



# A multi-scale method for urban tree canopy clustering recognition using high-resolution image



Jian-Nong Cao<sup>a</sup>, Zhenfeng Shao<sup>b,\*</sup>, Jia Guo<sup>a</sup>, Bei Wang<sup>a</sup>, Yuwei Dong<sup>a</sup>, Pinglu Wang<sup>a</sup>

<sup>a</sup> The School of Earth Science and Resources, Chang'an University, No. 126, Yanta South Road, Xi'an 710054, China

<sup>b</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

## ARTICLE INFO

### Article history:

Received 10 February 2014

Accepted 18 February 2015

### Keywords:

Image segmentation  
High-resolution image  
Constraints mean shift  
Wavelet  
Tree canopy

## ABSTRACT

In this paper a constraint mean shift method is proposed for extracting urban tree canopy with the use of high-resolution image. Through establishing multi-scale pyramid image feature space by wavelet, features between layers can be combined as the constrained route for mean shift to realize self-adaptive decomposition and scale transfer in multi-scale feature space, then differences of internal and external structure in urban tree canopy and differences of average spectral radiant intensity are used as multi-scale feature space of wavelet to realize the preliminary clustering segmentation, finally we apply the supervised segmentation to extract tree canopy based on clustering feature. Experiments demonstrate that the proposed method can eliminate the over-detailed effect of image accurate extraction of the urban tree canopy can be achieved.

© 2015 Published by Elsevier GmbH.

## 1. Introduction

Urban vegetation coverage has a great impact on the environment and its accurate estimation is of significance to guide the future urban planning management and environmental protection. Urban tree canopy is an important part of eco-system and an important biophysical index for ecological evaluation. Living vegetation volume (LVV) is defined as the volume of space occupied by stems and leaves. It is also the first 3-dimensional index of the urban greening index system. LVV has great significance in the research of greening layout, quantity and their correlation with atmospheric environment [1]. With the rapid change of urban environment, there is an urgent need for using the remote sensing technique to calculate large area coverage.

The aerial remote sensing, multi-spectral remote sensing, high-resolution remote sensing images are the main data sources for information extraction of urban tree canopy. Conventional supervised and unsupervised classification is the primary method to extract the urban tree canopy information [2]. Literatures [3–5] studied on extracting the information of tree canopy by high-resolution sensing images using methods including learning and supervision, region-based growing, watershed, template matching

and the mixture of these methods, but few consideration has been focused on the role of image scale on tree canopy recognition.

Literature [6] presented a multi-scale texture segmentation method whose multi-scale pyramid feature space was based on the mean of low-frequency and standard variances of high-frequency decomposed by the wavelet transform coefficient, applying mean shift algorithm to decompose the feature space, separating between different areas in coarse scale, locating the edge and other details in finer scale, and transferring the clustering results from coarse to finer scale layer-by-layer with a proper strategy, which led to a successful multi-scale texture segmentation. However, this method destroyed the smoothness of the feature space at different scales during transferring multi-scale features and increased the difficulty of parameters selecting, such as the bandwidth of mean shift (MS).

Thus, this paper integrates the spectral information (the mean of wavelet transform low-frequency coefficient) and texture information (the standard variances of wavelet transform high-frequency coefficient) in multi-scale structures (wavelet transform pyramid structure). Pyramid structure of the multi-scale feature space is established by wavelet decomposition. Image features are calculated by differences of internal and external urban tree canopy structure and difference of the average spectral radiant intensity, presenting the constraints mean shift (CMS) method to solve the complexity of feature space caused by multi-scale clustering transfer, and overcome the defects of target recognition caused by the lack of basic feature factors (spectrum, texture or scale). Self-adaptive decomposition of feature space is conducted with the

\* Corresponding author. Tel.: +86 15827188114.

E-mail addresses: [caojiannong@126.com](mailto:caojiannong@126.com) (J.-N. Cao), [shaozhenfeng@whu.edu.cn](mailto:shaozhenfeng@whu.edu.cn) (Z. Shao).

constraints mean shift (CMS) algorithm, which reduces the resolution of the image segmentation. Moreover, this paper presents a supervised segmentation method based on clustering feature, which further refines the segmentation results.

The proposed method works better on segmentation of high-resolution image. The rest of this paper is organized as follows: the first section discusses the constraints mean shift clustering; the second section is the extraction algorithm of the urban tree canopy based on clustering features; the third section is the experiment and analysis; the fourth section is the conclusion.

## 2. Constraints mean shift clustering

Wavelet transformation provides a comprehensive method texture analysis, which integrated of frequency spectrum, image structure and pixel statistics. Through decomposing remote sensing image via computational functions involving flexible, shift, etc., wavelet transformation holds strong capability in spatial and frequency decomposition. As an image analysis tool with high performance, wavelet transformation could give unified framework for image texture features representation and extraction. In this study, low pass image signal is recursive decomposition, each scale features are combined with low pass and high pass signals, respectively. Texture features of image regions would be discomposed in the successive layer, the pyramid structure of wavelet coefficients feature space can be constructed by salient feature regions in each scales.

$$W_f(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t)\psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

where  $\psi(x) \in L^2(R)$  is wavelet basis,  $a$  is flexible factor,  $b$  is shift factor. If binary discretization is operated for scale factor of continued wavelet transformation and shift factor ( $a = 2^j, b = 2^j k, (j, k \in Z)$ ), there would be produced a wavelet with scale and time discretization.

