International Journal of Remote Sensing
Vol. 33, No. 8, 20 April 2012, 2641–2656

Taylor & Francis
Taylor & Francis Group

Technical Note

A multiscale urban complexity index based on 3D wavelet transform for spectral–spatial feature extraction and classification: an evaluation on the 8-channel WorldView-2 imagery

XIN HUANG* and LIANGPEI ZHANG
The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, Hubei 430079, PR China

(*Received 2 September 2010; in final form 10 August 2011*)

The three-dimensional wavelet transform (3D-WT) processes a multispectral remotely sensed image as a cube and hence it is able to simultaneously represent variation information in joint spectral–spatial feature space. The urban complexity index (UCI) built on the 3D-WT is defined by comparing the amount of spectral and spatial variation, since natural features have relatively smaller spatial changes than spectral changes but urban areas show more variation in the spatial domain. The calculation of the UCI is subject to the selection of window sizes; therefore, in this study, a multiscale UCI (M-UCI) is proposed by integrating the UCI features in different moving windows and decomposition levels. The performance of the M-UCI was evaluated on two WorldView-2 data sets over urban and suburban areas, respectively. Experimental results showed that the M-UCI was effective in integrating multiscale information contained in different windows and gave higher accuracies than the single-scale UCI. In experiments, the proposed M-UCI was compared with a pixel shape index (PSI), which is a texture measure extracted from the spatial domain alone. It was revealed that the PSI was more effective for the classification of urban areas than natural landscapes, whereas the M-UCI was applicable for both urban and natural areas since it represented the joint spectral–spatial domains.

1. Introduction

High spatial resolution remotely sensed imagery is able to provide rich and detailed ground information, and this kind of data source has received more and more attention for applications such as environment assessment, territorial planning, land cover and land use mapping at the regional level. Although the high-resolution images show huge application potential to Earth observation, their feature extraction and classification methods are quite different from low or median spatial resolution images. The traditional classification methods are related to the pixel-wise processing and spectral analysis (Landgrebe 2003). However, in recent years it has been largely agreed that the spectral information alone is not adequate for the classification of high-resolution images. Therefore, researchers proposed to extract spatial features for complementing the spectral feature space and discriminating spectrally similar classes, such as the co-occurrence matrix (Puissant *et al*. 2005), pixel shape index (PSI) (Huang *et al*.

*Corresponding author. Email: huang_whu@163.com

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2012 Taylor & Francis
http://www.tandf.co.uk/journals
http://dx.doi.org/10.1080/01431161.2011.614287

2007a), morphological profiles (Pesaresi and Benediktsson 2001) and segmentation-based classification (Huang and Zhang 2008). It should be noted that the existing textural and structural features are related to the spatial domain alone, but few algorithms refer to feature extraction from the joint spectral–spatial domains.

The three-dimensional wavelet transform (3D-WT) is of interest in this study since it is able to describe the dependency of joint spectral–spatial domains. The 3D-WT is implemented in separable fashion by performing a 1D wavelet transform (1D-WT) separately in the spatial-row, spatial-column and spectral directions. The 3D-WT has been applied to video compression (Lewis and Knowles 1990, Levy and Wilson 1999) and texture analysis of biomedical images (Pinnamaneni *et al*. 2001, Saladi *et al*. 2001). However, most applications of the 3D-WT to remote sensing refer to image compression (Fowler and Rucker 2007, Hou and Liu 2008). Very recently, Yoo *et al*. (2009) proposed an urban complexity index (UCI) based on energy parameters of 3D wavelet coefficients for the discrimination of urban and natural areas. The basic idea of a UCI is that natural features (e.g. water, forest, grass and soil) have relatively smaller spatial changes than spectral changes, while urban areas (e.g. buildings, roads) have more variations in spatial domain than spectral domain. The UCI was originally implemented using a moving window with a fixed size, and hence its application is subject to selection of window sizes. Consequently, in this study a multiscale UCI (M-UCI) is proposed by integrating the 3D wavelet coefficients in different windows and decomposition levels. Moreover, the M-UCI is evaluated on the 8-channel WorldView-2 images that show rich information in both spectral and spatial domains. Based on the WorldView-2 experiments, we focus on the following research questions.

- Does the M-UCI give better results than the single-scale UCI by integrating the multiscale information in different windows and decomposition levels?
- Does the spectral–spatial joint feature index outperform the feature index that is extracted from the spatial domain alone (e.g. PSI; Huang *et al*. 2007a)?
- What is the effect of different mother wavelets on the performance of the UCI and M-UCI?
- Does the UCI improve the classification of 8-channel WorldView-2 images?

2. Methodology

2.1 *3D wavelet transform*

The WT is an effective tool for signal analysis at various scales and shifts. A WT can be constituted by a series of scaling functions $\varphi(x)$ and wavelet functions $\psi(x)$:

$$\varphi(x) = 2^{\frac{j}{2}}\varphi\left(2^j x - k\right), \tag{1}$$

$$\psi(x) = 2^{\frac{j}{2}}\psi\left(2^j x - k\right), \tag{2}$$

where j and k represent the scaling and translation parameters, respectively. In practice, the scaling functions are considered as low-pass filters, whereas wavelet functions are viewed as high-pass filters. Similarly to the construction of the 2D discrete WT (Mallat 1989), the 3D-WT can be constructed by a tensor product as follows:

$$\mathrm{I}^{(x,y,z)} = \left(\mathrm{L}^x \oplus \mathrm{H}^x\right) \otimes \left(\mathrm{L}^y \oplus \mathrm{H}^y\right) \otimes \left(\mathrm{L}^z \oplus \mathrm{H}^z\right) = \begin{cases} \mathrm{L}^x\,\mathrm{L}^y\,\mathrm{L}^z \\ \oplus\, \mathrm{L}^x\,\mathrm{L}^y\,\mathrm{H}^z \\ \oplus\, \mathrm{L}^x\,\mathrm{H}^y\,\mathrm{L}^z \\ \oplus\, \mathrm{L}^x\,\mathrm{H}^y\,\mathrm{H}^z \\ \oplus\, \mathrm{H}^x\,\mathrm{L}^y\,\mathrm{L}^z \\ \oplus\, \mathrm{H}^x\,\mathrm{L}^y\,\mathrm{H}^z \\ \oplus\, \mathrm{H}^x\,\mathrm{H}^y\,\mathrm{L}^z \\ \oplus\, \mathrm{H}^x\,\mathrm{H}^y\,\mathrm{H}^z \end{cases}, \tag{3}$$

