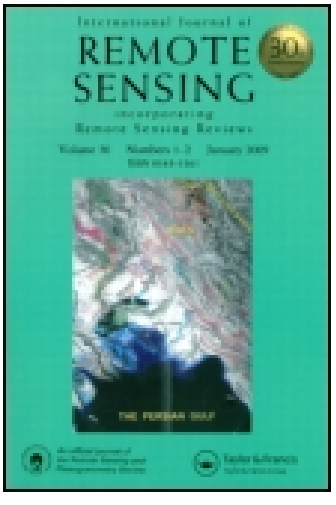


International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

# A multiscale feature fusion approach for classification of very high resolution satellite imagery based on wavelet transform

X. Huang [a], L. Zhang [a] & P. Li [a]

[a] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China

Published online: 20 Sep 2008.

To cite this article: X. Huang, L. Zhang & P. Li (2008) A multiscale feature fusion approach for classification of very high resolution satellite imagery based on wavelet transform, International Journal of Remote Sensing, 29:20, 5923-5941, DOI: 10.1080/01431160802139922

To link to this article: http://dx.doi.org/10.1080/01431160802139922

*International Journal of Remote Sensing*
Vol. 29, No. 20, 20 October 2008, 5923–5941

Taylor & Francis
Taylor & Francis Group

# A multiscale feature fusion approach for classification of very high resolution satellite imagery based on wavelet transform

X. HUANG, L. ZHANG* and P. LI
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China

(*Received 15 December 2006; in final form 15 April 2008*)



A novel methodology based on multiscale spectral and spatial information fusion using wavelet transform is proposed in order to classify very high resolution (VHR) satellite imagery. Conventional wavelet-based feature extraction methods employ single windows of a fixed size, which are not satisfactory as the VHR imagery contains complex and multiscale objects. In this paper, spectral and spatial features are extracted based on a set of concentric windows around a central pixel in order to integrate the information across different windows/scales. The proposed method is made up of three blocks: (1) the conventional wavelet-based feature extraction methods are extended from single band processing to multispectral bands, and from single window to multi-windows, (2) two multiscale fusion algorithms are proposed to exploit the multiscale spectral and spatial information and (3) a support vector machine (SVM), a relatively new method of machine learning, is used to classify the multiscale spectral–spatial feature sets. The proposed classification method is evaluated on two VHR datasets and the results show that the multiscale approach can improve the classification accuracy in homogeneous areas while simultaneously preserving accuracy in edge regions.

## 1. Introduction

Traditional image processing techniques have been proven to be inadequate for classification of very high resolution (VHR) remotely sensed imagery (Gong and Howarth 1990, Gong *et al.* 1992). Detailed features and small objects can be detected in VHR images and, consequently, the spectral signatures inside an information class become more heterogeneous and different objects become more spectrally similar (Myint *et al.* 2004). The resulting high intraclass and low interclass variabilities lead to a reduction in the statistical separability of the different land-cover classes in the spectral domain (Bruzzone and Carlin 2006). The introduction of spatial features has been found effective for addressing this problem. Therefore, many spatial feature extraction methods have been proposed and integrated with spectral features for classification of VHR satellite data. One commonly applied statistical procedure is the grey level co-occurrence matrix (GLCM), which is a widely used texture analysis technique for satellite data and has been successful to a certain extent (Barber and LeDrew 1991, Gong *et al.* 1992, Zhang 1999). A method based on straight lines to assess land development in high resolution satellite images was introduced and a set of statistical measures were

*Corresponding author. Email: zlp62@public.wh.hb.cn

*International Journal of Remote Sensing*
ISSN 0143-1161 print/ISSN 1366-5901 online 
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160802139922

extracted based on the sub-windows showing the regional line distribution (Unsalan and Boyer 2004a). An edge detection method, integrating a region-growing approach was used to improve classification for the images of the Indian remote sensing satellite 1C (IRS-1C) (Sun *et al.* 2003). Benediktsson *et al.* (2003, 2005) presented a technique of extended morphological profiles to describe multiscale spatial features and to interpret VHR satellite imagery. A length–width extraction algorithm (LWEA) was developed to extract the length and width of spectrally similar connected groups of pixels (Shackelford and Davis 2003). Huang *et al.* (2007) proposed a pixel shape index (PSI), which was a modified version of the LWEA. The PSI was calculated by a predetermined number of equally spaced lines (direction-lines) radiating from the central pixel.

This paper studies wavelet-based spectral and spatial features. Wavelet theory is well suited in any area of study where signals are complex and non-periodic, and it is particularly good at describing a scene in terms of the scale of the textures in it (Acharyya *et al.* 2003). Myint *et al.* (2004) compared the wavelet transform approach with spatial autocorrelation, the fractal approach and the GLCM. It was found that the wavelet approach is the most accurate of all four approaches for urban feature discrimination in VHR images. Acharyya *et al.* (2003) applied the M-band wavelet and neurofuzzy hybridization to segment the images of the IRS-1A. They found that the use of wavelet theory provided an effective representation of these images in terms of frequencies in different directions and orientations at different resolutions. Fukuda and Hirosawa (1999) proposed a wavelet-based texture feature set via the energy of sub-bands by the wavelet decomposition of a local window in an image. In their experiments, the wavelet feature set was successfully applied to the classification of SAR images. They carried out the classification by the simple minimum distance classifier for the purpose of examining the efficiency of the wavelet-based texture feature, and they thought that if more powerful classifiers such as neural networks were used, classification accuracy might be further improved.

Based on some approaches previously presented in the literature, this paper proposes a novel wavelet-based multiscale spectral–spatial feature fusion technique for classification of VHR satellite imagery. The salient aspects of the strategy are the following:

(1) A set of concentric windows with different sizes around a pixel is defined for identifying objects and structures on different scales. In addition, the wavelet-based multiscale spectral–spatial feature set is based on multispectral bands instead of a single channel.
(2) Two multiscale information fusion algorithms are proposed. The first is an adaptive window method, which adaptively determines the optimal window for each pixel, and the other is a multiple window pyramid approach, which considers the information at all the windows.
(3) A support vector machine (SVM) is employed to classify the multiscale spectral–spatial feature sets. The notable advantages of the SVM include self-adaptability, swift learning pace and its high-dimensional property in feature space.

The rest of this paper is organized as follows. Section 2 provides a detailed description of the extended wavelet-based feature extraction method. Sections 3 and 4 describe the two multiscale fusion algorithms and the SVM-based classification

technique, respectively. Two experiments on QuickBird and IKONOS multispectral images to evaluate the performance of the proposed methodology are described in §5, and §6 concludes the paper.

## 2. Multiscale spectral–spatial feature extraction based on wavelet transforms

The traditional feature extraction methods based on wavelet transform can be described as after a square local area around each pixel is decomposed, different statistical measures of each sub-image are calculated and are assigned to the components of the feature vector of the central pixel in the area (Fukuda and Hirosawa 1999). The conventional methods are extended in this paper, and the flow charts are shown and compared in figure 1.

The processing techniques in figure 1 can be analysed in terms of the following aspects:

(a) Band selection. Zhang *et al.* (2006) extracted the wavelet texture features using the first principal component (PCA1) based on principal component analysis (PCA) (Jollie 1986) on a set of multispectral Quickbird bands. Ouma *et al.* (2006) carried out an investigation to determine which band is optimal for urban tree texture extraction in IKONOS images. Myint *et al.* (2004) compared the classification performance of wavelet-based features for different spectral bands and found that the highest accuracies were achieved by bands 6 and 12 and PCA1 of the advanced thermal land application sensor (ATLAS) images. Traditionally, the optimal band or PCA1 is used as the base image for wavelet transform. However, it is important to consider all the spectral information because VHR sensors, such as IKONOS and Quickbird, include only a few spectral bands, and loss of spectral information may lead to a decrease in classification accuracy. Spatial information is similar among different bands and much redundancy exists if all the multispectral bands are used as base images for spatial features. Therefore, in this paper, the multiscale spectral features are built on all the available multispectral bands, while the spatial information is produced using only the PCA1 image.

(b) Window size. Myint *et al.* (2004) tested and compared different window sizes, and the largest window ($65 \times 65$) was found most effective for homogeneous texture regions. In some literature, an $8 \times 8$ window size was employed for VHR images since it was a trade-off between homogenous accuracy increase and detail preservation (Fukuda and Hirosawa 1999, Zhang *et al.* 2006). Acharyya *et al.* (2003) used a small window in order to detect the fine structures. A geostatistical analysis indicated that there was no single window size that would adequately characterize the range of textural conditions present in VHR images (Coburn and Roberts 2004). Therefore, it is not sound to use a fixed window size for feature extraction, especially for VHR images. In this paper, two algorithms are proposed to integrate spectral and spatial features at different scales and which are described in detail in §3.