Some literatures demonstrated the convergence [7,8] and sufficient condition of convergence [9] from mean shift algorithm on continuous and discrete function. Literature [9] successfully applied the algorithm on image segmentation and smoothing. It obtained the extensive application on image filtering, edge extraction, information fusion and other fields of image processing.

The Eq. (2) is applied to conduct iterative computations in literature [9].

$$y_{j+1} = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{y_i - x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{y_i - x_i}{h}\right\|^2\right)}, \quad j = 1, 2, \dots \quad (2)$$

where  $y_{j+1}$  is the process feature after the  $j$  time iteration, in Eq. (1),  $x$  is the current point, the process feature which is the closest to  $y_{j+1}$ , namely data center within bandwidth. Literature [9] demonstrated that sequence  $y_{j+1}$  can do convergence to  $y'$  when meeting certain conditions, then  $y'$  can be called corresponding pattern of  $x$ . The complexity of mean shift technique relies on two correlative controlled parameter, namely bandwidth  $h$  and density  $T$  [8].

In image processing, except for the feature space of transformation, the spatial information within every pixel is also important. So, it is necessary to apply the spatial information of image to feature vectors, then, the kernel function is changed as following:

$$K_{h_s^2, h_r^p}(x) = \frac{C}{h_s^2 h_r^p} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \quad (3)$$

where  $x^r$  is the feature of space location,  $x^s$  is feature vector that constitutes feature space,  $h_r$  and  $h_s$  is separately space-bandwidth

and feature-bandwidth, value of the window about space-width in image is  $h_r = (2n_r + 1) \times (2n_r + 1)$ ,  $C$  is normalization constant.

The density surface of discrete feature space in factual image is normally unsmooth, and density statistics is closely related to the selection of bandwidth in non-uniform discrete data [8–10]. If the bandwidth is too large, it may cause over-generalized density pattern distortion; if too small, it may easily cause density islands and other over-trivial density patterns in sparse data area, making the algorithm fall into local maximum. So, it is a reliable method [8] to utilize smaller bandwidth with attached density constraint. Density constraint cannot only avoid insignificance pattern's judgement procedure, but also improve clustering accuracy and speed up the convergence.

The difficulty of bandwidth selection is that both the narrow peak of statistical density model and the variation of details are related to bandwidth, literature [11] proposed two self-adaptive schemes: one is non-parametric method, which needs to define new self-adaption process of mean shift, applying insert rule and sampling point density estimation operator; the other is semi-parametric method, applying local data structure to extract scale information. Selection of clustering in self-adaptive bandwidth is superior to the fixed bandwidth method. However, it also increases the computation cost of mean shift, on the other hand. In addition complexity of image feature space is hard to be adapted by insert rule and structure information. Generally speaking, the peak-to-valley ratio method proposed by literature [8] is more reliable and practical, but the computation of peak-to-valley ratio needs to build density surface model which increases extra computing burden. Based on the above analysis, this paper proposed the constraints mean shift (CMS) method to solve the robustness of mean shift in complex feature space, bandwidth constraint information is from merging features transferred by scale, with no extra computing.

From Eq. (2), we can find that the conventional mean shift algorithm assumes that density smoothing is given with a limit of density change rate. As indicated in Fig. 1, the blue mass points can be used for obtaining a smooth feature space, in this way the algorithm can be successfully implemented. However, the red mass points in Fig. 1 are fragmented, discontinuous, stair-step unsmooth, indicating the difficulty to obtain a smooth feature space, so the reasonability and continuity of the mean shift route is invalid with Eqs. (1) and (2), and the algorithm is difficult to implement in this kind of feature space. Studying on the constraints mean shift (CMS), it

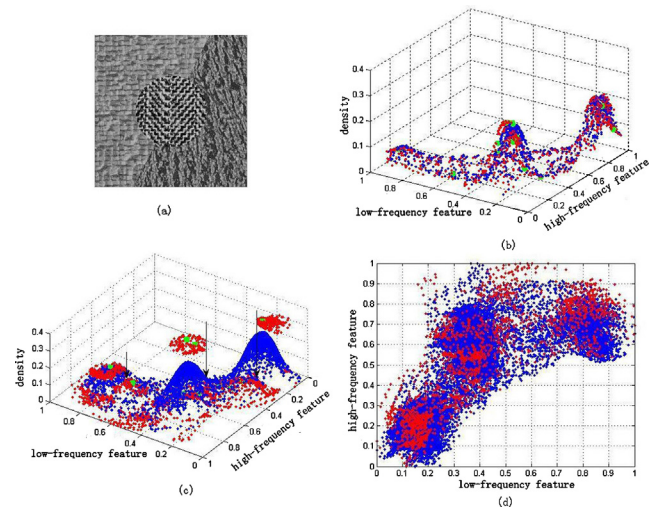


Fig. 1. Feature density distribution of scale feature transfer plot: (a) three composite image; (b) the density distribution of the fourth layer on the third layer; (c) the density distribution of the second layer on the first layer; (d) feature density distribution of the second layer on the first layer.

is necessary to discuss the reasons resulting in the feature space unsmooth, then the targeted route constraints can be proposed. The feature transfer strategy in literature [6] is the reason which causes the feature space unsmooth. The following briefly describes the principle.