where $\oplus$ denotes the space direct sum, and **L** and **H** represent the low- and high-pass filters along the x, y and z axes, respectively. In practice, x and y directions denote the spatial coordinates of an image, and z is the spectral axis. Equation (3) shows that one-level 3D-WT decomposition of an image cube generates eight subbands: LLL, LLH, LHL, LHH, HLL, HLH, HHL and HHH. For example, LLH represents the subband with low-frequency information in the spatial-row and spatial-column directions and high-frequency information in the spectral direction. In the implementation, the eight subbands involve the following eight wavelets, which are constructed through the tensor product of a 1D scaling function $\varphi(x)$ and a 1D wavelet function $\psi(x)$:

$$\begin{aligned} \Phi_{\mathrm{LLL}}(x,y,z) &= \varphi(x)\varphi(y)\varphi(z), \quad & \Psi_{\mathrm{LLH}}(x,y,z) &= \varphi(x)\varphi(y)\psi(z), \\ \Psi_{\mathrm{LHL}}(x,y,z) &= \varphi(x)\psi(y)\varphi(z), \quad & \Psi_{\mathrm{LHH}}(x,y,z) &= \varphi(x)\psi(y)\psi(z), \\ \Psi_{\mathrm{HLL}}(x,y,z) &= \psi(x)\varphi(y)\varphi(z), \quad & \Psi_{\mathrm{HLH}}(x,y,z) &= \psi(x)\varphi(y)\psi(z), \\ \Psi_{\mathrm{HHL}}(x,y,z) &= \psi(x)\psi(y)\varphi(z), \quad & \Psi_{\mathrm{HHH}}(x,y,z) &= \psi(x)\psi(y)\psi(z). \end{aligned} \tag{4}$$

A single-level decomposition of 3D-WT is demonstrated in figure 1, where the downward arrows stand for the down-sampling of the signals. It can be seen that the wavelets are implemented in the spatial-row, spatial-column and spectral directions, and hence 3D-WT provides an effective approach for joint spectral–spatial feature extraction.

2.2 *Multiscale urban complexity index*

Due to the ability of the 3D-WT for representation of the variation information in both spectral and spatial domains, Yoo *et al*. (2009) proposed a UCI based on the energy parameters of the 3D wavelet coefficients. The basic idea of the UCI is that in remotely sensed imagery, (1) natural features such as water, trees, grass and bare soil have relatively smaller spatial changes than spectral changes, while (2) the urban features including buildings and roads refer to more variability in row and column directions than the spectral direction. The eight subbands resulting from the 3D-WT can be separated into three categories:

- approximation – LLL;
- representations of spectral variation (**H)– LLH, LHH, HLH, HHH; and
- representations of spatial variation (**L)– LHL, HLL, HHL;

where the **H subbands denote the spectral variation since the high-pass filter is used in the spectral direction, and the **L subbands stand for the spatial variation as they

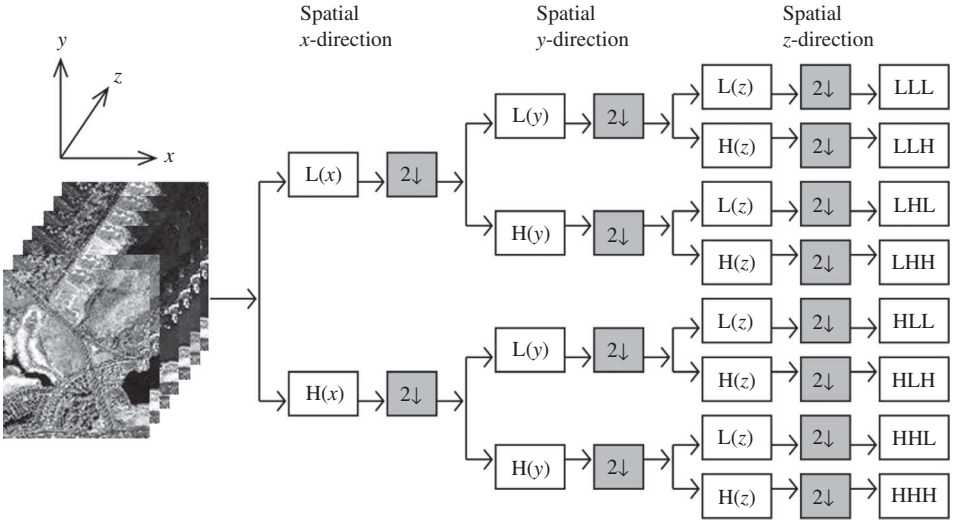

Figure 1. One-level decomposition of the 3D-WT for multispectral imagery (H(x) means high-frequency filtering in the x-direction and the down arrows mean the down-sampling of signals).

represent horizontal and vertical information in the spatial domain and low-frequency information in the spectral domain. Consequently, the UCI is defined as the sum of all H components (**H) divided by the sum of all L components (**L):

$$\mathrm{UCI} = \frac{E(\mathrm{HLL}) + E(\mathrm{LHL}) + E(\mathrm{HHL})}{E(\mathrm{LLH}) + E(\mathrm{LHH}) + E(\mathrm{HLH})}, \tag{5}$$

where the function $E(\cdot)$ denotes the energy or L2-norm of the subband f:

$$E(f) = \sum_i \sum_j \sum_m (f(i,j,m))^2. \tag{6}$$

In equation (6), i, j and m stand for the coordinates of x, y and z directions in a subband f(HLL, etc.), respectively.

