

# Combined multi-kernel support vector machine and wavelet analysis for hyperspectral remote sensing image classification

Kun Tan (谭 琨) and Peijun Du (杜培军)\*

Key Laboratory for Land Environment and Disaster Monitoring of State Bureau of Surveying and Mapping of China,  
China University of Mining and Technology, Xuzhou 221116, China

\*Corresponding author: dupjrs@cumt.edu.cn

Received October 25, 2010; accepted November 30, 2010; posted online January 1, 2011

Many remote sensing image classifiers are limited in their ability to combine spectral features with spatial features. Multi-kernel classifiers, however, are capable of integrating spectral features with spatial or structural features using multiple kernels and summing them for final outputs. Using a support vector machine (SVM) as classifier, different multi-kernel classifiers are constructed and tested using 64-band Operational Modular Imaging Spectrometer II hyperspectral image of Changping Area, Beijing City. Results show that by integrating spectral and wavelet texture information, multi-kernel SVM classifiers can obtain more accurate classification results than sole-kernel SVM classifiers and cross-information SVM kernel classifiers. Moreover, when the multi-kernel SVM classifier is used, the combination of the first four principal components from principal component analysis and wavelet texture provides the highest accuracy (97.06%). Multi-kernel SVM is therefore an effective approach to improve the accuracy of hyperspectral image classification and to expand possibilities for remote sensing image interpretation and application.

OCIS codes: 100.4145, 100.5010, 100.7410.

doi: 10.3788/COL201109.011003.

Hyperspectral remote sensing image classification remains a contentious topic in remote sensing<sup>[1–3]</sup>. With the increasing amounts of data from airborne and satellite hyperspectral sensors, numerous bands and low training samples pose a “curse of dimensionality”<sup>[1–3]</sup>. To address this issue and to increase the stability of classifiers, some feature selection algorithms have been proposed. However, the proposed approaches, time-consuming and scenario-dependent, usually lead to information loss<sup>[4]</sup>. In recent years, support vector machines (SVMs) have been widely used to classify hyperspectral images, demonstrating excellent performance in terms of accuracy, generalization, and robustness<sup>[5]</sup>.

According to related studies, SVMs have better performance compared with traditional classifiers<sup>[1,6,7]</sup>. This is observed when spectral information is used for a SVM classifier. Moreover, SVMs can account for both spectral information and spatial features; for example, a good framework of multi-kernels for hyperspectral image classification has been presented<sup>[5,8,9]</sup>. However, the researchers did not pay more attention to the parameter selection technique and more sophisticated texture techniques — two very important methods for improving the performance of multi-kernel approaches.

Many recent studies attempted to enhance texture analysis through different methods, such as texture classification<sup>[10]</sup>, texture segmentation<sup>[11,12]</sup>, and texture detection<sup>[11,13]</sup>. Various texture analysis techniques have also been developed. Traditional statistical approaches to texture analysis, such as co-occurrence matrices, second order statistics, and Gauss-Markov random fields<sup>[14]</sup>, are restricted to the analysis of spatial interactions over relatively small neighborhoods on a single scale.

The wavelet theory has developed rapidly since the

1980s, and has been used in fields such as signal processing<sup>[11,15–17]</sup>, image restoration<sup>[18]</sup>, image retrieval<sup>[19]</sup>, and pattern recognition<sup>[20]</sup>, among others. Uses of wavelet theory on the texture analysis of remote sensing images have also been investigated<sup>[21]</sup>.

In this letter, SVMs, which have demonstrated superior performance in the context of hyperspectral image classification, are adopted as the classifier. A series of multi-kernel classifiers based on SVM are constructed to account for both spectral and spatial features (described by wavelet texture information).

The SVM theory for a two-class problem is found in some references<sup>[22,23]</sup>. Some popular kernel functions include linear kernel, polynomial kernel, Gaussian radial basis function (RBF) kernel, and sigmoid kernel.

Considering some common SVM kernels, the bottleneck is defined as the kernel mapping function that attains similar samples. In general, we can reconstruct a new kernel if it fulfils Mercer’s condition.

Let  $\chi$  be any input space and  $K : \chi \times \chi \rightarrow \mathbb{R}$  be a symmetric function. In the expression,  $K$  is a Mercer’s kernel if, and only if, the kernel matrix formed by restricting  $K$  to any finite subset of  $\chi$  is positive semi-definite, having no negative eigen values.

We can then consider the convex combination of multiple kernels:

$$K(x_i, x_j) = \sum_{k=1}^K \beta_k K_k(x_i, x_j), \quad (1)$$

where  $1 > \beta_k \geq 0$  and  $\sum_k \beta_k = 1$ , and each kernel satisfies the condition of Mercer’s kernel.

