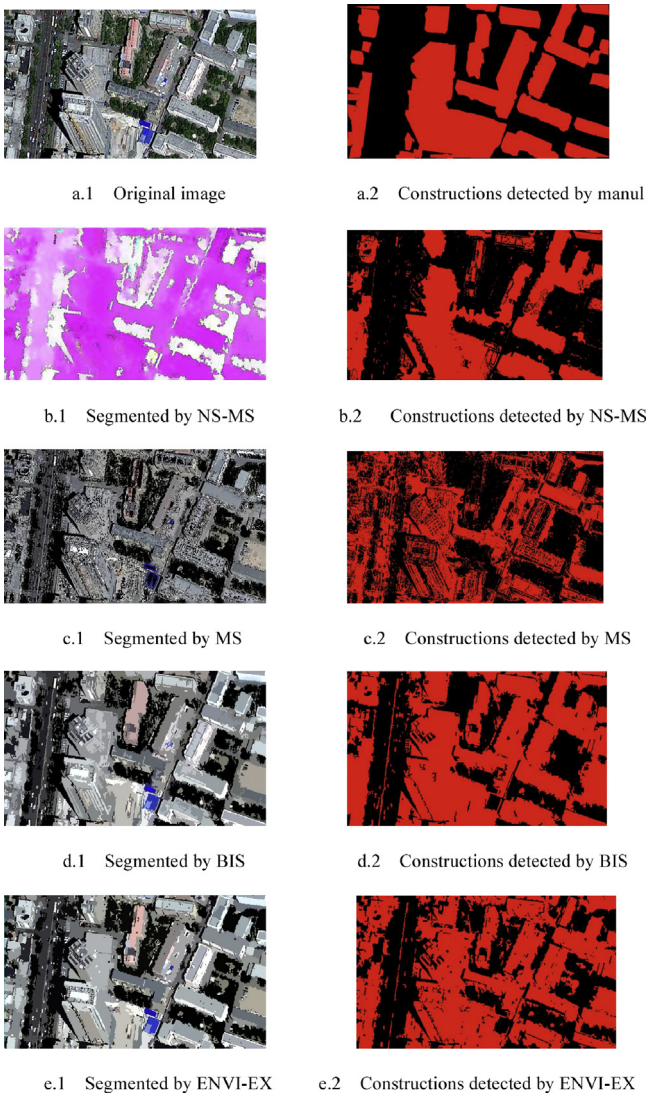

Fig. 4. Comparison of segmented results and extracted constructions.

space h_s and the one in color h_r are the same during the process of Mean Shift, and they are 20 and 16 apart. BIS has default parameters whose segmentation performance can be competitive. For the two weights, w_c and w are assigned both to be 0.5, and the merge scale is 50. Although the algorithm of ENVI-EX goes the same with BIS's, it has no default values, and with only two parameters in segmentation called scale level and merge scale. According to ENVI-EX tutorial, a scale level of 30.0 can best delineate the tops of constructions and at the same time, it can maintain details of them. As for merge scale, 94 would be a good choice [33].

When it comes to constructions detection, we divide the four methods into two groups according to their segmented results.

For group one, only NS-MS is included. From its segmentation results (Figs. 2b.1, 3b.1, 4b.1, 5b.1, 6b.1 and 7b.1), we can see that pixels of one feature have similar spectral information and larger differences from that of other features. Owing to this, spectral information can be used to extract constructions from six images with a fixed threshold.

The other three methods are grouped into another group. Their segmented results still preserve details of original image. When it comes to constructions extraction, characteristics of connexity regions can be helpful. But in order to guarantee the precision of extraction, supervised classification is adopted to extract constructions from images segmented, except for ENVI-EX. ENVI-EX