The UCI is calculated based on a moving window that is centralized by each pixel in an image cube; therefore, the selection of window sizes becomes a key problem when the UCI is applied to feature extraction and classification for remotely sensed imagery. Furthermore, it should be noted that equations (3) and (5) only consider one-level wavelet decomposition; however, in fact, the multilevel wavelet coefficients can be obtained by iteratively decomposing the low-frequency LLL subband. Consequently, the scale of the UCI is affected by two aspects: window size w and decomposition level l. Based on the above analysis, we propose a M-UCI by averaging the feature values at different windows and levels:

$$\text{M-UCI} = \frac{1}{N}\sum_{w,l} \mathrm{UCI}(w,l) = \frac{1}{N}\sum_{w,l} \left\{ \frac{E(\mathrm{HLL}) + E(\mathrm{LHL}) + E(\mathrm{HHL})}{E(\mathrm{LLH}) + E(\mathrm{LHH}) + E(\mathrm{HLH})} \right\}_{w,l}, \tag{7}$$

where N is the number of scales considered in the M-UCI, and UCI(w, l) denotes the UCI feature obtained by window size w and decomposition l. The calculation steps of the M-UCI are demonstrated in figure 2. The termination of the decomposition is determined by comparing the current decomposition level l and the maximum level $l_{\max}$, which is defined as

$$l_{\max} = \min(\log_2 w, \log_2 b), \tag{8}$$

where w is the size of the moving window and b denotes the number of available bands. For instance, a 4×4 local window ($w = 4$) in a WorldView-2 image ($b = 8$), resulting in a $4 \times 4 \times 8$ 3D structure, has at most a two-level 3D wavelet decomposition ($l_{\max} = 2$), since the subband at level 2 is equivalent to a $1 \times 1 \times 2$ 3D structure that cannot be decomposed any more.

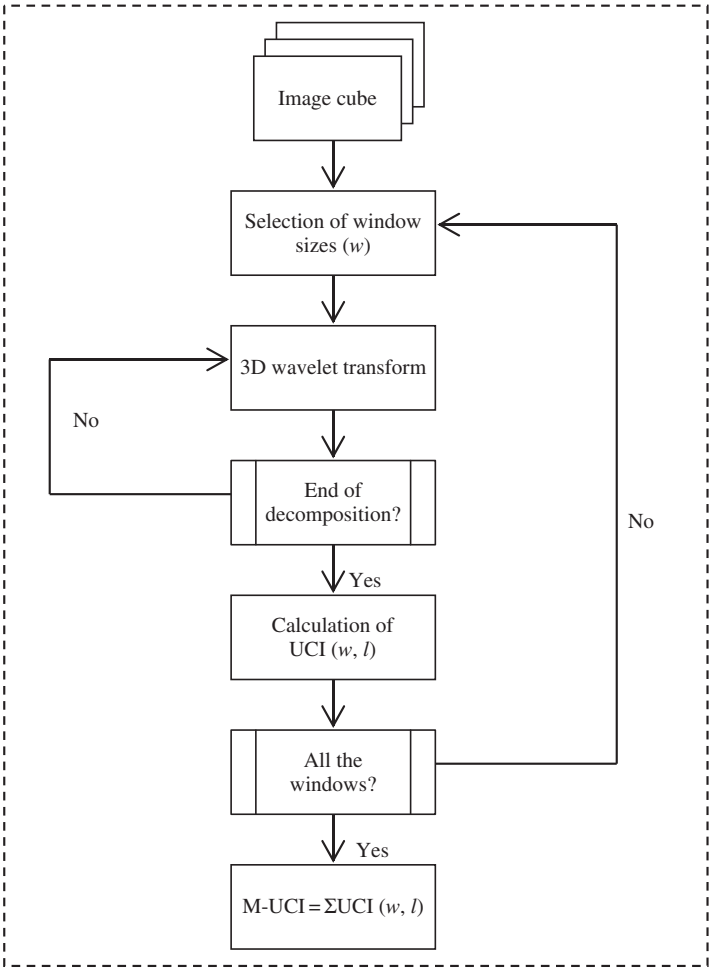

Figure 2. Processing flow of the M-UCI.

A graphical example of the M-UCI is shown in figure 3, from which we can obtain the following observations.

(1) UCI is effective for the indication of urban structures such as buildings and roads, which have larger brightness value than natural landscapes such as trees, grass, water and soil in the UCI feature images.