The proposed multiscale wavelet-based spectral–spatial feature extraction method can be described as follows:

Step1. For each pixel $(i, j)$, define a set of concentric windows around it: $\{w_1 \times w_1, \ldots, w_s \times w_s, \ldots, w_N \times w_N\}$ $(1 \leqslant s \leqslant N)$.

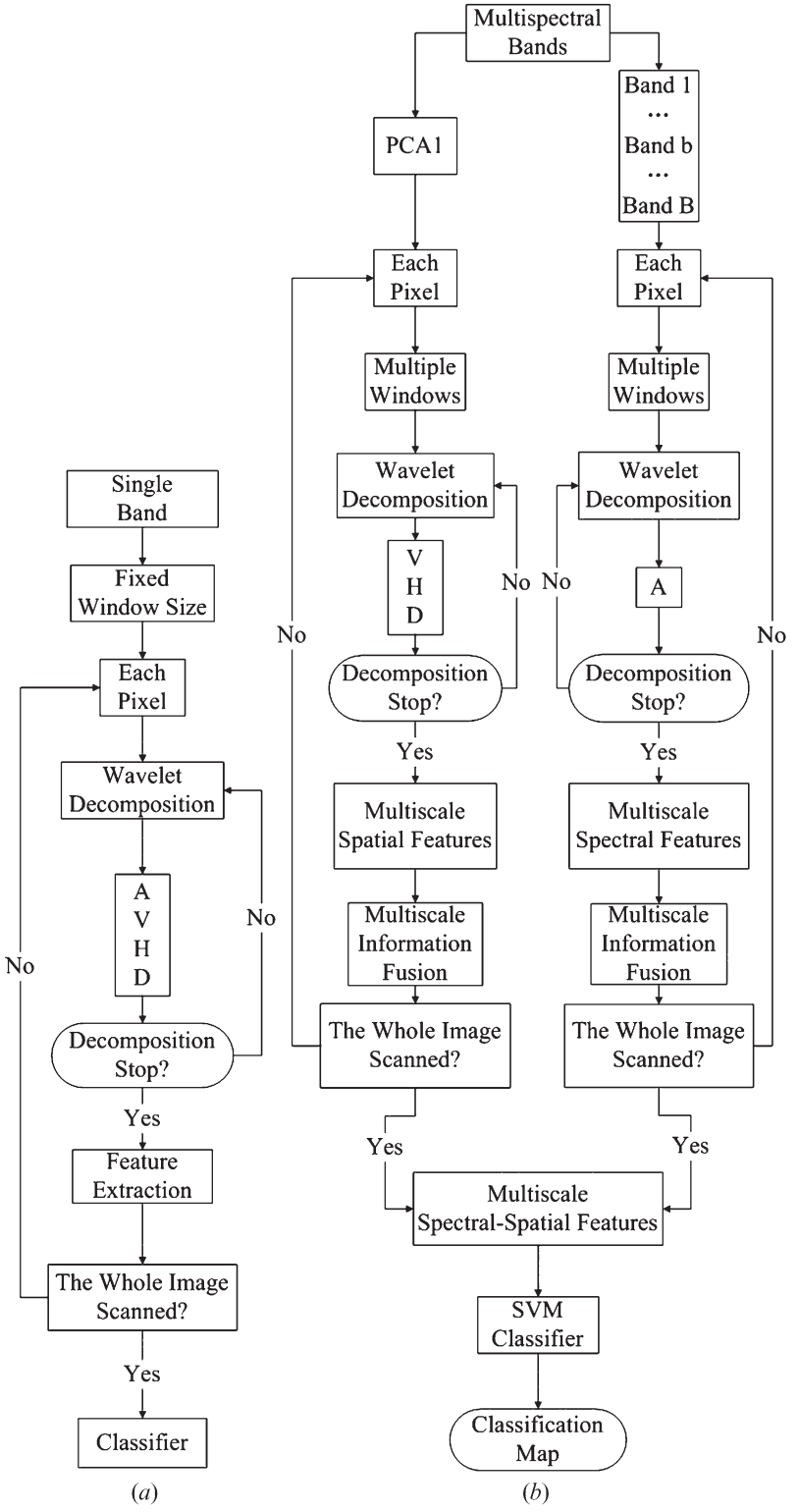


Figure 1. Flowchart of: (*a*) the conventional wavelet-based feature extraction algorithm and (*b*) the improved version in this paper.

Step2. Define the original multispectral images as scale $s=0$ and the spectral–spatial features are extracted using $w_1 \times w_1$, $w_s \times w_s$ and $w_N \times w_N$ moving windows.

Step3. (3a) The initializing value of $w$ is set to $w_s$;

(3b) Four sub-images can be obtained via wavelet transform (downsampling) of the $w \times w$ window region and the size of each sub-image is $(w/2) \times (w/2)$;

(3c) If $w/2=1$, then the wavelet decomposition stops and go to Step4; if $w/2>1$, then let $w=w/2$ and go to 3b.

This step aims to decompose the $w \times w$ area iteratively until four wavelet coefficients (approximate, horizontal, vertical and diagonal) are obtained. In this study, the Daubechies wavelet family (Daubechies 1990) is selected and the filter length is set to 4. It should be noted that the proposed algorithm is not sensitive to the wavelet transform approach and the filter kernels.

Step4. Let $s=s+1$ and repeat Step2 and Step3 until the wavelet coefficients for all the spectral bands at different scales have been calculated (i.e. $s=N$).

Step5. The multiscale spectral feature set for the central pixel $(i, j)$ can be defined as:

$$\begin{aligned} S_{\text{pe}}(i,j) &= \left\{ S^0_{\text{pe}}(i,j),\ \ldots,\ S^s_{\text{pe}}(i,j),\ \ldots,\ S^N_{\text{pe}}(i,j) \right\}, \\ \text{with } S^s_{\text{pe}}(i,j) &= \left\{ \left\| A^s_b(i,j) \right\| \right\}_{b=1}^{B} (0 \le s \le N, \quad 1 \le b \le B), \end{aligned} \tag{1}$$

where $A^s_b(i, j)$ denotes the approximation wavelet coefficient of pixel $(i, j)$ calculated by Steps 2 to 4. The parameters $b$ and $s$ represent the spectral channel and scale factor, respectively. It should be noted that $A^0_b(i, j)$ denotes the original spectral feature value in band $b$.

Step6. The multiscale spatial features are computed based on the PCA1 band. PCA is performed on the set of multispectral bands because it is optimal for data representation in the mean square sense and the first PCA band represents most of the information variation in the image. The multiscale spatial feature set for pixel $(i, j)$ can be defined as:

$$\begin{aligned} S_{\text{pa}}(i,j) &= \left\{ S^1_{\text{pa}}(i,j),\ \ldots,\ S^s_{\text{pa}}(i,j),\ \ldots,\ S^N_{\text{pa}}(i,j) \right\}, \\ \text{with } S^s_{\text{pa}}(i,j) &= \| V^s(i,j) \| + \| H^s(i,j) \| + \| D^s(i,j) \| \quad (1 \le s \le N), \end{aligned} \tag{2}$$

where $V^s(i, j)$, $H^s(i, j)$ and $D^s(i, j)$ are high-frequency wavelet coefficients for pixel $(i, j)$ at scale $s$ for the vertical, horizontal and diagonal directions, respectively.

Step7. Go through the whole image and calculate the multiscale spectral and spatial feature sets for each pixel. It is well known that combining spatial and spectral information can improve land use classification for VHR satellite imagery (Dell'Acqua *et al.* 2004). Hence, equations (1) and (2) are combined and the multiscale spectral–spatial feature set for the whole image can be written as:

$$\begin{aligned} F &= \left\{ F^0,\ \cdots,\ F^m,\ \cdots,\ F^N \right\}, \\ \text{with } F^0 &= S^0_{\text{pe}} \text{ and } F^s = \left\{ S^s_{\text{pe}},\ S^s_{\text{pa}} \right\} \quad (1 \le s \le N). \end{aligned} \tag{3}$$

## 3. Multiscale feature fusion

Coburn and Roberts (2004) showed that there was no single window size that would adequately characterize the range of textural conditions present in VHR data.

Hence, in this paper, the multiscale feature fusion algorithm is proposed to utilize information across different scales adaptively. It aims to increase the classification accuracy in homogeneous areas and at the same time preserve the accuracy in edge regions. In order to address this problem, two methods are proposed to fuse multiscale information. The first is an adaptive window (AW) algorithm and the second is a multiple window (MW) pyramid algorithm.