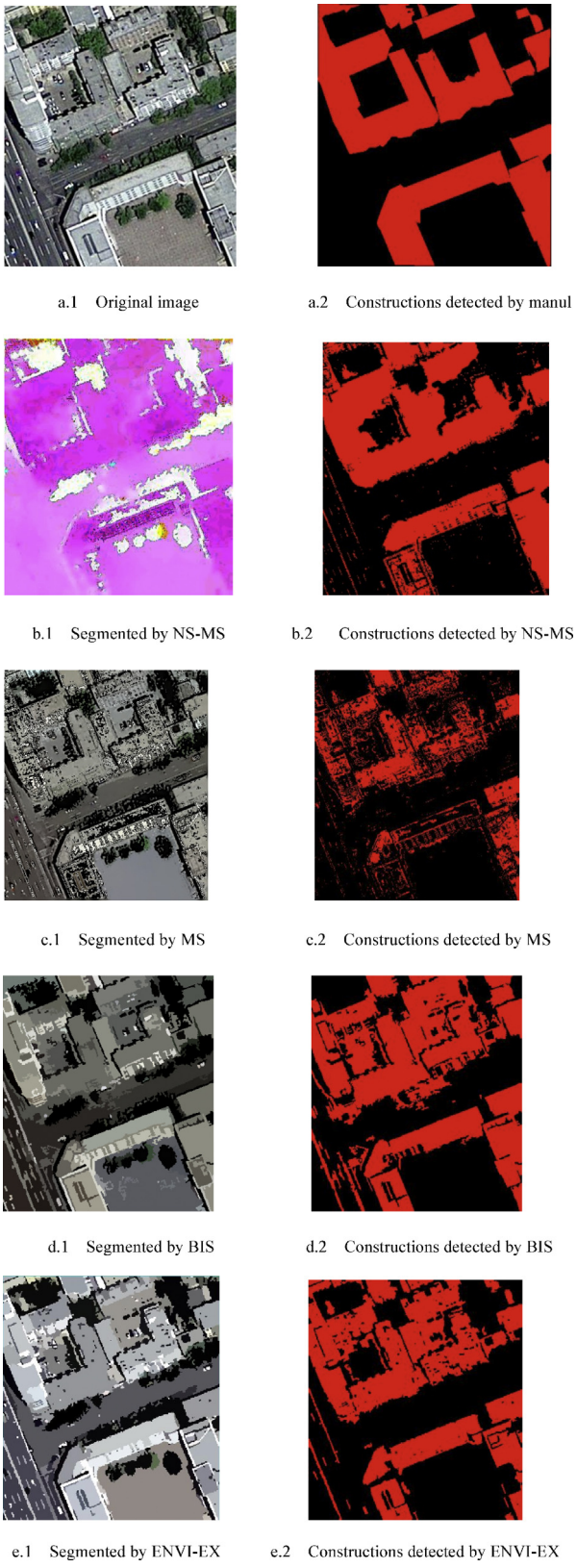

Fig. 5. Comparison of segmented results and extracted constructions.