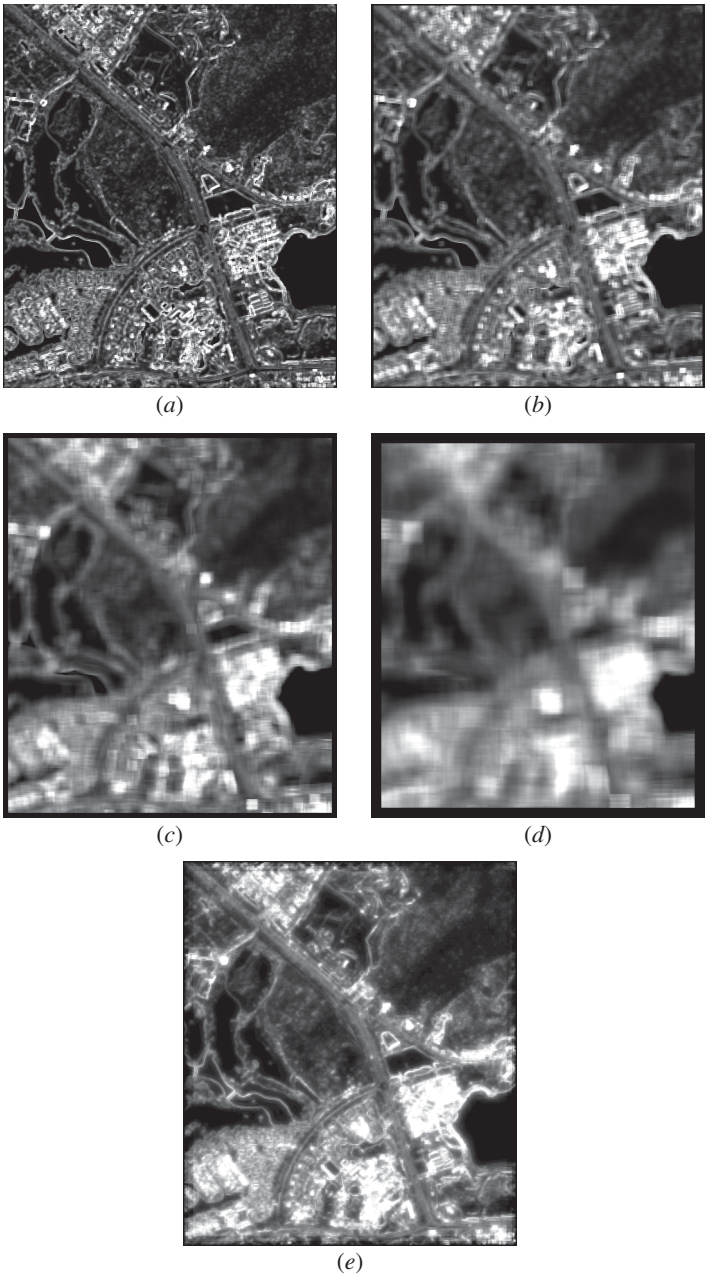

Figure 3. A graphical example of the M-UCI for the WorldView-2 image covering a suburban region in Hainan Island, in the south of China: (*a*) UCI ($w = 4$, $l = 1$); (*b*) UCI ($w = 8$, $l = 1$); (*c*) UCI ($w = 16$, $l = 1$); (*d*) UCI ($w = 32$, $l = 1$); and (*e*) M-UCI.

(2) As the window size increases, although the UCI feature values in the hardcore of the urban regions become more steady and insensitive to local variation, the values in the boundaries become more uncertain because a larger moving window is able to blur the end edges of local image structures.
(3) It can be clearly seen that the M-UCI feature image integrates the UCI information extracted from different window sizes and at the same time alleviates the negative effects resulting from large windows. Furthermore, previous work has shown that there was not the so-called optimal window in such a complex urban landscape, especially for the high-resolution images (Huang *et al*. 2007b). Therefore, from this point of view, the M-UCI is more reasonable than the single-scale version.

3. Experiments and analysis

3.1 *Data sets and experimental set-up*

In order to validate the effectiveness of the proposed M-UCI, experiments are conducted on the WorldView-2 images. The WorldView-2 satellite, launched on 8 October 2009, is able to provide eight multispectral bands with 2 m spatial resolution. It is the first high-resolution satellite to provide eight narrowly focused multispectral channels ranging from blue to near-infrared. In addition to the standard colours, four new bands (coastal, yellow, red edge and near-infrared 2) are included. The spectral information of WorldView-2 is summarized in table 1.

WorldView-2 images are very suitable for the 3D-WT, since they provide rich information in both spectral and spatial domains and the eight multispectral bands support multilevel decomposition of the 3D-WT in the spectral direction. Two WorldView-2 data sets are used in experiments. The first test image (figure 4(*a*)) covers a suburban region in Hainan Island, in the south of China, and the second test image (figure 4(*c*)) shows a typical urban area in Hangzhou City, in the east of China. The experiments are conducted on two images that present, respectively, the natural and urban landscapes because it is possible to evaluate the ability of the UCI for feature extraction in joint spectral–spatial domains. To this aim, in experiments, the UCI is compared with our previous algorithm, the PSI (Huang *et al*. 2007a), which represents the feature in the spatial domain alone. The test samples for the two data sets are shown in figure 4(*b*) and (*d*), respectively, and the numbers of training and test samples are listed in table 2.

Table 1. Spectral information of WorldView-2 imagery.

Spectral bands		Wavelength (nm)
Conventional spectral channels	Panchromatic	450 – 800
	Red	630 – 690
	Green	510 – 580
	Blue	450 – 510
	NIR1	770 – 895
New spectral channels	Yellow	585 – 625
	Coastal	400 – 450
	Red edge	705 – 745
	NIR2	860 – 1040

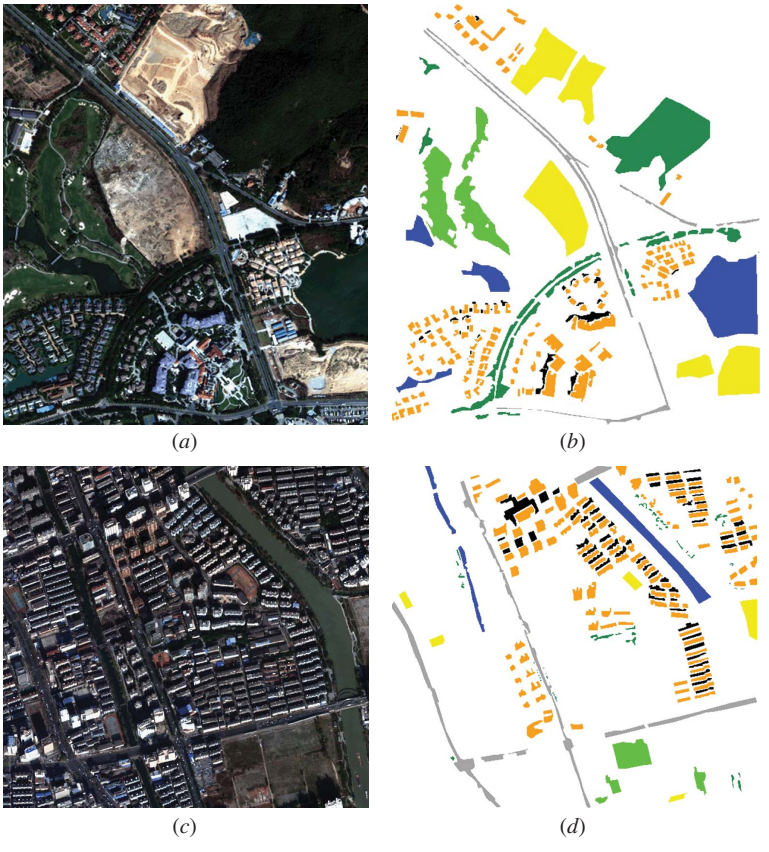

Figure 4. (*a*) and (*b*) are test image 1 (Hainan Province) and test samples, respectively; (*c*) and (*d*) are test image 2 (Hangzhou City) and its test samples, respectively. The legend is shown in table 2.