Literature [6] proposed the feature transfer principle of multi-scale image pyramid, which decomposes the image into four-layer channels of the wavelet coefficient to obtain the feature layers gradually altering from coarse to fine scales. Applying coarser scale feature during texture segmentation is implemented by using larger window of feature analysis. This can effectively segment the main outlines of different texture regions, but the local edge details among different texture regions are lost. Moreover, the window of feature analysis would be small, when the window being close to the edges of texture, which contains multiple texture information, may cause falsely segmenting. This small window based analysis can improve the accuracy of segmentation in the edges, but it will cause a weak consistency between image regions. So, combining both effectively must rely on proper segmentation scale selection for different image areas as well as self-adaptive feature analysis. A basic principle is: if containing no image edge in window, larger feature analysis window should be used to segment; if containing edge, reduce the size till it has no any edge [12]. The thought can be realized by searching nature of the neighborhood at the beginning of segmentation in current window, namely if the segmentation result from current window and its neighbor window is divided into the same class in coarse scale feature space, then considering the window has no edge and marking it, otherwise no marking. The unmarked areas can be regarded as containing image edge, namely border zone between different textures, then apply finer scale feature to segment them. In other words, the window size of feature analysis is up to area's consistency of segmentation in upper level and its nature of the neighborhood. This paper uses eight neighborhoods to judge and search. The feature of wavelet coefficient meets the Markov property. Namely property in current window is only up to that in neighbor. Size of each scale coefficient matrix decomposed by orthogonal wavelet is one quarter of previous finer one, which means every wavelet coefficient is corresponding to four coefficients in the next level, equal to divide the parent feature window into four parts in the next scale. Segmentation process begins with the first segmentation and mark the image in the coarsest scale feature of wavelet decomposition using mean shift, then quadruple marked result, apply corresponding finer scale feature to unmarked areas, marked areas as a whole can segment further by using feature mean in same segmentation areas, till obtain segmentation result with the finest scale.

For the convenient analysis of the variation in feature space before and after scale transfer, we take three texture composite images as examples (Fig. 1(a)), and diagrammatize change of density distribution about feature space points. In Fig. 1, blue is the feature points in current channel, red is the clustering feature points after feature transfer from upper channel, green points are the clustering center, which is the mean of quadruple clustering result from wavelet coefficient in upper level and corresponding mean of homogeneous areas feature in next level. The edge feature points stay the same. Clustering feature space consists of clustering center and edge feature points. Fig. 1(b) is the density distribution of the fourth layer feature channel on the third-layer feature transfer result, due to the fourth feature space is sparser with more green clustering center points (20 classes) and disperse clustering, so in Fig. 1(b), the discontinuity and stair-step in red mass points is slighter than them in blue; but the lower mean shift is, the closer to the actual class the clustering is, and more significant the feature transfer result is, Fig. 1(c) is the density distribution of the result of the second layer feature channel on the first layer feature transfer, the discontinuity and stair-step in clustering

feature space is significant. Literature [6] made selection of feature clustering step by step among multi-scale through different scale feature of multi-channel transfer come true, though clustering feature space implies clustering information among scales, its smoothness is badly broken, showing discontinuity and stair-step distribution, make regular mean shift method hard to be carried out, as the red points shown in Fig. 1(b) and (c).

The reason on discontinuity and stair-step distribution of scale transfer feature space is analyzed. The space contains two kinds of points: clustering center and edge points. On the one hand, computing mass points needs to assume the bandwidth, the density computation of every clustering center must base on certain range, assuming that bandwidth is small enough so that it cannot contain two or more clustering center. Then neighbor points of clustering center in certain range should be edge points. Meanwhile, density of clustering center mainly consists of the iteration count on similar feature pixel, which is corresponding to clustering center, so its density must be higher than density of edge points. Similarly, the density computation of every edge point must also base on certain range of bandwidth, the density of edge point closer to clustering center must be raised by density of clustering center, so it causes stair-step variation in feature space, and edge of image area must be at the edge of stair-step variation, as black arrow shown in Fig. 1(c). Because of transmitting to the lower channel, iteration count of clustering center is bigger, so the discontinuity and stair-step phenomenon is more significant. On the other hand, because clustering center is mainly consist of the iteration count on similar feature pixel which is corresponding to clustering center, so the horizontal projection in clustering feature space shows dis-sparse, it aggravates the discontinuity on horizontal direction, as the red points shown in Fig. 1(d).

Obviously, applying regular mean shift algorithm to clustering feature space (red mass points) is hard to make convergence to the optimized pattern with higher density. Analyzing feature clustering space can obtain three characters, first, the horizontal projection in clustering feature space is more shattered and discontinuous; second, it has large stair-step and discontinuity on density direction (vertical); third, edge points are at the border zone between discontinuity and longitudinal stair-step of horizontal direction. According to the three characters, this paper proposed the route's constraints mean shift (CMS) method, it can not only use clustering information between scales, but also carry out the constraints mean shift method successfully, and realize robust clustering during clustering information of coarse scale transmit to the fine one.