### 3.1 *Adaptive window algorithm*

The AW algorithm employs an adaptive scale fixing approach for different textural regions. It is preferable to characterize a texture using a large window so that a sufficient amount of information is considered. Yet, a small window is necessary to accurately locate the boundaries between homogeneous regions. In order to determine the suitable scale for a pixel adaptively, the edge feature is used as supplementary information and is defined as:

$$\begin{cases} E_t \text{ if pixel } (i,j) \text{ is at the edge,} \\ E_b(i,j) = 0 \quad \text{otherwise,} \end{cases} \tag{4}$$

where $b$ is the number of spectral band ($1 \leqslant b \leqslant B$). The Sobel edge detector, the Prewitt edge detector, the TBL edge detector (Sun *et al.* 2003) and the Canny filter (Solaiman *et al.* 1998) have been tested for satellite imagery. In this study, edge detection is performed using the Canny filter, since it is an optimal edge detector, having a low probability of false or missing edges and a high accuracy of edge positioning. After edge detection, a concept of edge density is defined for each pixel:

$$E^s_{\text{density}}(i,j) = \frac{\sum\limits_{(i,j)\in R_s} E(i,j)}{w_s \times w_s}, \tag{5}$$

where $R_s$ is a $w_s \times w_s$ window around the pixel $(i, j)$ and $E(i,j) = \frac{1}{B}\sum\limits_{b=1}^{B} E_b(i,j)$.

Based on the multispectral and edge information in different local windows, a scale selection index ($S_{\text{index}}$) is defined for choosing the optimal scale for each pixel:

$$S^s_{\text{index}}(i,j) = \left[\sum_{b=1}^{B} \frac{\sigma^s_b(i,j)}{\sigma^s_b}\right] \times E^s_{\text{density}}(i,j), \tag{6}$$

where $\sigma^s_b(i,j)$ and $\sigma^s_b$ are the local standard deviation computed for pixel $(i, j)$ and the global standard deviation for the whole image in band $b$ at scale $s$, respectively. The $S_{\text{index}}$ is a measure of local heterogeneity: low values correspond to homogeneous areas, while high values represent heterogeneous areas. A larger value of $S^s_{\text{index}}(i,j)$ shows that the information at scale $s$ contains more detailed features and hence, if a large window is used for the central pixel $(i, j)$, some important edge and structural information may be smoothed. Therefore, $S_{\text{index}}$ is used to adaptively select the optimal scale $O_{\text{scale}}(i,j)$ and the corresponding optimal window $O_{\text{window}}(i,j)$ for each pixel. The optimal scale $O_{\text{scale}}(i,j)$ for a pixel $(i, j)$ can be determined by:

$$O_{\text{scale}}(i,j) = \underset{1\leq s\leq N}{\arg\min}\left[S^s_{\text{index}}(i,j)\right], \tag{7}$$

where the operator 'arg min' means to find the optimal parameter $s$ by minimizing the function. A low value of $O_{\text{scale}}(i,j)$ shows that edge or geometrical detail is around the

central pixel and hence the information in a small window is more reliable for classification. A large $O_{\text{scale}}(i, j)$ denotes that the central pixel lies in a homogeneous area and a large window is used to improve the classification. Given that the optimal scale $O_{\text{scale}}(i, j)$ for pixel $(i, j)$ corresponds to the optimal window $O_{\text{window}}(i, j)$, some new feature sets are computed by averaging all the sequential combinations on the windows smaller than or equal to the optimal window $O_{\text{window}}(i, j)$:

$$\overline{S_{\text{pe}}^{O_{\text{scale}}}}(i, j) = \frac{1}{O_{\text{scale}}(i, j) + 1} \sum_{x=0}^{O_{\text{scale}}(i, j)} S_{\text{pe}}^{x}(i, j) \tag{8}$$

and

$$\overline{S_{\text{pa}}^{O_{\text{scale}}}}(i, j) = \frac{1}{O_{\text{scale}}(i, j) + 1} \sum_{x=1}^{O_{\text{scale}}(i, j)} S_{\text{pa}}^{x}(i, j), \tag{9}$$

where the optimal scale represents the largest window size included in the average operation and $x$ denotes the scales smaller than or equal to the optimal one ($x \leqslant O_{\text{scale}}(i, j)$). In equations (8) and (9), the features at the windows $w_s \leqslant O_{\text{window}}(i, j)$ are averaged because this can enhance the homogeneity and at the same time preserve the detailed and edge information. The parameters $\overline{S_{\text{pe}}^{O_{\text{scale}}}}$ and $\overline{S_{\text{pa}}^{O_{\text{scale}}}}$ denote the multiscale fusion feature sets of spectral and spatial information for the whole image, respectively, and $\overline{F^{O_{\text{scale}}}} = \left\{\overline{S_{\text{pe}}^{O_{\text{scale}}}}, \overline{S_{\text{pa}}^{O_{\text{scale}}}}\right\}$ denotes the multiscale spectral–spatial feature set. The classification rule for the adaptively fused multiscale feature can be described as:

$$C(i, j) = c_k \Leftrightarrow \text{C la}\left(\overline{F^{O_{\text{scale}}}}(i, j)\right) = c_k, \tag{10}$$

where $\text{C la}\left(\overline{F^{O_{\text{scale}}}}(i, j)\right)$ denotes the class label for the pixel $(i, j)$ classified with feature set $\overline{F^{O_{\text{scale}}}}$, $C(i, j)$ is the class label of pixel $(i, j)$ and $\{c_k\}_{k=1}^{p}$ represent the class labels in classification maps ($C(i, j) \in \{c_1, \ldots, c_k, \ldots c_p\}$ with $1 \leqslant k \leqslant p$).

### 3.2 *Multiple windows pyramid algorithm*

The AW algorithm integrates multiscale features by selecting an optimal scale and averaging the information at the windows smaller than or equal to the optimal one. A $(B+1)$-dimensional fusion feature set $\overline{F^{O_{\text{scale}}}} = \left\{\overline{S_{\text{pe}}^{O_{\text{scale}}}}, \overline{S_{\text{pa}}^{O_{\text{scale}}}}\right\}$ is used for classification in the AW algorithm, while the MW algorithm aims to exploit the feature set $F$ (see equation (3)), which contains all the multiscale spectral and spatial information. The MW algorithm fuses multiscale features automatically by integrating a set of concentric windows in a classifier. Considering that human interpreters may simultaneously use multiple windows of a variety of sizes to build evidence of the image pattern (Binaghi *et al.* 2003), the MW algorithm is an attempt to mimic human perception in identifying objects of different shape and structure on different scales. The classification rule for the MW algorithm can be described as:

$$C(i, j) = c_k \Leftrightarrow \text{Cla}(F(i, j)) = c_k, \tag{11}$$

where $F(i, j)$ denotes the spectral/spatial features in the multiple windows for pixel $(i, j)$. It is worth noting that, as the MW algorithm provides a large amount of

information, it always leads to a high-dimensional feature vector ($(N+1)\times B+N$). For this reason, a SVM is used to classify the feature sets extracted by the multiple windows as it is a classification technique intrinsically less sensitive to the high dimensionality of the feature space (Bruzzone and Carlin 2006).

## 4. Support vector machine

Conventional classifiers, such as the maximum likelihood classifier, are not capable of achieving a satisfactory accuracy for VHR data. This is because the estimated distribution function usually employs the normal distribution, which may not represent the actual distribution of the data. In recognition of this, the fuzzy ARTMAP (Dell'Acqua *et al.* 2004), BP neural networks (Benediktsson *et al.* 2005), probabilistic relaxation (Unsalan and Boyer 2004b) and probability neural networks (Tian *et al.* 1999) have been tested for integrating the spectral and spatial features. In this paper, a SVM is employed to interpret the multiscale spectral–spatial features because of its computational simplicity, superior accuracy and high-dimension property in feature space when compared to other classifiers (Cortes and Vapnik 1995).