Table 2. Numbers of the training and test samples for the two test data sets.

	Test image 1 (Hainan)		Test image 2 (Hangzhou)	
Land cover classes	Training	Test	Training	Test
Roads	56	4 459	48	8 800
Buildings	59	11 148	55	17 890
Water	53	11 209	52	5 680
Shadow	52	1 427	49	7 075
Trees	52	14 086	55	1 047
Grass	51	7 417	58	3 359
Soil	55	17 541	42	2 651
Total	378	67 287	359	46 502

Four window sizes, $w = 4$, 8, 16 and 32, are used in both test data sets, considering that the chosen local windows correspond to the characteristics of the land cover classes and the spatial resolution. According to equation (8), the maximum of

decomposition levels are $l_{max} = 2$, 3, 3 and 3 for the four window sizes, respectively. In experiments, the generated UCI feature images concatenated with the eight spectral bands are fed into a support vector machine (SVM) for classification. The SVM is used due to its swift learning pace and adaptability to the hybrid spectral–spatial feature space.

3.2 *Validation of the M-UCI*

The classification accuracies of the single-scale UCI and M-UCI are reported in table 3 for the Hainan and Hangzhou data sets, respectively. The overall accuracy (OA) extracted from the confusion matrix is used for accuracy comparison. From the experimental results of the two test images, we can obtain the following comments.

(1) The M-UCI outperformed all the single-scale operators, which shows that the proposed M-UCI is effective in integrating the information contained in different moving windows.
(2) The addition of the UCI images significantly improved the accuracy of spectral classification with only multispectral bands as input features.
(3) In most cases, the first decomposition level ($l = 1$) gave the highest classification accuracies because the first level contains the majority of the energy of wavelet coefficients. Consequently, in this study, the M-UCI was calculated based on level 1, that is, l was set to 1 in equation (7).

The class-specific accuracies of the UCI with different window sizes are shown in table 4 in order to analyse the effects of window sizes in detail. The F-measure of the producer's and user's accuracies (Shah *et al*. 2010) based on the confusion matrix was used to describe the class-specific accuracies:

$$A(c) = \frac{2P(c)U(c)}{P(c) + U(c)}, \tag{9}$$

Table 3. Comparison between single-scale and M-UCI for the two data sets.

	Overall accuracy (%)		
Window sizes	Level 1	Level 2	Level 3
WorldView-2 Hainan image (suburban)			
$w = 4$	89.3	84.6	NA
$w = 8$	92.0	85.1	84.5
$w = 16$	92.6	91.9	83.3
$w = 32$	91.5	92.1	78.7
Spectral only	84.7		
M-UCI	93.6		
WorldView-2 Hangzhou image (urban)			
$w = 4$	81.8	78.4	NA
$w = 8$	92.4	90.0	92.2
$w = 16$	92.4	91.1	90.9
$w = 32$	86.3	86.0	88.9
Spectral only	77.6		
M-UCI	93.6		

Note: NA, not available.

Table 4. Class-specific accuracies (%) of M-UCI for the two data sets.

Classes	Features					
	Spectral	$w=4$	$w=8$	$w=16$	$w=32$	M-UCI
WorldView-2 Hainan image (suburban)						
Roads	71.3	70.1	73.1	76.1	73.8	74.2
Buildings	62.8	75.8	83.3	86.2	78.8	86.2
Water	99.2	99.5	97.4	96.1	99.3	99.0
Shadow	85.5	82.8	70.7	65.4	83.9	78.1
Trees	97.3	97.3	97.6	97.5	97.6	97.4
Grass	96.8	96.8	97.3	97.1	97.2	96.9
Soil	78.6	88.1	94.4	96.2	93.0	96.4
Ave	84.5	87.2	87.7	87.8	89.1	89.7
WorldView-2 Hangzhou image (urban)						
Roads	87.2	87.6	88.7	90.0	77.3	88.0
Buildings	71.0	77.9	93.3	94.4	89.9	94.4
Water	95.8	95.9	96.7	90.7	86.9	97.3
Shadow	95.8	95.8	96.6	93.0	95.3	97.2
Trees	90.5	88.6	90.1	89.5	77.3	90.8
Grass	89.9	89.5	93.0	96.9	93.1	95.0
Soil	23.5	40.2	81.2	85.0	68.0	88.9
Ave	79.1	82.2	91.4	91.3	84.0	93.1

where $P(c)$ and $U(c)$ are the producer's and user's accuracies of class c, respectively and $A(c)$ represents all the accuracies in table 4.

By analysing the results of the WorldView-2 Hainan data set shown in table 4, it can be seen that the spectral-only classification gave satisfactory accuracies for natural landscapes such as water, shadow, trees and grass, but relatively poor accuracies for man-made structures such as roads and buildings. Moreover, the accuracy of the class of soil was not high since the spectral information alone is not adequate for discrimination between soil, buildings and roads due to their very similar spectral response. On the other hand, when the UCI images were integrated in the feature space, the accuracies of urban structures were significantly enhanced. By comparing the performance of UCI in different scales, it can be found that it is difficult to determine the optimal window since the window sizes of 8 and 16 did not give satisfactory results to shadow, and the window size of 32 was not accurate for buildings and soil. In this context, the performance of the M-UCI showed that the multiscale feature fusion was an effective approach for information combination in different windows, since the M-UCI gave satisfactory results to most of classes and achieved the highest average accuracy (Ave).

By looking at the results of the Hangzhou test image in table 4, a similar conclusion was drawn. The spectral classification gave poor accuracies to buildings and soil, and the exploitation of UCI features enhanced the spectral classification significantly as the UCI was able to distinguish the spectrally similar urban structures. The class-specific accuracies revealed that classifications of buildings and soil for window size 4, soil for window size 8, water and shadow for window size 16, and roads, water, trees and soil for window size 32 were not accurate. Once again, the M-UCI achieved the highest average accuracy (93.1%), and the M-UCI gave higher accuracies for all the information classes compared to the spectral-only classification.