Specific processes:

**Step 1:** starting points of mean shift are restrained to be clustering center points in feature space of upper level channel (green points in Fig. 1). Taking Eq. (4) for iterative calculations, the first value  $x_1$  of the current point  $x$  must satisfy  $x_1 \in Cet_{Le}$ ,  $Cet_{Le}$  in equation it is the clustering center points' collection after the current channel gets the upper clustering information, subscript  $Le = 3, 2, 1$  is the pyramid layers, 4 layers in this paper, the top layer (fourth) has no route constraints.  $Cet_{Le}$  is the contour points of the new feature space after clustering with higher density, mean shift route constraints in  $Cet_{Le}$  consistently drifts toward the higher density feature points, avoiding the route accessing to low density areas, overcoming local maximum clustering error caused by unsmooth data, which can save convergence time and exclude low density clustering.

**Step 2:** after iterative calculation,  $x_i$  is the current point, subscript  $i \geq 2$ , the density models satisfy  $y_1 < y_i < y_{i+1} \cdots \leq y'$ , start a new iteration when converging to  $y'$ , repeat step 1 and 2. Mean shift route can access to the edge points with higher density, then become new clustering center that do not fall into the edge points with lower density, conform to the character of scale transfer.

**Step 3:** given proper density threshold constraint can obtain correct clustering pattern, because the intermediate layer has a larger change, no meaning of convergence, so threshold is just set at the lowest level. Density threshold selection is experiential, but has its own regularity, around the density jump location is often density threshold with larger value range and no sensibility.

Fig. 2 takes extracting tree canopy and green lands in dense urban image as an example, makes experiments using constraints mean shift, then obtains segmentation process of clustering in each wavelet decomposition level from coarse to fine scale, as shown in Fig. 2(b) and (e). It can be seen that the final clustering result is shown in Figure 2(e), which has error only at the edge, and shows a better effect. Experiment shows that bandwidth variation is insensitive in a certain range when using constraints mean shift method, so the method is more robust.

### 3. The feature extraction algorithm based on clustering for urban canopy

Using the multi-channel and multi-scale mean shift clustering to segment the texture depends on the consistency of multi-scale feature transfer and the separability of each distribution features in each channel. However, it should be noted that using the mean-shift method for texture segmentation has some inherent technical defects: firstly, the down-sampling of the wavelet feature extraction leads the fine-scale spatial channel and the size of original image difference between 1/4 pixel. Therefore, its segmentation and the original image spatial resolution are away 1/4 pixel; secondly, feature clustering accuracy depends on the separability of the feature space, and the wavelet feature extraction is difficult to ensure completely separation of the texture space, so the feature confusion is difficult to avoid, the fault texture segmentation for edge is 4 times over the original image.

The key difficulty of texture segmentation lies in the texture sampling and the determination of category, CMS technology effectively solves the determination of category and the sampling. This

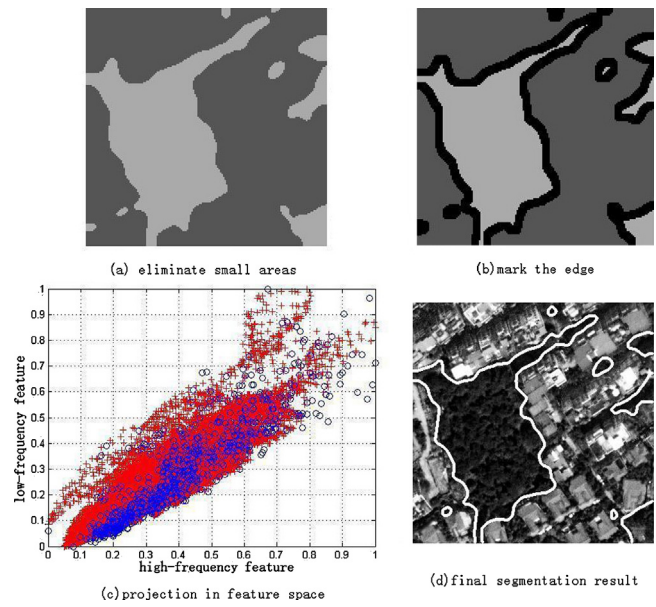


Fig. 3. Algorithm principle for supervised feature based on clustering.

paper makes the CMS clustering as the results of the coarse texture segmentation, using the image space texture or variation, presents the supervised segmentation method based on clustering feature to refine the edge segmentation.

Principles and steps are as follows:

**Step1:** Sampling image preparation: mean shift clustering segmentation results can determine the number of texture categories, eliminate small areas and extend clustering image to four times, obtain the clustering image as the same size as original image, which is called the sample image. The neighborhood window size for eliminating small areas is  $(2n_E + 1) \times (2n_E + 1)$ . Fig. 3(a) is the result after eliminating small areas for 2(e).