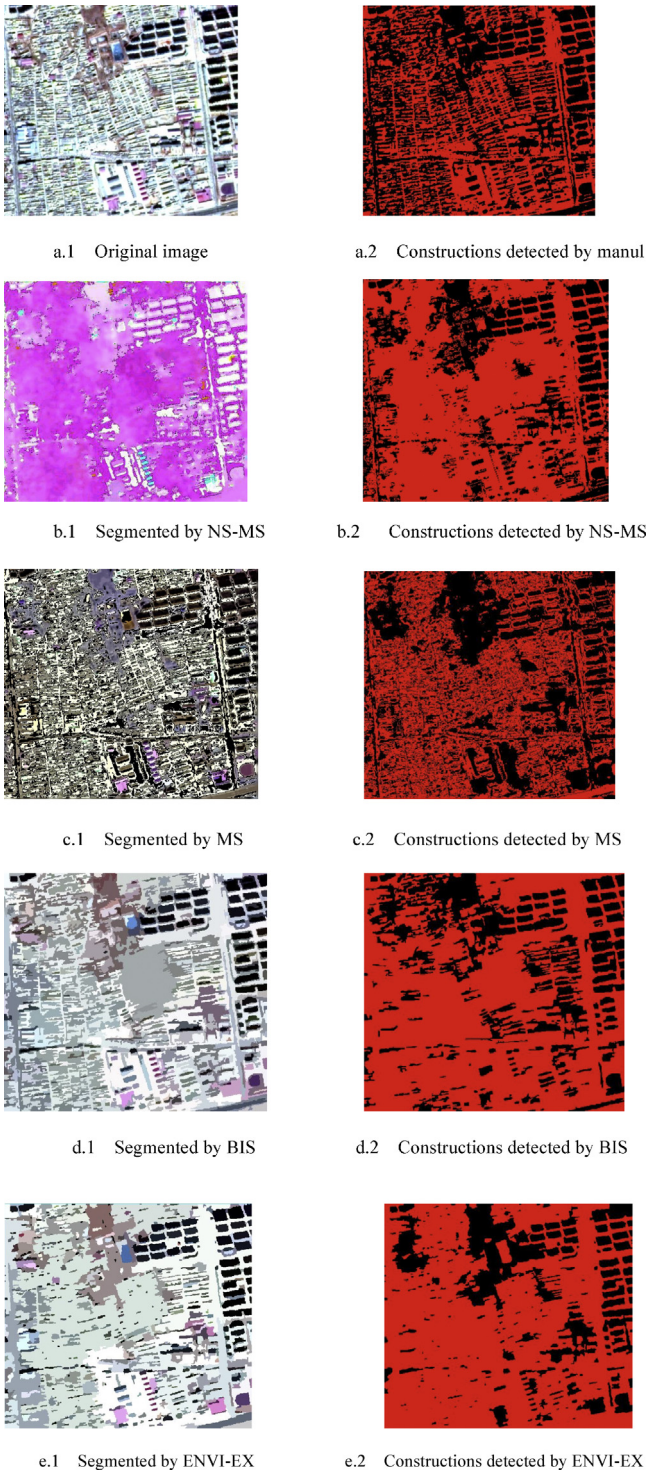

Fig. 6. Comparison of segmented results and extracted constructions.

is special for its design of software package, and several features of connexity regions (i.e., shape, texture, band ratio) are used to extract constructions out. On the basis of ENVI-EX Tutorial [32], four features can be used to extract buildings and rooftops out. The first one is band ratio, because normalized difference vegetation index (NDVI) of buildings and rooftops is next to zero. The second one is rectangle-fit, it represents how much the shape of buildings and rooftops approximates a rectangle. Area is the third one, it is a separation of buildings from other industrial or other sorts of buildings. Finally is the band ratio, the rooftops' color is always dark