SVM classifiers of the form $f(x)=w\Phi(x)+b$ learn from the data $\{(x_i, y_i)|x_i \in R^d, y_i \in \{-1, 1\}, i=1, \ldots, D\}$, where $x_i$ is a $d$-dimensional feature vector in a $d$-dimensional space of real numbers, $f(x)$ denotes a hyperplane that separates samples label $y_i$ on each side and the weight vector $w$ and bias term $b$ are the parameters of the hyperplane. The hyperplane calculation can be formulated into a constrained optimization problem as:

$$\min_{w,b,\xi_i}\left\{\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\xi_i\right\} \text{ subject to } y_i(w\Phi(x_i)+b)\geq 1-\xi_i, \quad \xi_i\geq 0, \tag{12}$$

where $C$ is a regularization parameter and $\xi_i$ is a slack variable. Based on the Lagrange formulation, the optimal discriminant function can be expressed in terms of Lagrange dual variables $\alpha_i$:

$$f(x)=\sum_{i\in \mathrm{SV}}\alpha_i y_i K(x_i, x_j)+b, \tag{13}$$

where SV (support vector) is the set of training samples with associated dual variables $\alpha_i$ satisfying $\alpha_i>0$. A kernel function $K(x_i, x_j)$ is introduced into the SVM so that the original input space can be transformed nonlinearly to a higher dimensional feature space where linear methods may be applied (Liu *et al.* 2006). A specific application for a SVM needs to handle several issues:

(1) Multiclass problem. While the SVM was originally designed for binary classification, most remote sensing applications involve multiple classes. Two approaches are commonly used for an $M$-class SVM (Foody and Mather 2004, Melgani and Bruzzone 2004): (a) one against all (OAA) involves a parallel architecture made up of $M$ SVMs, one for each class. Each SVM solves a two-class problem defined by one information class against all the others and the winner-takes-all rule is used to decide the class label for each pixel; (b) one against one (OAO), $M(M-1)/2$ SVMs are applied on each pair of classes. In this case, each SVM carries out a binary classification in which two information classes are analysed against each other and the most commonly computed class label is reserved for each pixel. In this paper, the OAO method is employed.

(2) Selection of kernel function. Based on the kernel function in equation (13), the SVM finds an optimal linear hyperplane in a higher dimensional feature space that is nonlinear in the original input space. The kernel trick avoids direct evaluation in the higher dimensional feature space by computing it via the kernel function with data vectors in the input space. The commonly used kernel functions are the radial basis function (RBF) and the inhomogeneous polynomial function (POLY):

$$\begin{cases} \text{POLY} : K(x_i, x_j) = (x_i \cdot x_j + 1)^q \\ \text{RBF} : K(x_i, x_j) = \exp\left[-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right] \end{cases}, \tag{14}$$

where $q$ is the order of the POLY kernel function and $\sigma$ is the kernel parameter. In this paper, the SVM classifier with a RBF kernel is used, because it has been proven effective in a number of different classification problems.

(3) Parameter optimization. In addition, the kernel based implementation of the SVM involves problems pertaining to the selection of multiple parameters, including the kernel parameters $\sigma$ and the regularization parameter $C$. Some standard methods exist that can facilitate the selection of parameters in the SVM classifier design. In our method, these parameters were selected using the leave one out (LOO) algorithm, which minimizes the estimate of the expected generalization error using a gradient descent search over the space of the parameters (Melgani and Bruzzone 2004).



## 5. Experiments

Two VHR datasets were tested for the proposed multiscale classification technique. In experiments, the statistics employed to measure classification accuracies were the overall accuracy (OA) and the Kappa coefficient based on the confusion matrix. In this paper, a more precise accuracy assessment procedure is adopted to evaluate the performance of different feature sets at different scales. The test samples were divided into two subsets: homogeneous areas and edge regions. This approach was used because feature sets on larger window sizes (or coarser resolution) usually perform well in homogeneous areas, but produce low accuracies in edge regions (Myint *et al.* 2004). Therefore, the effectiveness of the multiscale classification algorithm should be evaluated based on both homogeneous and edge accuracies.

### 5.1 *IKONOS dataset*

The imagery used for this experiment was acquired by the IKONOS commercial remote sensing satellite over Shenzhen City in the south of China and consists of four multispectral (MS) bands with 4 m resolution. The four MS bands collect data at the red, green, blue and near-infrared channels. The test image comprised 512 lines and 512 columns and is shown in figure 2(*a*). Table 1 shows the relationship between scale, resolution and window size and feature set.

The multiscale spectral and spatial feature images are shown in figure 2. According to table 1 and figure 2, an $8 \times 8$ window size represents a local area of $32 \times 32$ m, which is larger than the size of many objects in the image, such as *road*, *building*, *grass*, *tree*, etc. Therefore, in this experiment, only the information at pixel

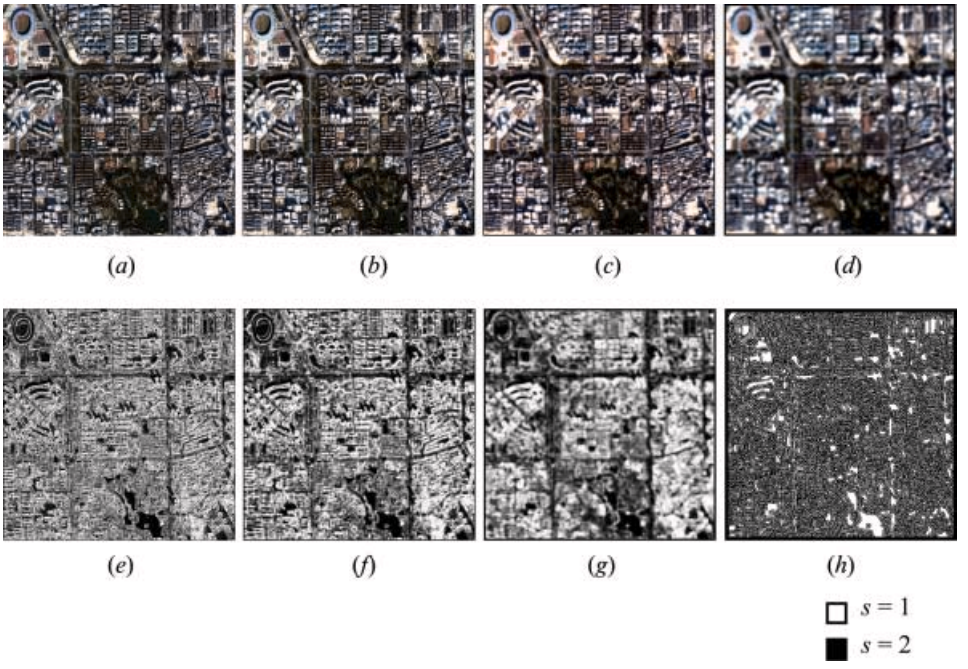


Figure 2. (*a*) IKONOS multispectral image in Shenzhen City, south of China. Multiscale spectral feature images for scale: (*b*) $s=1$, (*c*) $s=2$ and (*d*) $s=3$. Multiscale spatial feature images for scale: (*e*) $s=1$, (*f*) $s=2$ and (*g*) $s=3$. (*h*) The optimal window size for each pixel.

level $2\times2$ and $4\times4$ window sizes is employed for feature fusion. The optimal scale $O_{\text{scale}}(i, j)$ (see equation (7)) for each pixel is shown in figure 2(*h*), where larger windows are assigned to homogeneous areas and smaller ones are used for edge pixels.

Table 2 shows the numbers of samples in the training and test sets for different information classes. The accuracy statistics for different feature sets are provided in table 3.

From the classification accuracies in table 3, it can be observed that:

(a) When spectral information is used alone for classification, the features extracted by the $2\times2$ and $4\times4$ windows enhance the accuracies both in homogeneous and edge regions, which shows that the contextual information can increase the accuracy of classification of VHR data effectively. It is worth noting that the spectral classification at scale $s=1$ outperforms that at scale $s=2$ because a large local window may change or destroy some fine structures and geometrical details in the VHR image. Although spectral features are used alone, the proposed AW and MW algorithms significantly increase the

Table 1. The relationship between scale, resolution, window size and feature set.

| Scale ($s$) | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| Window size | $1\times1$ | $2\times2$ | $4\times4$ | $8\times8$ |
| Resolution (m) | 4 | 8 | 16 | 32 |
| Feature set | $F^0=\left\{S^0_{\text{pe}}\right\}$ | $F^1=\left\{S^1_{\text{pe}}, S^1_{\text{pa}}\right\}$ | $F^2=\left\{S^2_{\text{pe}}, S^2_{\text{pa}}\right\}$ | $F^3=\left\{S^3_{\text{pe}}, S^3_{\text{pa}}\right\}$ |

Table 2. Number of training and test samples for different information classes in the IKONOS dataset.

| Class | Number of training samples | Test set on homogeneous area | Test set on edge area |
|---|---|---|---|
| Water | 72 | 710 | 198 |
| Tree | 74 | 1141 | 359 |
| Road | 97 | 269 | 454 |
| Building | 85 | 2005 | 634 |
| Bare soil | 75 | 162 | 74 |
| Shadow | 74 | 830 | 708 |
| Grass | 72 | 238 | 188 |
| Total | 549 | 5355 | 2615 |



classification results at pixel level. Compared with the results of pixel level classification, the accuracy improvements for homogeneous areas are 9.8% and 12.1% in OA and 11.5% and 13.8% in Kappa for the AW and MW algorithms, respectively. The improvements for edge regions are 5.5% and 3.0% in OA and 6.4% and 3.4% in Kappa, respectively.