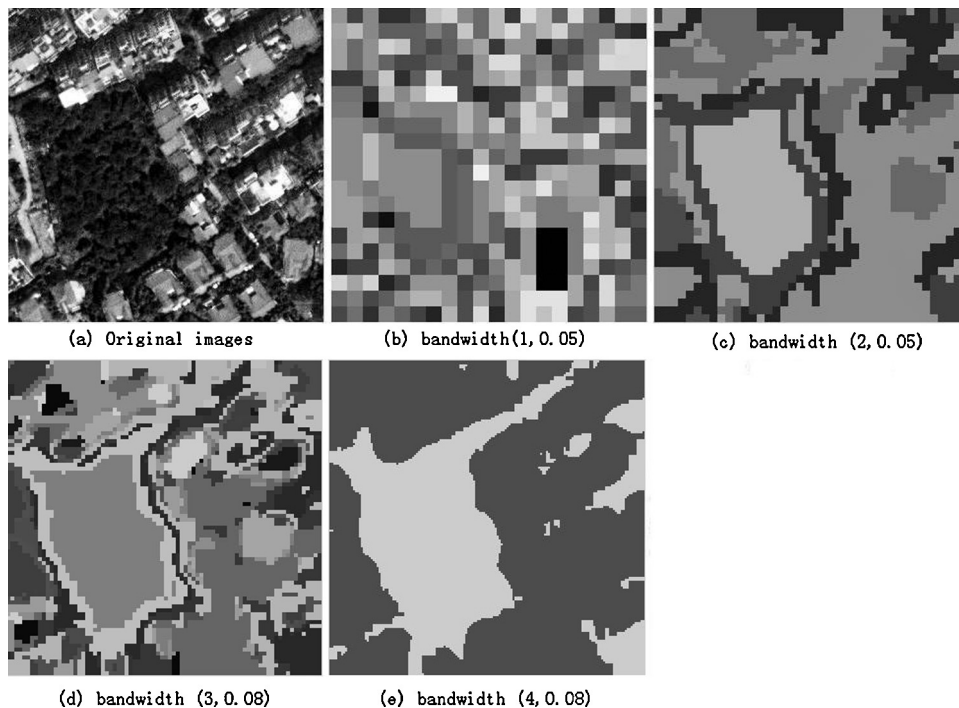


Fig. 2. Segmentation process of clustering in kinds of scale under constraints mean shift method, b and e are the segmentation results from coarse to fine scale.

**Step2:** Marking the edge to prepare for unrecognized pixels: sampling image presents the errors of edge segmentation, so make 8-adjacent connection marking with the principle in literature [6], labeling the pixels within the window having no edge(save classification results), and unlabeled pixel(value equals 0) is the unrecognized pixel. The marking window size is  $(2n_l + 1) \times (2n_l + 1)$ . Fig. 3(b) is the labeling result for Fig. 3(a).

**Step3:** Extraction of sample feature: clustering regions of sampling image are irregular shapes, under constraints of the marked image, after calculating the texture feature, it needs to divide image into the non-overlapping and regular image blocks, window of the image block is  $(2n_3 + 1) \times (2n_3 + 1)$ , supposing the number of clusters is  $L$ , each class can be divided into  $W_{L,p}$  windows, the total number of sampling image windows is  $S_p = \sum_{i=1}^{i=L} W_{L,p}$ , constituting

the calculation unit for the classified pixel texture feature, the mean of low-frequency and standard deviation of high-frequency constitutes the feature points, whose number is  $S_p$ , to supervise the attribution of unrecognized pixels, as the blue points in 3(c).

**Step4:** Feature extraction of unrecognized pixels: set unrecognized pixels as the center, calculate the pixel texture, the window is  $(2n_3 + 1) \times (2n_3 + 1)$ , the number of the feature points equals to the number of unrecognized pixels, as the red points in Fig. 3(c).

**Step5:** Supervised calculation: in the feature space, calculate the feature distance between the unrecognized pixels (the red points in Fig. 3(c)) and the sample feature points (the blue points in Fig. 3(c)), take the genus of the unrecognized pixels, which has the minimum distance to sample feature points.

**Step6:** Image segmentation: after calculating all the unrecognized pixels, then project the classification features into the image space to complete the image segmentation.

The computational complexity is determined by the number of unrecognized pixels and the number of sampling feature points. Among them, the number of unrecognized points can be reduced in the mean shift clustering stage, and sampling feature points can be properly controlled, meanwhile, the algorithm process can be computed parallelly, so the efficiency can be guaranteed in realization of software engineering.

#### 4. Experiment and analysis

The constraints mean shift algorithm of this paper share the same theoretical foundation with the normal mean shift while it improves to make constraint on mean shift route after complexity of the feature space caused by scale transfer. Thus, the route constraints thought of the constraints mean shift algorithm, whose substance is a calculative strategy according to the mean shift on multi-scale applications. Literature [6] compared the performance, accuracy and computational efficiency between the constraints mean shift algorithm and normal mean shift and the results showed the sensitivity of bandwidth selection in the constraints mean shift algorithm dropped significantly. The degree of automation increased greatly. Accuracy was basically the same between the constraints and normal mean shift algorithm both on the segmentation of the texture edge; the constraints mean shift algorithm in the feature space avoided making invalid search process and avoided its time consumption on lower-density data through inheriting the clustering information in the feature channel with the coarser scale, thus, the efficiency of clustering could be increased to the most 2.06 times, the least 1.38 times, average 1.73 times.

We select the high-resolution remote sensing images in which the urban tree canopy distribution is typical to implement the extraction experiment. Meanwhile direct supervised and unsupervised classification methods are also selected as the comparison purpose. Separately, the quantity error of regional pixels and the distance error of boundary pixels are utilized to estimate the accuracy of segmentation. The experiments use 0.5m spatial resolution aerial remote sensing images in some region of China, and the extraction results are shown in Figs. 4–6. Fig. 4 is the result processed by the proposed method, experimental parameters of Fig. 4 are listed in Table 1. Fig. 5 is the result of direct supervised classification, Fig. 6 is the result of the unsupervised classification, the algorithm in Figs. 5 and 6 both make experiments in ERDAS, the segmentation results are post-processed by eliminating the small areas noise, and the window size of filtering is  $9 \times 9$ , equal to  $4.5 \text{ m}^2$  in actual ground.