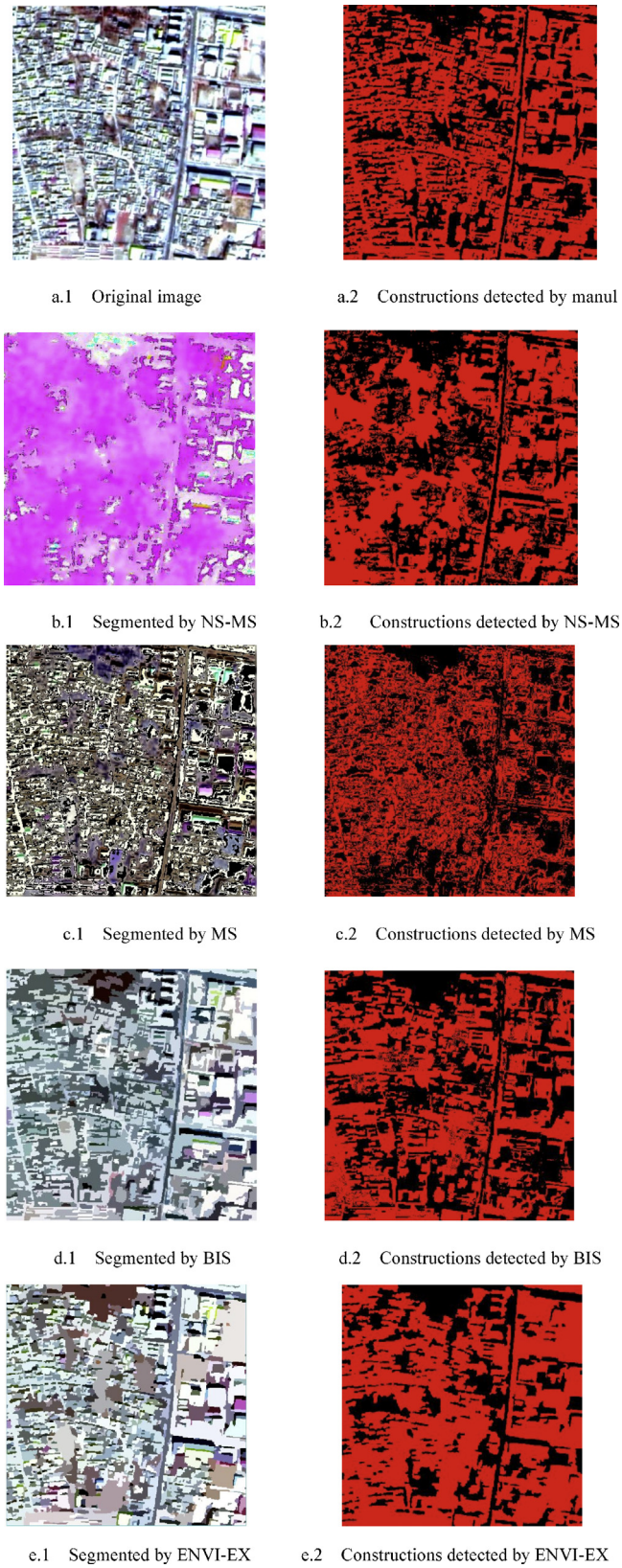

Fig. 7. Comparison of segmented results and extracted constructions.

Table 1
Statistical analysis of constructions extracted by four methods.

		Producer's accuracy (%)	User's accuracy (%)	Overall accuracy (%)	Kappa
Fig. 2	NS-MS	78.66	72.93	78.1418	0.5588
	MS	57.15	65.27	68.3200	0.3449
	BIS	84.71	57.81	66.6528	0.3563
	ENVI-EX	80.82	65.73	73.4835	0.4742
Fig. 3	NS-MS	89.15	79.51	89.8054	0.7660
	MS	87.01	76.12	87.8599	0.7230
	BIS	89.96	67.98	84.2021	0.6564
	ENVI-EX	73.84	84.31	87.9751	0.7040
Fig. 4	NS-MS	76.93	87.91	83.7949	0.6740
	MS	61.23	63.15	64.1237	0.2807
	BIS	92.66	73.95	80.7532	0.6180
	ENVI-EX	87.13	71.70	77.2414	0.5477
Fig. 5	NS-MS	89.15	83.95	88.8741	0.7704
	MS	59.38	80.82	78.1802	0.5237
	BIS	83.98	71.00	79.9314	0.5940
	ENVI-EX	82.09	80.98	85.1682	0.6914
Fig. 6	NS-MS	92.54	66.11	73.6326	0.4797
	MS	52.11	48.13	50.0273	0.0021
	BIS	93.14	63.81	71.3304	0.4356
	ENVI-EX	94.50	59.75	66.7784	0.3489
Fig. 7	NS-MS	81.42	86.68	81.7092	0.6272
	MS	53.10	62.55	53.6944	0.0741
	BIS	79.09	88.15	81.4376	0.6251
	ENVI-EX	92.36	79.68	81.6494	0.6077

and the spectral value in green band is relatively low. But the specific values for each image have to be adjusted, according to human knowledge and reasoning about specific feature types.

5.2.2. Performance evaluation by vision

Figs. 2a.2, 3a.2, 4a.2 and 5a.2 are generated by visual interpretation as evaluation criterion of the four methods. Figs. 6a.2 and 7a.2 are produced by supervised classification, because of low resolution of Resources Satellite number one 02C images, and we can hardly figure out exactly the outlines of constructions by naked eyes. We hold a detailed comparison with the evaluation criterion generated in each figure.