3.3 *Comparison with PSI*

In order to evaluate the ability of the M-UCI for spectral–spatial feature extraction, it was compared with our previous algorithm (Huang *et al*. 2007a), the PSI, which is an urban feature index extracted from the spatial domain alone. Classification maps of the Hainan and Hangzhou data sets are presented in figures 5 and 6, respectively, for visual inspection. The accuracies of spectral classification, the PSI and the M-UCI are compared in table 5. In the suburban data set, where the image scene is dominated by natural landscapes, the addition of the PSI in the feature space resulted in a slight accuracy decrease compared to the spectral classification, but exploitation of the M-UCI improved the OA by 10.4%. In the urban data set, where the image scene is dominated by complex urban structures, both the PSI and M-UCI improved the spectral classification significantly, and the improvements were 11.6% and 16%, respectively. By comparing the different performances of the PSI and M-UCI in different image

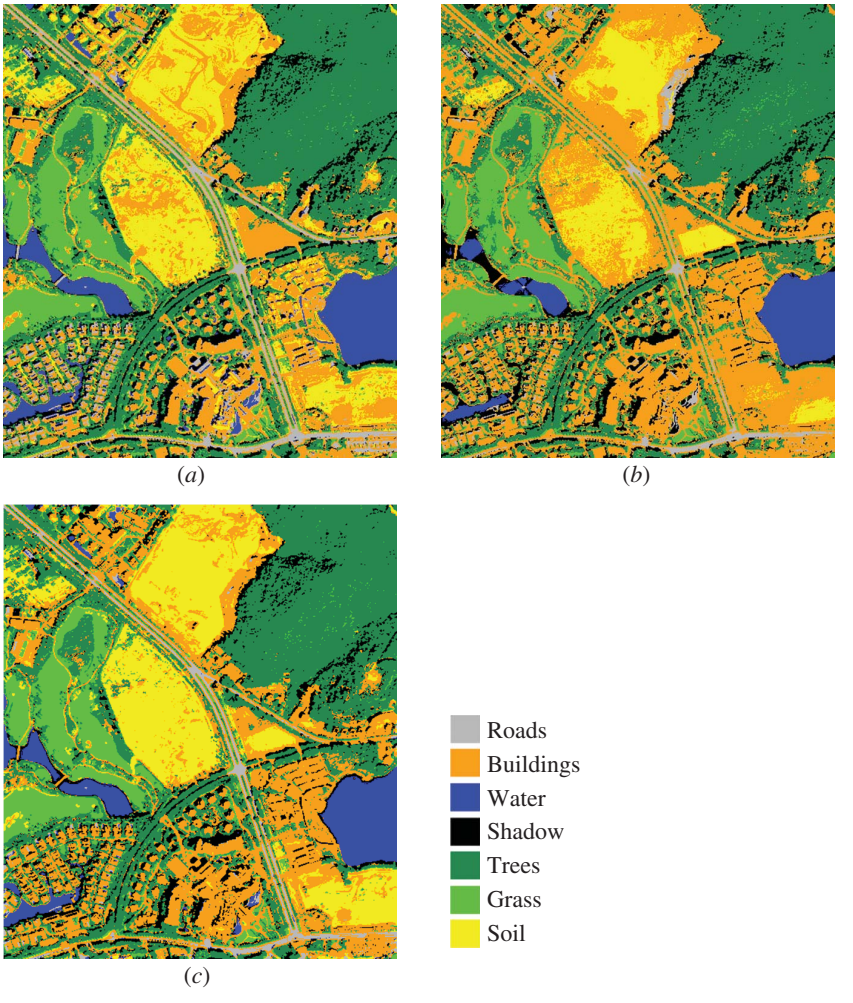

Figure 5. Classification maps of the WorldView-2 Hainan data set: (*a*) spectral classification, (*b*) PSI and (*c*) M-UCI.

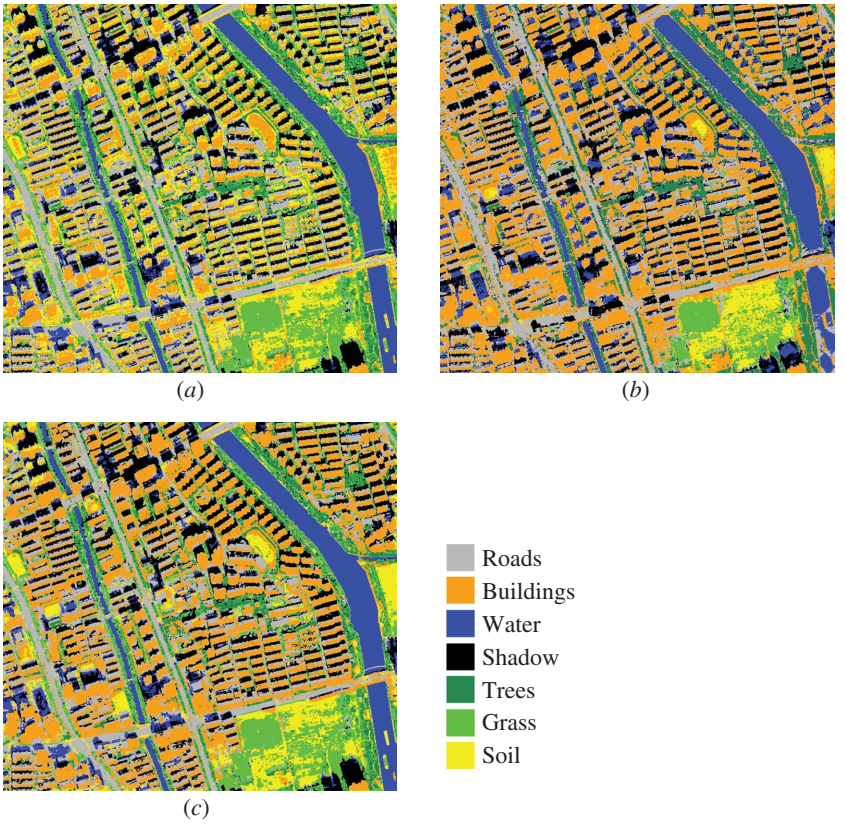

Figure 6. Classification maps of the WorldView-2 Hangzhou data set: (*a*) spectral classification, (*b*) PSI and (*c*) M-UCI.