The extraction results in Fig. 4 indicate that the contiguous tree canopy could be extracted effectively. The scattered tree canopy is more difficult to extract, which is susceptible to omit. The shadow between the tall buildings further has a greater impact on the lower tree canopy though the proposed method obtains a better

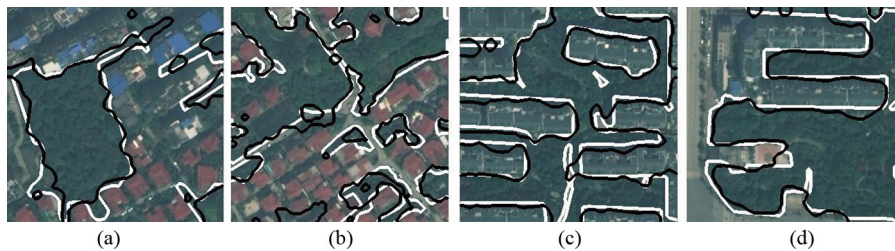


Fig. 4. The extraction results of the tree canopy in typical urban processed by the proposed method (320\*320).

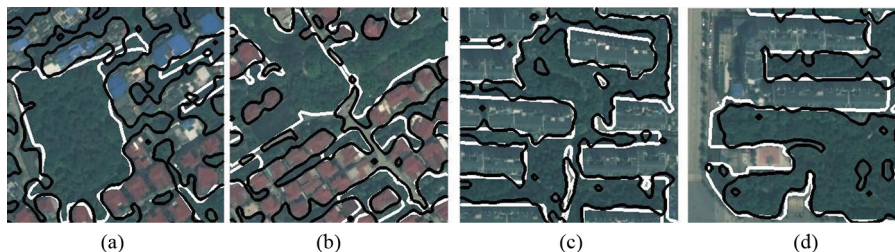


Fig. 5. The extraction results of the tree canopy in typical urban processed by direct supervised method (320\*320).

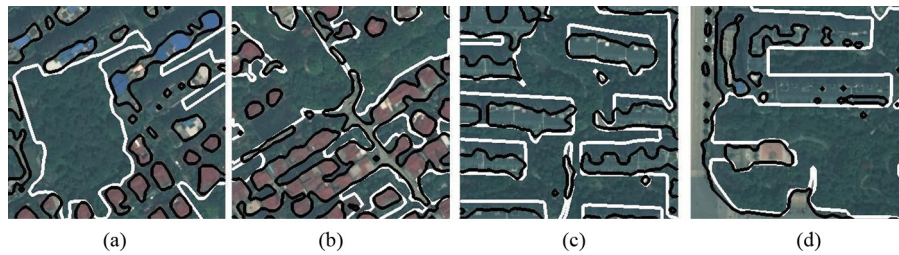


Fig. 6. The extraction results of the tree canopy in typical urban processed by the unsupervised method (320\*320).

Table 1  
Experimental parameters.

Image		Bandwidth		Density	Features window			Supervised classification parameters		
Drawing no.	Stratification	$h_s$	$n_r$	$T$ (number of pixels)	Low-frequency $n_1$	High-frequency $n_1$	Smooth $n_2$	$n_E$	$n_L$	$n_3$
Fig. 4a	4	0.05	1	1	1	3	1	2	2	6
	3	0.05	2	1	3	3	1			
	2	0.08	2	1	3	3	3			
	1	0.08	4	1800	5	5	3			
Fig. 4b	4	0.05	1	1	1	3	1	1	1	6
	3	0.05	2	1	3	3	1			
	2	0.08	2	1	3	3	3			
	1	0.1	4	2600	5	5	3			
Fig. 4c	4	0.1	1	1	1	3	1	5	5	6
	3	0.1	1	1	3	3	1			
	2	0.1	1	1	3	3	3			
	1	0.1	1	1680	5	5	3			
Fig. 4d	4	0.05	1	1	1	3	1	3	3	6
	3	0.08	2	1	3	3	1			
	2	0.08	2	1	3	3	3			
	1	0.1	4	2500	5	5	3			

result on the recognition of integrated scale and texture, because of only taking account of panchromatic spectral intensity, therefore there is a further improvement on the spectral recognition. Compared with Fig. 4, in the extraction results of Fig. 5, the mis-segmentation increases significantly. The main reason is that it is difficult to accurately distinguish the shadow of the tree canopy from the tree canopy in the spectral sampling of the tree canopy, the shadow of buildings and the shadow of the tree canopy is very similar in direct supervised segmentation, thus causes severe mis-segmentation. Meanwhile, because direct supervised segmentation takes no account of scale and texture, it causes many small speckles within segmentation region. Moreover, it still has more broken patches after post-processing. Compared with Fig. 5, in extraction results of Fig. 6, the mis-segmentation is too large to describe, the main reason is that the unsupervised segmentation method has too great blindness. It takes no account of scale and texture, so the unsupervised segmentation is basically an invalid method. The qualitative analysis considers that the proposed method has the best result, followed by the direct supervised method, the unsupervised method is basically invalid.