The extracted constructions by NS-MS are more neat and the roads and space extracted by mistake are less than the other three methods. Furthermore, the blocks generated by NS-MS are more real-world objects shaped and smoothed. While the other three techniques perform relatively bad. MS extracts less information compared with other methods (see Figs. 2c.2, 3c.2, 4c.2, 5c.2, 6c.2 and 7c.2) and its results are badly influenced by shades (see Figs. 4c.2 and 5c.2). Some obvious buildings are missed (see Fig. 4c.2). Redundancy information is a mechanical damage for BIS (see Figs. 2d.2 and 3d.2) and the distribution of constructions is not clear. Both BIS and ENVI-EX have recognized space and roads as constructions to a large extent (see Fig. 5d.2 and e.2). Moreover, extractions by ENVI-EX are fragmentary, they badly ruin the real features' shape (see Fig. 3e.2).

From segmentation results we can see that NS-MS is special (see Figs. 2b.1, 3b.1, 4b.1, 5b.1, 6b.1 and 7b.1), it contains various blocks whose spectral information is quite different from each other and the pixels in one block share similar spectral characteristics, which lays solid foundation for detecting construction and extracting them from other features with the help of spectral thresholds. For the other three algorithms, their segmented images mainly maintain the spectral signature of original image.

5.2.3. Performance evaluation by statistics

An accuracy evaluation is performed for the extraction results of each image in Table 1. It contains user's accuracies and producer's, overall accuracy and Kappa Coefficient of Agreement. Cohen's Kappa coefficient is a statistical measure of inter-rater agreement or inter-annotator agreement [33] for qualitative (categorical) items. Kappa coefficient is more stable for the reason that it takes the factor when agreement occurs by accident into consideration. It measures the agreement between two raters, one is recognized as ground truth classification, while the other is the figure that needs to be evaluated.

Kappa coefficient is defined below:

$$\kappa = \frac{P_1 - P_2}{1 - P_2} \tag{37}$$

$$P_1 = \frac{N_s}{N} \tag{38}$$

$$P_2 = \frac{N_{t1} \times N_{s1} + N_{t0} \times N_{s0}}{N \times N} \tag{39}$$

where N is the total number of pixels in each image, N_S is the number of pixels which are grouped into the same category in both images, one of which is considered as ground truth image while the other is the classification image, evaluated by the rater. The number of pixels of target object in reality is N_{t1}, and simulated pixels of target object is N_{s1}. Accordingly, the number of pixels of non-target object in reality is N_{t0}, and the simulated is N_{s0}.

Table 1 documents detailed accuracy assessment of the four methods (NS-MS, MS, BIS, ENVI-EX) extracting constructions, based on reference data in each figure. Kappa coefficient is grouped according to the theory proposed by Fleiss [35] that numerical area above 0.75 is regarded as perfect, from 0.40 to 0.75 is supposed to be good, and below 0.40 is poor.

From the analysis of Kappa coefficients of every image in Table 1,we can come to the conclusion that NS-MS shows the best performance with the highest kappa coefficient compared with the other three methods in extracting constructions from every image. And all of them are higher than 0.40, even higher than 0.75 which can be regarded as perfect. BIS and ENVI-EX perform neck and neck, both of them show bad performance in one image with a kappa coefficient less than 0.40. Interestingly, all of their kappa

coefficients are not higher than 0.75. MS performs badly in four images (see Figs. 2, 4, 6 and 7) with a kappa coefficient less than 0.40.

Overall accuracies share the same trend with Kappa coefficients. NS-MS can reach 89.8054% and it is averaged at 82.6597%. BIS and ENVI-EX share and share alike at about 78%. As for MS, its average overall accuracy is 67.03%, indicating its bad performance in extracting constructions.

Although there are two sorts of images with different spatial resolutions, NS-MS insists on behaving well, extracting constructions with real-world object shaped and smoothed, while MS performs worse with lower overall accuracy and Kappa coefficient in both images (see Figs. 6c.2 and 7c.2). BIS and ENVI-EX are both influenced with resolution decreasing, and their performances are just about the same.

All the four techniques show bad performance with low overall accuracies in Figs. 2 and 6 compared with other figures, one possible reason is that both the original images of Figs. 2 and 6 have numerous blocks confused with roads and space, which are obstacles for constructions detection. However, NS-MS still performs very well dealing with these images, owing that NS-MS segments image in neutrosophic set domain, which can enlarge the difference among pixels' spectral information of different category and smoothen the difference among pixels' spectral characteristics of the same category.