Table 5. Comparison of spectral classification, the PSI and the M-UCI for the two WorldView-2 data sets.

	Hainan data set (Suburban)			Hangzhou data set (Urban)		
Classes	Spectral	PSI	M-UCI	Spectral	PSI	M-UCI
Roads	71.3	45.8	74.2	87.2	86.0	88.0
Buildings	62.8	68.3	86.2	71.0	91.6	94.4
Water	99.2	97.5	99.0	95.8	88.9	97.3
Shadow	85.5	69.6	78.1	95.8	88.9	97.2
Trees	97.3	95.7	97.4	90.5	91.5	90.8
Grass	96.8	94.3	96.9	89.9	94.4	95.0
Soil	78.6	80.4	96.4	23.5	74.2	88.9
OA	84.7	83.2	93.6	77.6	89.2	93.6

scenes, it can be revealed that the PSI is more applicable in urban areas than natural areas because it highlights the spatial variation that is more sensitive to urban structures. On the other hand, it can be seen that the M-UCI is applicable for both urban and natural areas since it is able to represent the spectral and spatial features at the same time.

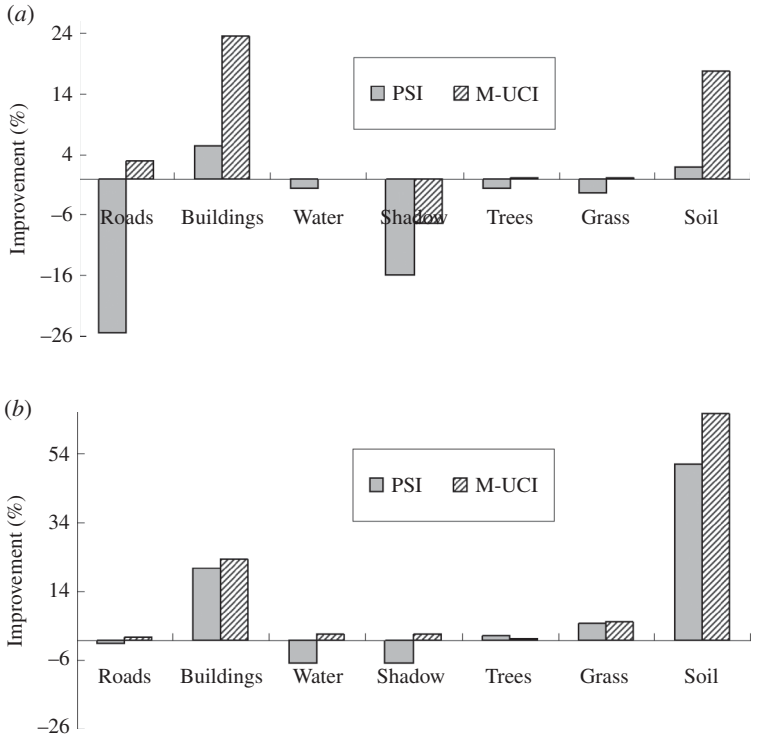

Figure 7. The positive and negative increments of class-specific accuracies for the PSI and M-UCI compared to the spectral classification for (*a*) Hainan data set (suburban) and (*b*) Hangzhou data set (urban).

The positive and negative increments of class-specific accuracies for the PSI and M-UCI compared to the spectral classification are demonstrated in figure 7. It can be observed that the M-UCI is able to enhance the classification for most of the information classes including both urban and natural landscapes, whereas the PSI is more useful in urban areas and not very accurate for grass, trees, water and shadow.

3.4 *Discussion of wavelet basis*

The 3D-WT performs wavelet decomposition separately in the spatial and spectral directions; therefore, different wavelet basis functions can be used in different directions. In the above experiments, the Haar basis function was applied to both spectral and spatial domains since it was carried out in the original UCI algorithm (Yoo *et al*. 2009) and it proved to be efficient for high-resolution images (Myint *et al*. 2004). In this study, the effects of wavelet basis functions are discussed by implementing different wavelets on spectral and spatial directions. To this aim, the wavelet basis of the fourth-order Daubechies (db4) is considered in the comparative study since Haar is actually the first-order Daubechies wavelet (db1). Experimental results are compared in table 6, from which we can obtain some information about the selection of the wavelet basis for the UCI:

Table 6. Comparison of different wavelet basis functions (db1 and db4) for the M-UCI.

	Wavelet basis function					
	Spectral db1	Spatial db1	Spectral db1	Spatial db4	Spectral db4	Spatial db1
UCI ($w=4$)	89.3		85.8		84.9	
UCI ($w=8$)	92.0		84.8		91.3	
UCI ($w=16$)	92.6		93.3		92.6	
UCI ($w=32$)	91.5		91.4		91.3	
M-UCI	93.6		94.3		93.1	

(1) In the spatial direction: when increasing the number of vanishing moments from db1 to db4 in the spatial direction, the accuracies of the UCI with large window sizes ($w=16$, 32) increased, but accuracies of small window sizes ($w=4$, 8) decreased. This phenomenon showed that the selection of wavelet basis in the spatial domain was related to the window size. The wavelet basis with a large number of vanishing moments was more suitable for large windows, and vice versa.
(2) In the spectral domain: employment of the db4 wavelet basis in the spectral direction resulted in slightly lower accuracies. This observation signified that the Haar or db1 function was appropriate for an 8-channel WorldView-2 image because the number of bands is not large enough to decompose the image in band direction by a function with a large value of vanishing moments.