(b) Purely spectral information cannot discriminate spectrally similar objects effectively, especially for VHR data. Therefore, the introduction of wavelet high-frequency features evidently improves the spectral classification, both in homogeneous and edge regions. This tendency can be clearly seen from table 3. Compared to the results of pixel level classification, the accuracy improvements for homogeneous areas are 14.1% and 14.2% in OA and 16.5% and 16.3% in Kappa for the AW and MW algorithms, respectively. The improvements for edge regions are 8.8% and 6.4% in OA and 9.8% and 7.4% in Kappa, respectively.

The classification maps for different feature sets for the same area as figure 2 are shown in figure 3. The confusion matrices for figures 3(*a*), (*d*) and (*e*) are provided in table 4.

Table 3. Accuracy statistics for different feature sets in the IKONOS experiment (OA represents overall accuracy and the highest accuracies for different feature sets are highlighted).

| | | | Homogeneous accuracy | | Edge accuracy | |
|---|---|---|---|---|---|---|
| Feature | Scale | Feature set | OA (%) | Kappa | OA (%) | Kappa |
| Spectral | 0 | $S_{pe}^{0}$ | 74.08 | 0.685 | 74.61 | 0.690 |
| | 1 | $S_{pe}^{1}$ | 79.31 | 0.744 | 76.90 | 0.715 |
| | 2 | $S_{pe}^{2}$ | 78.19 | 0.729 | 74.99 | 0.689 |
| | AW | $\overline{S_{pe}^{O_{scale}}}$ | 83.83 | 0.800 | 80.15 | 0.754 |
| | MW | $S_{pe}$ | 86.13 | 0.823 | 77.59 | 0.724 |
| Spectral and spatial | 1 | $F^{1}=\left\{S_{pe}^{1}, S_{pa}^{1}\right\}$ | 82.35 | 0.778 | 78.70 | 0.734 |
| | 2 | $F^{2}=\left\{S_{pe}^{2}, S_{pa}^{2}\right\}$ | 82.88 | 0.785 | 78.36 | 0.730 |
| | AW | $\overline{F^{O_{scale}}}=\left\{\overline{S_{pe}^{O_{scale}}}, \overline{S_{pa}^{O_{scale}}}\right\}$ | 88.22 | 0.850 | 82.91 | 0.788 |
| | MW | $F=\{S_{pe}, S_{pa}\}$ | 88.27 | 0.848 | 80.99 | 0.764 |

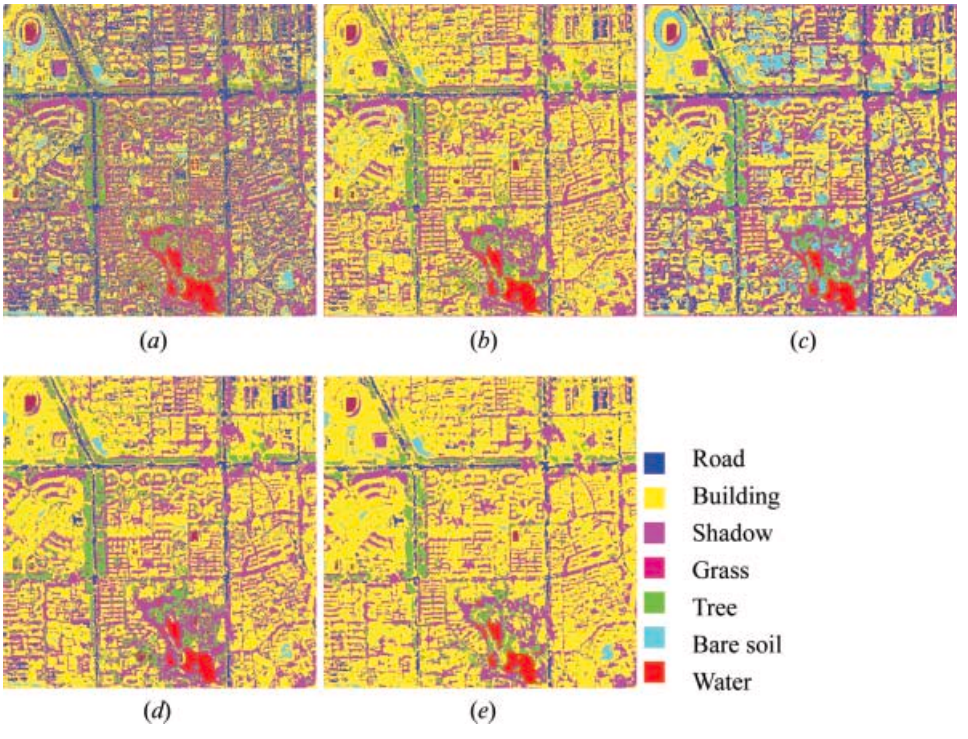


Figure 3. (*a*) Classification map for spectral information at pixel level. Classification results of spectral–spatial feature sets at scale: (*b*) $s=1$ and (*c*) $s=2$. Classification maps of multiscale fusion feature sets for (*d*) AW and (*e*) MW algorithms.

Comparing the AW and MW algorithms with pixel level classification (figures 3(*a*), (*d*) and (*e*)), the improvements in homogeneous areas for the producer's accuracy (PA) of *building* are 24.6% and 36.3% for the AW and MW algorithms, respectively; the improvements for *bare soil* are 14.8% and 14.2% for the PA and 19.5% and 30% for the user's accuracy (UA), respectively. For the accuracy statistics in edge regions, improvements occurred in *water*, *tree*, *bare soil*, *building*, *road* and *grass* for the PA and in *water*, *tree*, *road*, *building*, *bare soil*, *shadow* and *grass* for the UA.

### 5.2 *Quickbird dataset*

The image used in this experiment is of the urban area of the city of Beijing. As is well known, China is a developing country that has shown rapid economic and social development in recent years. In particular, because Beijing is to be the host city of the 2008 Olympic Games, additional investment has stemmed from governments and enterprises to accelerate the city construction, especially for the infrastructure. The VHR satellite images will play an important part in regional planning, project programming and construction monitoring, etc. The Quickbird image used in this experiment consists of three multispectral bands (RGB) with 2.44 m spatial resolution (the infrared channel is unfortunately not available). The image, of size $512 \times 512$ pixels, is shown in figure 4(*a*). Table 5 shows the number of samples in the training and test sets for different information classes.

Table 4. Accuracy statistics for the classification maps in figure 3.
W – water, T – tree, R – road, B – building, BS – bare soil, S – shadow, G – grass, UA – user's accuracy and PA – producer's accuracy.

(*a*) Homogeneous and edge accuracies for pixel level classification (figure 3(*a*)).

| | Homogeneous areas | | | | | | | | Edge regions | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | T | R | B | BS | S | G | UA | W | T | R | B | BS | S | G | UA |
| W | 681 | 0 | 0 | 0 | 0 | 31 | 0 | 95.7 | 171 | 1 | 0 | 0 | 0 | 15 | 0 | 91.4 |
| T | 0 | 972 | 1 | 34 | 0 | 45 | 8 | 91.7 | 0 | 297 | 14 | 45 | 0 | 50 | 10 | 71.4 |
| R | 0 | 0 | 245 | 581 | 0 | 2 | 0 | 29.6 | 0 | 4 | 402 | 208 | 0 | 3 | 0 | 65.2 |
| B | 0 | 37 | 22 | 979 | 24 | 22 | 8 | 89.7 | 0 | 10 | 32 | 264 | 18 | 31 | 26 | 69.3 |
| BS | 0 | 0 | 0 | 100 | 138 | 0 | 0 | 58.0 | 0 | 0 | 0 | 15 | 56 | 0 | 0 | 78.9 |
| S | 29 | 86 | 0 | 130 | 0 | 730 | 0 | 74.9 | 27 | 21 | 0 | 10 | 0 | 609 | 0 | 91.3 |
| G | 0 | 46 | 1 | 181 | 0 | 0 | 222 | 49.3 | 0 | 26 | 6 | 92 | 0 | 0 | 152 | 55.1 |
| PA | 95.9 | 85.2 | 91.1 | 48.8 | 85.2 | 87.9 | 93.3 | | 86.4 | 82.7 | 88.6 | 41.6 | 75.7 | 86.0 | 80.9 | |