The evaluation of the results can be divided into two types: qualitative and quantitative criteria. Qualitative evaluation methods are combined with the application and visual effects, it has made a qualitative evaluation above. Quantitative evaluation methods also have two kinds of metrics: based on the amount error of pixels and based on the boundary pixel distance error. In order to accurately compare the recognition accuracy, this paper uses two kinds of quality evaluation criteria: based on region and based on edge. Take Fig. 4 (a) as an example, the black border line region in the picture is the extraction results by the proposed algorithm, whose surrounded areas are automatic measurement areas,

supposed as  $M_1$ ; and the white border lines boundaries are the measurement results by artificial visual interpretation, whose surrounded areas are the actual measurement areas, supposed as  $M_2$ . Both overlay, we can determine the pixels' range and location of misclassification. Quantitative evaluation methods: 1. The method based on the amount error of regional pixels, which is the pixel difference values that is the comparison between the experimental results and measurement results, the regional pixel error rate is  $Err = |M_1 - M_2|/M_2$ , when  $Err$  is more than 1, the value of  $Err$  equals 1. The correct rate is  $Cor = 1 - Err$ ; 2. Pixel distance error evaluation methods based on boundary [13],  $F$  and  $Z$  separately represents the borders processed by the algorithm and measurement boundaries, then the minimum distance which each point from  $F$  to  $Z$  and the minimum distance which each point from  $Z$  to  $F$  constitutes two distributions the deviation  $D_Z^F$  and  $D_F^Z$ , their means  $\mu D_Z^F$  and  $\mu D_F^Z$ , and their variances  $\sigma D_Z^F$  and  $\sigma D_F^Z$  can be used to measure the deviation between  $F$  and  $Z$ . In order to facilitate the evaluation, separately take the average of the mean and deviation:  $\mu \bar{D} = (\mu D_Z^F + \mu D_F^Z)/2$  and  $\sigma \bar{D} = (\sigma D_Z^F + \sigma D_F^Z)/2$ , which can express the integrated deviation between  $F$  and  $Z$  in pixels. The smaller the value of  $F$  and  $Z$  is, the higher the degree of fit is achieved, namely the higher the accuracy of the boundary is. Precision comparison of experimental results is presented in Table 2.

Combined with the qualitative evaluation in Figs. 4–6 and quantitative analysis in Table 2, the experimental conclusions can be drawn. As a whole, factors affecting the recognition rate of the urban canopy include buildings shadow, the mixture of sporadic canopy and the texture of the background within complex building environments. The overall accuracy indicates that the more concentrated woody vegetation is, the higher recognition

**Table 2**  
Precision comparison of different extraction algorithm and actual measured results.

Image	Method	Region-based (pixel)		Edge-based (pixel)					
		Err (%)	Cor (%)	$\mu D_F^2$	$\sigma D_F^2$	$\mu D_Z^2$	$\sigma D_Z^2$	$\mu \bar{D}$	$\sigma \bar{D}$
Fig. 4a	The proposed	7.52	92.48	5.5471	5.2132	6.6414	8.6774	6.0943	6.9453
Fig. 5a	Direct supervised	42.57	57.43	3.9988	3.0534	20.6580	19.0148	12.3284	11.0341
Fig. 6a	Unsupervised	100	0	10.2575	6.3598	23.1963	16.5834	16.7269	11.4716
Fig. 4b	The proposed	4.69	95.31	4.6606	4.7209	7.3476	13.1757	6.0041	8.9483
Fig. 5b	Direct supervised	21.93	78.07	3.7887	4.6209	12.5944	15.7999	8.1915	10.2104
Fig. 6b	Unsupervised	37.52	62.48	3.7764	3.8157	14.2954	16.5191	9.0359	10.1674
Fig. 4c	The proposed	6.42	93.58	5.6639	5.2232	4.4383	3.1319	5.0511	4.1776
Fig. 5c	Direct supervised	17.36	82.64	3.7290	3.3733	4.9573	5.6355	4.3431	4.5044
Fig. 6c	Unsupervised	44.94	55.06	7.3279	6.2403	8.9553	8.3530	8.1416	7.2966
Fig. 4d	The proposed	4.42	95.58	9.9980	15.7715	4.6100	3.0720	7.3040	9.4217
Fig. 5d	Direct supervised	19.36	80.64	4.8593	4.9270	8.9709	11.5746	6.9151	8.2508
Fig. 6d	Unsupervised	50.01	49.99	13.7332	15.1008	18.3548	19.8460	16.0440	17.4734

accuracy can be achieved while the more fragmented, the lower correct rate is, and less building blocked, the higher correct rate is. Comparing different extraction methods, the direct supervised segmentation method just takes account of the spectral multidimensional vector, no scale and texture, so its accuracy is much lower than the proposed method; the unsupervised segmentation method not only but also fails to consider the goal directness of the multidimensional vector of spectra, scale and texture. Therefore its error rate is basically 37.52%. In the Fig. 6a, the error rate of the unsupervised segmentation is unexpectedly up to 110.90%, so it is basically invalid to directly use the unsupervised segmentation method on target recognition. The proposed method is based on the assumption that only the spectral intensity (grey level) is considered, which yields the highest extraction accuracy. It can be seen that the scale and texture are the main factors of target recognition in the extraction of urban tree canopy and if considering the spectral multidimensional vector, it will obtain a better result.