Compared with MS, BIS and ENVI-EX, NS-MS is a robust algorithm to segment image and do constructions extraction, regardless of the resolution of the image and the distribution of constructions and roads in the image. BIS and ENVI-EX perform relatively well, except for some conditions, i.e., objects are confused with each other to a large extent. MS can segment images well, but influenced by resolution of the image and the confusion between various objects when extracting constructions.

6. Summary

It is commonly believed that there are mainly two sorts of techniques to proceed after segmenting images in extracting constructions. One is supervised classification, and the other is extracting connected components based on geometrical characteristics and texture features, but the thresholds of each character can be one-off because of the various styles of buildings and variable study images. Faced with such problems, an unsupervised new algorithm which can extract buildings directly based on segmented images is proposed in this paper. It synthesizes neutrosophic set and mean shift to segment images, creating a new style of formats to display segmented results, and owing to this, constructions can be extracted by spectral information with a stable threshold. NS-MS has two key steps, one is transformation from color space to neutrosophic set domain and the other is mean shift segmentation in neutrosophic set domain.

Compared with three commonly used methods, MS, BIS and ENVI-EX, there are six main characteristics of NS-MS in constructions detection.

- Real-world object shaped and smoothed: constructions detected by this algorithm are real-world object shaped, and the pixels of one category are smoothed
- Robust: images, whether with high quality or relatively low quality, can be used to extract constructions with this method
- Unsupervised: the technique can 'one-step' extract constructions without human intervention
- One parameter: the parameter used in this algorithm does not vary with the images
- Dependable: Roads and space which are frequently confused with constructions can be got rid of as well
- Less redundant information: extracted constructions are neat and obvious, they are extracted as blocks with less trivial spots around.

Acknowledgements

This research is conducted with the help of 'Major projects of high resolution earth observation system', Major State Basic Research Development Program of China (2010CB950603), Public service sectors (meteorology) Special Fund Research (GYHY201006042), National Natural Science Foundation of China (41001209), European Commission (Call FP7-ENV-2007-1 Grant no. 212921) as part of the CEOP-AEGIS project (http://www.ceop-aegis.org/) coordinated by the University de Strasbourg and National Natural Science Foundation of China (40971202).