OA=74.1%, Kappa=0.685 OA=74.6%, Kappa=0.690

(*b*) Homogeneous and edge accuracies for multiscale spectral classification (figure 3(*d*)).

| | Homogeneous areas | | | | | | | | Edge regions | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | T | R | B | BS | S | G | UA | W | T | R | B | BS | S | G | UA |
| W | 699 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 164 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| T | 0 | 1039 | 0 | 3 | 0 | 33 | 2 | 96.5 | 0 | 328 | 0 | 0 | 0 | 14 | 15 | 91.9 |
| R | 0 | 0 | 220 | 267 | 0 | 0 | 0 | 45.2 | 0 | 0 | 400 | 111 | 0 | 0 | 0 | 78.3 |
| B | 0 | 1 | 49 | 1657 | 3 | 83 | 0 | 92.4 | 0 | 4 | 37 | 499 | 13 | 134 | 9 | 71.7 |
| BS | 0 | 0 | 0 | 13 | 159 | 0 | 0 | 92.5 | 0 | 0 | 0 | 0 | 61 | 0 | 0 | 100 |
| S | 11 | 69 | 0 | 65 | 0 | 714 | 0 | 83.1 | 34 | 20 | 17 | 24 | 0 | 560 | 8 | 84.5 |
| G | 0 | 32 | 0 | 0 | 0 | 0 | 236 | 88.1 | 0 | 7 | 0 | 0 | 0 | 0 | 156 | 95.7 |
| PA | 98.5 | 91.1 | 81.8 | 82.6 | 98.2 | 86.0 | 99.2 | | 82.8 | 91.4 | 88.1 | 78.7 | 82.4 | 79.1 | 83.0 | |

OA=88.2%, Kappa=0.850 Overall accuracy=82.9%, Kappa=0.788

(*c*) Homogeneous and edge accuracies for multiscale spectral–spatial approach (figure 3(*e*)).

| | Homogeneous areas | | | | | | | | Edge regions | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | T | R | B | BS | S | G | UA | W | T | R | B | BS | S | G | UA |
| W | 708 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 180 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| T | 0 | 1004 | 0 | 0 | 0 | 14 | 0 | 98.6 | 1 | 309 | 0 | 34 | 0 | 31 | 4 | 81.5 |
| R | 0 | 0 | 185 | 132 | 0 | 0 | 0 | 58.4 | 0 | 4 | 388 | 81 | 0 | 0 | 0 | 82.0 |
| B | 0 | 61 | 84 | 1706 | 1 | 63 | 28 | 87.8 | 0 | 21 | 65 | 477 | 0 | 98 | 73 | 65.0 |
| BS | 0 | 0 | 0 | 22 | 161 | 0 | 0 | 88.0 | 0 | 0 | 0 | 0 | 74 | 0 | 0 | 100 |
| S | 2 | 76 | 0 | 145 | 0 | 753 | 0 | 77.2 | 17 | 14 | 0 | 33 | 0 | 579 | 0 | 90.1 |
| G | 0 | 0 | 0 | 0 | 0 | 0 | 210 | 100 | 0 | 11 | 1 | 9 | 0 | 0 | 111 | 84.1 |
| PA | 99.7 | 88.0 | 68.8 | 85.1 | 99.4 | 90.7 | 88.2 | | 90.9 | 86.1 | 85.5 | 75.2 | 100 | 81.8 | 59.0 | |

OA=88.3%, Kappa=0.848 OA=81.0%, Kappa=0.764

The multiscale spectral–spatial feature images are displayed in figure 4. In the experiment, $1 \times 1$ (the pixel level), $2 \times 2$, $4 \times 4$ and $8 \times 8$ window sizes are used for multiscale information fusion. The optimal window for each pixel is shown in figure 4(*h*). The classification accuracies for different feature sets are shown in table 6.

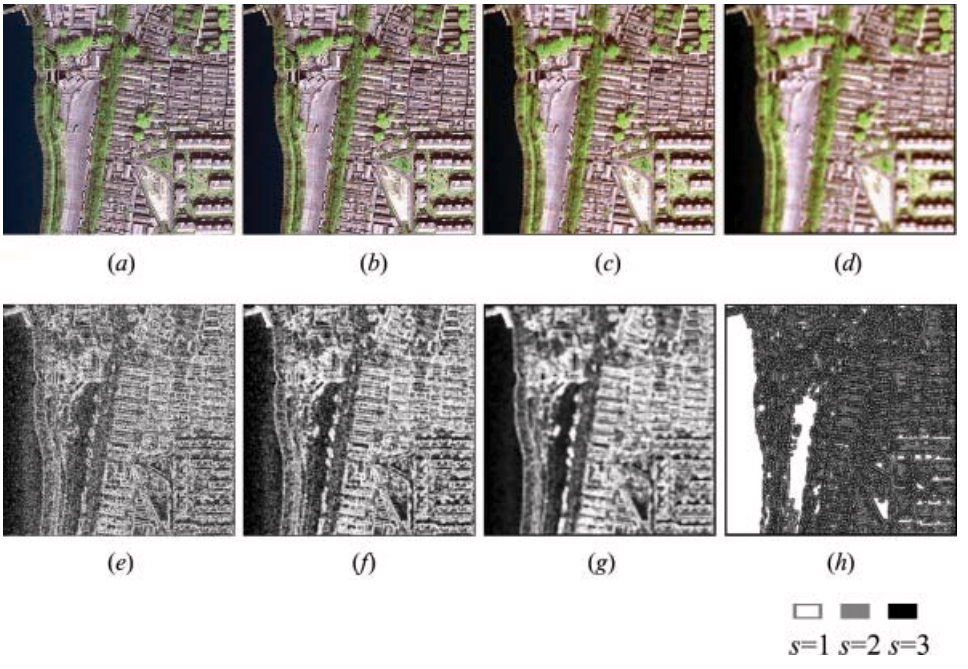


Figure 4. (*a*) Quickbird multispectral image in Beijing, China. Multiscale spectral feature images for scale: (*b*) $s=1$, (*c*) $s=2$ and (*d*) $s=3$. Multiscale spatial feature images for scale: (*e*) $s=1$, (*f*) $s=2$ and (*g*) $s=3$. (*h*) The optimal window size for each pixel.

From the results in table 6, it can be seen that:

(a) Comparing the classification results for scales $s=0$, $s=1$ and $s=2$, the feature sets at the larger windows achieve higher accuracies in homogeneous areas, but lower accuracies in edge regions. Therefore, we should evaluate the effectiveness of different window sizes considering both homogeneous and edge accuracies. This is especially important for VHR image processing. The information at low resolution is not reliable for edge areas as, at those scales, details and edge information have been smoothed by the decomposition process in a large window area. From the accuracy statistics in table 6, it can be seen that the two multiscale fusion approaches obtain better results than

Table 5. Number of the training and test samples for different information classes in the Quickbird dataset.

| Class | Number of training samples | Test set on homogeneous area | Test set on edge area |
|---|---|---|---|
| Water | 70 | 858 | 792 |
| Tree | 80 | 712 | 948 |
| Road | 85 | 1554 | 1159 |
| Building | 79 | 2679 | 1923 |
| Bare soil | 70 | 532 | 221 |
| Shadow | 76 | 1133 | 880 |
| Grass | 78 | 278 | 216 |
| Total | 538 | 7746 | 6139 |

Table 6. Accuracy statistics for different feature sets in the Quickbird experiment (highest accuracies for different feature sets are highlighted).

| Feature | Scale | Feature Set | Homogeneous accuracy | | Edge accuracy | |
|---|---|---|---|---|---|---|
| | | | OA (%) | Kappa | OA (%) | Kappa |
| Spectral | 0 | $S_{pe}^{0}$ | 83.72 | 0.800 | 74.33 | 0.689 |
| | 1 | $S_{pe}^{1}$ | 84.65 | 0.811 | 73.12 | 0.673 |
| | 2 | $S_{pe}^{2}$ | 84.77 | 0.811 | 71.18 | 0.647 |
| | AW | $\overline{S_{pe}^{O_{scale}}}$ | 86.81 | 0.836 | 73.48 | 0.674 |
| | MW | $S_{pe}$ | 89.53 | 0.869 | 78.84 | 0.736 |
| Spectral and spatial | 1 | $F^{1}=\left\{S_{pe}^{1}, S_{pa}^{1}\right\}$ | 85.00 | 0.815 | 73.81 | 0.681 |
| | 2 | $F^{2}=\left\{S_{pe}^{2}, S_{pa}^{2}\right\}$ | 86.42 | 0.831 | 73.76 | 0.677 |
| | AW | $\overline{F^{O_{scale}}}=\left\{\overline{S_{pe}^{O_{scale}}}, \overline{S_{pa}^{O_{scale}}}\right\}$ | 93.64 | 0.920 | 78.27 | 0.730 |
| | MW | $F=\{S_{pe}, S_{pa}\}$ | 90.16 | 0.876 | 80.29 | 0.754 |

the fixed window algorithm. Hence, it is necessary to integrate multiscale information in order to increase the homogeneous accuracy and at the same time keep the detail features.