Accuracy analysis of the data in Table 2 shows that the correct rate of the automatic extraction for urban canopy is above 90% processed by the proposed method, and the average accuracy rate is 93.90%; edge-based mean variance and the average deviation variance is 6.11 and 7.37 pixels, separately equal to 3.06 and 3.69 m in actual ground respectively, which can meet the actual application requirements under the data updating on the amount of urban greening and the rapid development of urban, such as the objective evaluation of the degree on change of afforestation scale, planning, management, environmental protection.

## 5. Conclusions

This paper studies the target recognition model taking account of the spectrum, texture and scale together. To highlight the research model mechanism, reduce the difficulty for describing the model, only consider spectral intensity information of the spectral features. Obviously, if multispectral or high spectral information is provided, a better effect can be expected due to the richness of texture information.

The factors influencing the tree canopy recognition include: the shadow of buildings and structures, the weakness of fragmentary tree canopy features which lead to the mixture of the spectrum and the texture. Two solutions can solve the former problem: one is using shadow detection and correction technique to eliminate or compensate the spectral mixture caused by the shadow [14], further to improve the recognition rate; the other one is using LiDAR data to correct the true-ortho image, then, recognize the multi-scale clustering for the true-ortho image to increase the recognition rate. The latter way needs to be exclusively studied on small target protection and identification technology.

The related issues that need to continue to study include: (1) mean shift is a kind of statistical method, in the image blocking, the regional areas of target and background determine the scale of relative mode in feature space, if some targets have large difference of areas with the background, it may lead to the failure of target recognition, this is one reason for the lower recognition rate in fragmentary pieces of tree canopy, hence, the study on small target recognition is needed; (2) the effectiveness of the mean shift on feature space decomposition depends on the separability of the feature space, so it needs more effectively technical study on the feature measurement method of the texture difference between urban tree canopy and background. For example, building the multi-scale feature space based on multispectral or hyperspectral data can enhance the separability of feature space. In this case, the feature transfer strategy, extraction of integrated texture feature and so on, has a certain complexity, which will raise new problems to be studied; (3) parameters involved in this paper such as bandwidth, density are empirical threshold, which are a common problem in pattern recognition by using mean shift technique, its self-adaptive confirmation technique needs a further study.

## Acknowledgements

This work was supported in part by National Natural Science Foundation of China under Grant 40971217 and 61172174, National Science & Technology Specific Projects under Grant 2012YQ16018505 and 2013BAH42F03, Open Fund for a National Key Laboratory in Geographic Information Engineering (No. SKLGIE2013-M-3-2).

## References

- [1] M.R. Shi, Z.X. Zhou, Q.J. Xie, Y.Y. Wang, J. Qin, Research progress on urban forest, *J. Landscape Res.* 2 (3) (2010) 30–34.
- [2] J. Müllerová, J. Pergl, P. Pyšek, Remote sensing as a tool for monitoring plant invasions: testing the effects of data resolution and image classification approach on the detection of a model plant species *Heracleum mantegazzianum* (giant hogweed), *Int. J. Appl. Earth Obs. Geoinf.* 25 (2013) 55–65.
- [3] X.L. Qin, Z.Y. Li, H.R. Yi, Extraction method of tree crown using high-resolution satellite image, *Remote Sens. Technol. Appl.* 20 (2) (2005) 228–232.
- [4] A. Chang, Y. Eo, Y. Kim, Y. Kim, Identification of individual tree crowns from LiDAR data using a circle fitting algorithm with local maxima and minima filtering, *Remote Sens. Lett.* 4 (1) (2013) 29–37.
- [5] S.W. Myint, N.S.N. Lam, J.M. Tylor, Wavelets for urban spatial feature discrimination: comparisons with fractal, spatial autocorrelation, and spatial co-occurrence approaches, *Photogr. Eng. Remote Sens.* 70 (7) (2004) 803–812.
- [6] S. Wang, Y. Xia, L.C. Jiao, Mean shift based adaptive texture image segmentation method, *J. Softw.* 21 (6) (2010) 1451–1461.
- [7] Y.Z. Cheng, Mean shift, mode seeking, and clustering, *Pattern Anal. Mach. Intell.* 17 (8) (1995) 790–799.
- [8] C. Dorin, M. Peter, Distribution free decomposition of multivariate data, *Pattern Anal. Appl.* 2 (1) (1999) 22–30.

- [9] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, *Pattern Anal. Mach. Intell.* 24 (5) (2002) 603–619.
- [10] S. Sheather, M. Jones, A reliable data-based bandwidth selection method for kernel density estimation, *J. R. Stat. Soc. Ser. B* 53 (3) (1991) 683–690.
- [11] D. Comaniciu, V. Ramesh, P. Meer, The variable bandwidth mean shift and data-driven scale selection, in: *Proceedings 8th IEEE International Conference on Computer Vision*, Vancouver, 2001, pp. 438–445.
- [12] K.H. Liang, T. Tjahjadi, Adaptive scale fixing for multiscale texture segmentation, *IEEE Trans. Image Process.* 15 (1) (2006) 249–256.
- [13] Q. Huang, B. Dom, Quantitative methods of evaluating image segmentation, in: *Proceedings of the 1995 International Conference on Image Processing*, IEEE Computer Society, Washington, 1995, pp. 53–56.
- [14] Z. Pu, P. Liu, L. Yang, J. Bai, Shadow rectification of remote sensing images in information extraction for urban green space, *J. Beijing For. Univ.* 31 (2) (2009) 80–85.