References

[1] S. Noronha, R. Nevatia, Detection and Modeling of Buildings from Multiple Aerial Images, IEEE Trans. Pattern Anal. Mach. Intell. 23 (5) (2001) 501–518.
[2] G.J. Hay, T. Blaschke, D.J. Marceau, A. Bouchard, A comparison of three image-object methods for the multiscale analysis of landscape structure, ISPRS J. Photogramm. Remote Sens. 57 (2003) 327–345.
[3] J. Tian, D.M. Chen, Optimization in multi-scale segmentation of high-resolution satellite images for artificial feature recognition, Int. J. Remote Sens. 28 (20) (2006) 4625–4644.
[4] K.V. Mardia, T.J Hainsworth, A spatial thresholding method for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 10 (6) (1988) 910–927.
[5] Jähne Bernd, Digital Image Processing, Springer, Berlin, 2005.
[6] A. Jain, Fundamentals of Digital Image Processing, Prentice Hall, Englewood Cliffs, NJ, 1989.
[7] J.L. Moigne, J.C. Tilton, Refining image segmentation by integration of edge and region data, IEEE Trans. Geosci. Remote Sens. 33 (3) (1995) 605–615.
[8] S.-Y. Chen, W.-C. Lin, C.-T. Chen, Split-and-merge image segmentation based on localized feature analysis and statistical tests, Graph. Model. Image Process. 53 (5) (1991) 457–475.
[9] F. Smarandache, A Unifying Field in Logics: Neutrsophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability, American Research Press, Rehoboth, 2005.
[10] J. Allen, S. Singh, Neurofuzzy and neutrosophic approach to compute the rate of change in new economies., University of New Mexico, 2002.
[11] M. Khoshnevisan, S. Bhattacharya, A short note on financial data set detection using neutrosophic probability, Florentin Smarandache, 2002, 75 pp.
[12] F. Smarandache, R. Sunderraman, H. Wang, Y. Zhang, Interval Neutrosophic Sets and Logic: Theory and Applications in Computing, HEXIS Neutrosophic Book Series, No.5, Books on Demand, Ann Arbor, MI, 2005.
[13] P. Kraipeerapun, C.C. Fung, Binary classification using ensemble neural networks and interval neutrosophic sets, Neurocomputing 72 (13–15) (2009) 2845–2856.
[14] R. Umberto, Neutrosophic logics: prospects and problems, Fuzzy Sets Syst. 159 (14) (2008) 1860–1868.
[15] M. Zhang, L. Zhang, H.D. Cheng, A neutrosophic approach to image segmentation based on watershed method, Signal Process. 90 (5) (2010) 1510–1517.
[16] H.D. Cheng, Y. Guo, A new neutrosophic approach to image thresholding, New Math. Nat. Comput. 4 (3) (2008) p291.
[17] Y. Guo, H.D. Cheng, New neutrosophic approach to image segmentation, Pattern Recogn. 42 (5) (2009) 587–595.
[18] L. Ma, R. Staunton, A modified fuzzy C-means image segmentation algorithm for use with uneven illumination patterns, Pattern Recogn. 40 (11) (2007) 3005–3011.
[19] A. Sengur, Y. Guo, Color texture image segmentation based on neutrosophic set and wavelet transformation, Comput. Vis. Image Understand. 115 (8) (2011) 1134–1144.
[20] Y. Cui, et al., An adaptive mean shift algorithm based on LSH, Procedia Eng. 23 (2011) 256–269.
[21] Y. Guo, H.D. Cheng, J. Tian, Y. Zhang, A novel approach to speckle reduction in ultrasound imaging, Ultrasound Med. Biol. 35 (2009) 628–640.
[22] C. Dorin, An algorithm for data-driven bandwidth selection, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2) (2003) 281–288.
[23] C. Dorin, V. Ramesh, P. Meer, The variale bandwidth mean shift and data-driven scale selection, in: Eighth IEEE International Conference on Computer Vision, 2001, pp. 438–445.
[24] D. Dubois, et al., Terminological difficulties in fuzzy set theory – the case of, Fuzzy Sets Syst. 156 (3) (2005) 485–491.
[25] P. Grzegorzewski, E. Mrówka, Some notes on (Atanassov's) intuitionistic fuzzy sets, Fuzzy Sets Syst. 156 (3) (2005) 492–495.
[26] C. Dorin, M. Peter, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 603–619.

[27] B.C. Li, Z.G. Guo, C. Wen, Multi-level fuzzy enhancement and edge extraction of images, Fuzzy Systems Math. 14 (4) (2000) 77–83.
[28] M. Möllera, L. Lymburnerb, M. Volkc, The comparison index: a tool for assessing the accuracy of image segmentation, Int. J. Appl. Earth Observ. Geoinform. 9 (3) (2007) 311–321.
[29] Q. Zhan, M. Molenaar, K. Tempfli, W.Z. Shi, Quality assessment for geo-spatial objects derived from remotely sensed data, Int. J. Remote Sens. 26 (14) (2005) 2953–2974.
[30] UW, Contribution to the assessment of segmentation quality for remote sensing applications, in: International Archives of Photogrammetry and Remote Sensing XXXVII (Part B7), Beijing, 2008, pp. 479–484.
[31] U.C. Benz, et al., Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, J. Photogramm. Remote Sens. 58 (2004) 239–258.
[32] ENVI EX Tutorial: Feature Extraction with Rule-Based Classification. Internet: http://geology.isu.edu/dml/ENVI_Tutorials/Feature_Extraction_RuleBased.pdf
[33] J. Strijbos, et al., Content analysis:what are they talking about? Comput. Educ. 46 (2006) 29–48.
[34] http://en.wikipedia.org/wiki/Kappa_coefficient
[35] J.L. Fleiss, Statistical Methods for Rates and Proportions, 2nd ed., John Wiley, New York, 1981.