(b) The accuracies achieved by integrating spectral and spatial features are higher than those obtained using purely spectral information. The introduction of wavelet-based structural features obviously improves the spectral classification, both in homogeneous and edge regions. This is because spectral information alone cannot discriminate spectrally similar objects effectively and wavelet-based high-frequency coefficients represent the edge and geometrical structures. The proposed wavelet-based spectral–spatial feature sets are effective for classifying VHR imagery.

The classification maps for different feature sets are shown in figure 5, and the confusion matrices for figures 5(*a*), (*d*) and (*e*) are provided in table 7.

From figure 5 and table 7, it can be observed that the multiscale fusion algorithms can improve the classification and reduce the pepper-salt effect in homogenous areas, and at the same time preserve the geometrical and structural information effectively. Compared to figure 5(*a*), the improvements in figures 5(*d*) and (*e*) are obvious, especially for the spectrally similar *road*, *building* and *bare soil* because of the introduction of multiscale wavelet-based spatial features.

Comparing the AW and MW algorithms with pixel level classification (figures 5(*a*), (*d*) and (*e*)), the improvements in homogeneous areas are 9.9% and 6.4% in OA and 12.0% and 7.6% in Kappa, respectively. The multiscale fusion algorithms also increase the classification accuracies in edge regions, the improvements in OA are 3.9% and 6.0% and 4.1% and 6.5% in Kappa for the AW and MW algorithms, respectively.

## 6. Conclusions

In this paper, a novel wavelet-based technique exploiting multiscale spectral and spatial information is proposed to classify very high resolution (VHR) satellite imagery. The limitation of the conventional wavelet-based feature extraction

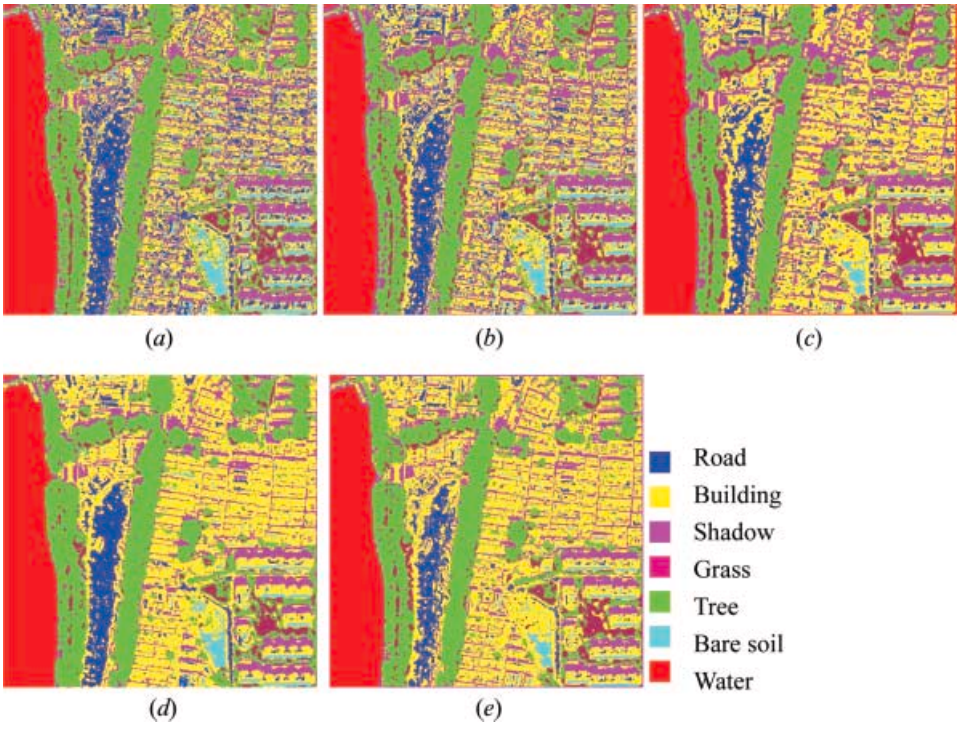


Figure 5. (*a*) Classification map for spectral information at pixel level. Classification results of spectral–spatial feature sets at scale: (*b*) $s=1$ and (*c*) $s=2$. Classification maps of multiscale fusion feature sets for (*d*) AW and (*e*) MW algorithms.

approach is the fixed shape and choice of size of the analysis window. The proposed methodology aims to extract multiscale features by integrating the information in a set of concentric windows around each pixel in the image. The proposed approach is made up of three blocks:

(1) The conventional wavelet-based feature extraction method is improved. Multispectral bands are considered because VHR satellite sensors only include a few spectral bands and hence spectral information is very important for classification. The multiscale spectral and spatial feature sets are extracted. The multiscale spectral feature sequence is obtained based on the approximate sub-bands via wavelet transform for different window areas, while the multiscale spatial feature sets are calculated by integrating the directional sub-bands (horizontal, vertical and diagonal).

(2) In order to exploit the multiscale spectral–spatial feature sets, two information fusion algorithms are proposed. The first is an adaptive window (AW) algorithm. It aims to select the optimal window for each pixel and the multiscale features are incorporated by averaging the feature sets at reliable scales for each pixel. The second is a multiple window (MW) pyramid algorithm, which aims to exploit the multiscale spectral and spatial feature sets automatically by the machine learning approach.

(3) A support vector machine (SVM) is then used to interpret the multiscale feature sets. The choice of an SVM-based classification architecture is

Table 7. Accuracy statistics for the classification maps in figure 5.
W – water, T – tree, R – road, B – building, BS – bare soil, S – shadow, G – grass, UA – user's accuracy and PA – producer's accuracy.

(*a*) Homogeneous and edge accuracies for pixel level classification (figure 5(*a*)).

| | Homogeneous areas | | | | | | | | Edge regions | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | T | R | B | BS | S | G | UA | W | T | R | B | BS | S | G | UA |
| W | 858 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 792 | 2 | 0 | 0 | 0 | 0 | 0 | 99.8 |
| T | 0 | 693 | 0 | 304 | 0 | 0 | 41 | 66.8 | 0 | 719 | 8 | 68 | 0 | 0 | 53 | 84.8 |
| R | 0 | 0 | 1403 | 556 | 0 | 0 | 0 | 71.6 | 0 | 0 | 810 | 293 | 0 | 2 | 0 | 73.3 |
| B | 0 | 0 | 149 | 1664 | 27 | 8 | 0 | 90.0 | 0 | 23 | 324 | 1035 | 36 | 17 | 1 | 72.1 |
| BS | 0 | 0 | 2 | 79 | 505 | 0 | 0 | 86.2 | 0 | 0 | 17 | 473 | 184 | 0 | 0 | 27.3 |
| S | 0 | 0 | 0 | 65 | 0 | 1125 | 0 | 94.5 | 0 | 57 | 0 | 36 | 0 | 861 | 0 | 90.3 |
| G | 0 | 19 | 0 | 11 | 0 | 0 | 237 | 88.8 | 0 | 147 | 0 | 18 | 1 | 0 | 162 | 49.4 |
| PA | 100 | 97.3 | 90.3 | 62.1 | 94.9 | 99.3 | 85.3 | | 100 | 75.8 | 69.9 | 53.8 | 83.3 | 97.8 | 75.0 | |

OA=83.7%, Kappa=0.800 OA=74.3%, Kappa=0.689

(*b*) Homogeneous and edge accuracies for multiscale spectral classification (figure 5(*d*)).