4. Conclusion

The traditional UCI was implemented on a moving window with a fixed size. Therefore, the multiscale extension of the UCI is investigated in this study. The main advantage of the UCI compared to other spatial or textural measures lies in that it is able to simultaneously represent the spectral and spatial information by processing a multispectral image as a cube. The basic idea of the M-UCI is based on the fact that there is no single window size that would adequately characterize the complex and multiscale textural conditions present in high-resolution remotely sensed images (Puissant *et al*. 2005, Huang *et al*. 2007b). Specifically, the M-UCI is achieved by integrating the wavelet coefficients in different window sizes and decomposition levels. The proposed M-UCI was evaluated on the 8-channel WorldView-2 images that were of interest in this study because they provided rich information in both spectral and spatial domains and their 8 channels supported multilevel 3D wavelet decomposition.

Based on our experiments, the research questions proposed in the introduction can be answered as follows.

(1) The M-UCI gave better results than the single-scale UCI because the former was more suitable for the multiscale characteristics of high-resolution imagery. In addition, it was found that one-level decomposition of the 3D-WT was accurate enough for the construction of the UCI and M-UCI since the first decomposition level contained the majority of energy of wavelet coefficients.

This conclusion was supported by the two experiments, where the UCI generated on the second or third decomposition levels only gave marginal accuracy increments compared to the spectral classification.

(2) Based on the comparative experiments between the M-UCI and PSI, it was revealed that the traditional spatial index (e.g. PSI) was effective for the classification of urban structures but not for natural landscapes, whereas the M-UCI was applicable for both urban and natural areas because it was able to describe the joint spectral–spatial domains.

(3) By substituting the Haar or db1 basis function with the db4, which has larger number of vanishing moments, separately in spectral and spatial directions, it was revealed that in the spatial domain, the wavelet basis function should be selected according to its number of vanishing moments and the window size of the UCI. On the other hand, in the spectral direction, it was found that the Haar or db1 basis function was fine enough for the construction of the UCI from an 8-band WorldView-2 image.

(4) The experiments conducted on two WorldView-2 data sets over urban and natural areas showed that, although the WorldView-2 images provided additional four spectral bands than the traditional 4-channel high-resolution imagery, the M-UCI significantly enhanced the OA by 8.9% and 16.0% for Hainan (natural) and Hangzhou (urban) images, respectively. This phenomenon reveals that, although recently more spectral bands have become available for high spatial resolution imagery by courtesy of development of imaging techniques, spatial information is needed for accurate interpretation of high-resolution imagery. Furthermore, the joint spectral–spatial feature extraction approach such as the M-UCI points out an important research orientation for classification of multi-/hyperspectral imagery with high spatial resolution.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 41101336 and 41061130553, the Fundamental Research Funds for the Central Universities under Grant No. 3101016 and the LIESMARS Special Research Funding.

References

Fowler, J.E. and Rucker, J.T., 2007, 3D wavelet-based compression of hyperspectral imagery. In *Hyperspectral Data Exploitation: Theory and Applications*, C.-I. Chang (Ed.), pp. 379–407 (Hoboken, NJ: John Wiley).

Hou, Y. and Liu, G., 2008, Hyperspectral image lossy-to-lossless compression using the 3D embedded zeroblock coding algorithm. In *International Workshop on Earth Observation and Remote Sensing Applications*, 30 June–2 July 2008, Beijing (Piscataway, NJ: Institute of Electrical and Electronics Engineers), pp. 1–6.

Huang, X. and Zhang, L., 2008, An adaptive mean-shift analysis approach for object extraction and classification. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, pp. 4173–4185.

Huang, X., Zhang, L. and Li, P., 2007a, Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery. *IEEE Geoscience and Remote Sensing Letters*, **4**, pp. 260–264.

Huang, X., Zhang, L. and Li, P., 2007b, An adaptive multiscale information fusion approach for feature extraction and classification of IKONOS multispectral imagery over urban areas. *IEEE Geoscience and Remote Sensing Letters*, **4**, pp. 654–658.

Landgrebe, D., 2003, *Signal Theory Methods in Multispectral Remote Sensing* (Hoboken, NJ: John Wiley).

Levy, I.K. and Wilson, R., 1999, Three dimensional wavelet transform video compression. In *IEEE International Conference on Multimedia Computing and Systems*, 7–11 June 1999, Florence (Piscataway, NJ: Institute of Electrical and Electronics Engineers), pp. 924–928.

Lewis, A.S. and Knowles, G., 1990, Video compression using 3D wavelet transform. *Electronic Letters*, **26**, pp. 396–398.

Mallat, S.G., 1989, A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, pp. 674–693.

Myint, S.W., Lam, N.S.N. and Tyler, J., 2004, Wavelets for urban spatial feature discrimination: comparisons with fractal, spatial autocorrelation, and spatial co-occurrence approaches. *Photogrammetric Engineering and Remote Sensing*, **70**, pp. 803–812.

Pesaresi, M. and Benediktsson, J.A., 2001, A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, pp. 309–320.

Pinnamaneni, P., Saladi, S. and Meyer, J., 2001, 3-D Haar wavelet transformation and texture-based 3D reconstruction of biomedical data sets. *Visualization, Imaging and Image Processing (VIIP 2001)*. In *The International Association of Science and Technology for Development (IASTED)*, 3–5 September 2001, Marbella, Spain (Calgary, AB: ACTA Press), pp. 389–394.

Puissant, A., Hirsch, J. and Weber, C., 2005, The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. *International Journal of Remote Sensing*, **26**, pp. 733–745.

Saladi, S., Pinnamaneni, P. and Meyer, J., 2001, Texture-based 3-D brain imaging. In *2nd IEEE International Symposium on Bioinformatics and Bioengineering*, 4–6 November 2001, Bethesda, MD (Piscataway, NJ: Institute of Electrical and Electronics Engineers), pp. 136–143.

Shah, V.P., Younan, N.H., Durbha, S.S. and King, R.L., 2010, Feature identification via a combined ICA–wavelet method for image information mining. *IEEE Geosciences and Remote Sensing Letters*, **7**, pp. 18–22.

Yoo, H.Y., Lee, K. and Kwon, B.D., 2009, Quantitative indices based on 3D discrete wavelet transform for urban complexity estimation using remotely sensed imagery. *International Journal of Remote Sensing*, **30**, pp. 6219–6239.