| | Homogeneous areas | | | | | | | | Edge regions | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | T | R | B | BS | S | G | UA | W | T | R | B | BS | S | G | UA |
| W | 858 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 792 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| T | 0 | 693 | 0 | 100 | 0 | 0 | 52 | 82.0 | 0 | 840 | 0 | 21 | 0 | 1 | 73 | 89.8 |
| R | 0 | 0 | 1370 | 55 | 0 | 0 | 0 | 96.2 | 0 | 0 | 704 | 102 | 0 | 0 | 0 | 87.3 |
| B | 0 | 0 | 184 | 2517 | 62 | 14 | 0 | 90.7 | 0 | 18 | 451 | 1448 | 39 | 175 | 4 | 67.8 |
| BS | 0 | 0 | 0 | 0 | 470 | 0 | 0 | 100 | 0 | 0 | 4 | 347 | 178 | 0 | 0 | 33.7 |
| S | 0 | 0 | 0 | 4 | 0 | 1119 | 0 | 99.6 | 0 | 31 | 0 | 0 | 0 | 704 | 0 | 95.8 |
| G | 0 | 19 | 0 | 3 | 0 | 0 | 226 | 91.1 | 0 | 59 | 0 | 5 | 4 | 0 | 139 | 67.2 |
| PA | 100 | 97.3 | 88.2 | 94.0 | 88.4 | 98.8 | 81.3 | | 100 | 88.6 | 60.7 | 75.3 | 80.5 | 80.0 | 64.4 | |

OA=93.6%, Kappa=0.920 OA=78.3%, Kappa=0.730

(*c*) Homogeneous and edge accuracies for multiscale spectral–spatial approach (figure 5(*e*)).

| | Homogeneous areas | | | | | | | | Edge regions | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | W | T | R | B | BS | S | G | UA | W | T | R | B | BS | S | G | UA |
| W | 858 | 0 | 0 | 0 | 0 | 0 | 0 | 100 | 792 | 0 | 0 | 0 | 0 | 0 | 0 | 100 |
| T | 0 | 629 | 0 | 181 | 0 | 0 | 24 | 75.4 | 0 | 740 | 0 | 41 | 0 | 0 | 30 | 91.3 |
| R | 0 | 0 | 1303 | 53 | 0 | 0 | 0 | 96.1 | 0 | 0 | 687 | 33 | 0 | 0 | 0 | 95.4 |
| B | 0 | 0 | 251 | 2385 | 64 | 46 | 0 | 86.9 | 0 | 22 | 471 | 1567 | 123 | 20 | 1 | 71.1 |
| BS | 0 | 0 | 0 | 0 | 468 | 0 | 0 | 100 | 0 | 0 | 0 | 270 | 98 | 0 | 0 | 26.6 |
| S | 0 | 0 | 0 | 39 | 0 | 1087 | 0 | 96.5 | 0 | 91 | 1 | 12 | 0 | 860 | 0 | 89.2 |
| G | 0 | 83 | 0 | 21 | 0 | 0 | 254 | 71.0 | 0 | 95 | 0 | 0 | 0 | 0 | 185 | 66.1 |
| PA | 100 | 88.3 | 83.9 | 89.0 | 88.0 | 95.9 | 91.4 | | 100 | 78.1 | 59.3 | 81.5 | 44.3 | 97.7 | 85.7 | |

OA=90.2%, Kappa=0.876 Overall accuracy=80.3%, Kappa=0.764

motivated by its self-adaptability, swift learning pace and high-dimensional property in feature space.

In experiments on the two VHR datasets, the proposed classification technique performed better than the conventional algorithms. The multiscale fusion approach can obviously increase the accuracies in homogeneous areas and at the same time

preserve the detail and edge features. In experiments, two comparisons were made. In the first, the classification results for different scales were compared. The experimental statistics show that larger analysis windows produced higher homogeneous accuracies but lower edge accuracies, and smaller windows resulted in higher accuracies in edge regions but lower accuracies and pepper-salt effects in homogeneous areas. The proposed AW and MW algorithms can overcome this trade-off effectively. In the second comparison, classification results for spectral feature sets were compared with spectral–spatial feature sets. In experiments, the classification by integrating spectral and spatial features clearly outperformed the purely spectral classification. This result shows that the extended wavelet-based feature extraction method was effective for VHR image classification.



**Acknowledgements**

The authors would like to thank the editor and the two anonymous reviewers. Their insightful suggestions have significantly improved this paper. This work is supported by the 863 High Technology Program of China under grant 2007AA12Z148 and by the National Science Foundation of China under grant 40771139.

**References**

Acharyya, M., De, R.K. and Kundu, M.K., 2003, Segmentation of remotely sensed images using wavelet features and their evaluation in soft computing framework. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 2900–2905.

Barber, D.G. and Ledrew, E.F., 1991, SAR sea ice discrimination using texture statistic: a multivariate approach. *Photogrammetric Engineering and Remote Sensing*, **57**, pp. 949–958.

Benediktsson, J.A., Pesaresi, M. and Arnason, K., 2003, Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 1940–1949.

Benediktsson, J.A., Palmason, J.A. and Sveinsson, J.R., 2005, Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, pp. 480–491.

Binaghi, E., Callo, I. and Pepe, M., 2003, A cognitive pyramid for contextual classification of remote sensing images. *IEEE transactions on Geoscience and Remote Sensing*, **41**, pp. 2906–2922.

Bruzzone, L. and Carlin, L., 2006, A multilevel context-based system for classification of very high spatial resolution images. *IEEE transactions on Geoscience and Remote Sensing*, **44**, pp. 2587–2600.

Coburn, C.A. and Roberts, A.C.B., 2004, A multiscale texture analysis procedure for improved forest stand classification. *International Journal of Remote Sensing*, **25**, pp. 4287–4308.

Cortes, C. and Vapnik, V., 1995, Support vector networks. *Machine Learning*, **20**, pp. 273–297.

Daubechies, I., 1990, The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, **36**, pp. 961–1005.

Dell'Acqua, F., Gamba, P., Ferari, A., Palmason, J.A., Benediktsson, J.A. and Arnason, K., 2004, Exploiting spectral and spatial information in hyperspectral urban data with high resolution. *IEEE Geoscience and Remote Sensing Letters*, **1**, pp. 322–326.

Foody, G.M. and Mather, A., 2004, A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, pp. 1335–1343.

FUKUDA, S. and HIROSAWA, H., 1999, A wavelet-based texture feature set applied to classification of multifrequency polarimetric SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, pp. 2282–2286.

GONG, P. and HOWARTH, P.J., 1990, The use of structural information for improving land-cover classification accuracies at the rural–urban fringe. *Photogrammetric Engineering and Remote Sensing*, **56**, pp. 67–73.

GONG, P., MARCEAU, D.J. and HOWARTH, P.J., 1992, A comparison of spatial feature extraction algorithms for land-use classification with SPOT HRV data. *Remote Sensing of Environment*, **40**, pp. 137–151.

HUANG, X., ZHANG, L. and LI, P., 2007, Classification and extraction of spatial features in urban areas using high resolution multispectral imagery. *IEEE Geoscience and Remote Sensing Letters*, **4**, pp. 260–264.

JOLLIE, I.T., 1986, *Principal Component Analysis* (New York: Springer-Verlag).

LIU, D.S., KELLY, M. and GONG, P., 2006, A spatial-temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. *Remote Sensing of Environment*, **101**, pp. 167–180.

MELGANI, F. and BRUZZONE, L., 2004, Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, pp. 1778–1790.

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.

OUMA, Y.O., NGIGI, T.G. and TATEISHI, R., 2006, On the optimization and selection of wavelet texture for feature extraction from high-resolution satellite imagery with application towards urban-tree delineation. *International Journal of Remote Sensing*, **27**, pp. 73–104.

SHACKELFORD, A.K. and DAVIS, C.H., 2003, A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 1920–1932.

SOLAIMAN, B., KOFFI, R.K., MOUCHOT, M.C. and HILLION, A., 1998, An information fusion method for multispectral image classification postprocessing. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, pp. 395–406.

SUN, W.X., HEIDT, V., GONG, P. and XU, G., 2003, Information fusion for rural land-use classification with high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, pp. 2840–2850.

TIAN, B., SHAIKH, M.A., AZIMI-SADJADI, M.R., VONDER, T.H. and REINKE, D.L., 1999, A study of cloud classification with neural networks using spectral and textural features. *IEEE Transactions on Neural Networks*, **10**, pp. 138–151.

UNSALAN, C. and BOYER, K.L., 2004a, Classifying land development in high-resolution panchromatic satellite images using straight-line statistics. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, pp. 907–919.

UNSALAN, C. and BOYER, K.L., 2004b, Classifying land development in high-resolution panchromatic satellite imagery using hybrid structural-multispectral features. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, pp. 2840–2850.

ZHANG, L., HUANG, X., HUANG, B. and LI, P., 2007, A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, pp. 2950–2961.

ZHANG, Y., 1999, Optimisation of building detection in satellite images by combining multispectral classification and texture filtering. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**, pp. 50–60.