For definite pixel entity  $x_i$ , if  $x_i$  is recorded with  $x_i^s$  in the spectral domain and  $x_i^t$  is in the spatial domain after

texture extraction, we can rewrite the foregoing multi-kernel function as

$$K(x_i, x_j) = \sum_{k=1}^n \beta_k K_k(x_i^s, x_j^s) + \sum_{k=1}^m \beta_k K_k(x_i^t, x_j^t). \quad (2)$$

The multi-kernel equation constitutes a tradeoff between the spectral and spatial domains, which is actually implemented by designing the parameter of  $\beta_k$ . The preceding kernel classifiers can be conveniently modified to account for the relationship between spatial and spectral information<sup>[22]</sup>.

Similar to the convex kernel, another kernel can be defined as<sup>[5]</sup>

$$K(x_i, x_j) = K_s(x_i^s, x_j^s) + K_t(x_i^t, x_j^t) + K_{st}(x_i^s, x_j^t) + K_{ts}(x_i^t, x_j^s), \quad (3)$$

where  $x_i^s$  and  $x_j^s$  must have the same dimension. It is named cross-information kernel. We use a polynomial kernel for  $K_{st}(x_i^s, x_j^t)$  and  $K_{ts}(x_i^t, x_j^s)$  to decrease computational burden and simplify parameter selection. A detailed proof of this can be found in Ref. [5].

Wavelet analysis is based on the Fourier transform. Frequency analysis of stationary signals can be effectively achieved by projecting the signal onto a set of infinite spatial extent basis functions using the Fourier transform:

$$X(f) = \int_{-\infty}^{+\infty} x(t)e^{-2j\pi ft} dt, \quad (4)$$

where  $X(f)$  represents the global frequency of the signal. Similarly, effective frequency analysis of non-stationary signals can be achieved by projecting the signal onto a set of spatially localized basis functions using the wavelet transform. This is achieved by combining the inner products of signal  $X(t)$  with the translation and dilation of a prototype function:

$$\Psi(a, b) = \int_{-\infty}^{+\infty} x(t)\psi_{ab}^*(t)dt, \quad (5)$$

where  $a, b \in \mathbb{R}$ ,  $a$  is the dilatation factor and  $b$  is the translation factor. Here,  $\psi_{a,b}(t)$  is the translated and scaled version of the mother wavelet  $\psi(t)$  given by

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \quad (6)$$

Various choices of  $a$  and  $b$  result in different possible wavelet bases at different scales and translations. Notably, large values of the scale variable correspond to small frequencies and small lengths of the  $\psi_{a,b}(t)$  function. Conversely, small values correspond to low frequencies and large lengths of the  $\psi_{a,b}(t)$  function. This effectively shows that low frequencies need to be analyzed at large scales, and high frequencies at small scales. Haar, Daubechies, and Gabor are just a few of the wavelet functions.

From Mallat's theory, image data can be broken down into orthogonal components and translation invariant using the discrete wavelet transform (DWT) method<sup>[5]</sup>.

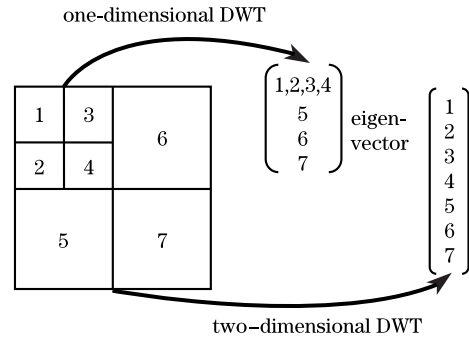


Fig. 1. DWT example for image data.

A DWT example for image data is shown in Fig. 1. To extend the wavelet transform to two dimensions, it is only necessary to filter separately horizontally and vertically. This produces four sub-bands at each scale. Denoting the horizontal frequency first and the vertical frequency second, this produces high-high (HH), high-low (HL), low-high (LH), and low-low (LL) image sub-bands.

A typical way of characterizing a texture using DWT is to extract energy measures from each sub-band<sup>[24]</sup>. If the sub-band is  $x(m, n)$ , then energy features from each sub-band can be calculated using the  $l_1$  norm:

$$e = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |x(i, j)|, \quad (7)$$

where  $M$  and  $N$  are the length and width of the sub-band, respectively;  $i$  and  $j$  are the rows and columns of the channel, respectively; and  $x$  is the wavelet coefficient. Each energy measure characterizes the magnitude of frequency content at the orientation and scale of each sub-band. Thereafter, each sub-band should be resampled to the original length and width of the image data.

By comparing the Haar wavelet and the Daubechies' wavelet, we see that the latter has continuous derivatives responding well to discontinuities in the texture, while the former does not allow sharp transitions and fast attenuation. Moreover, the Haar wavelet cannot efficiently separate the image signals into low- and high-frequency sub-bands. In contrast, Daubechies' wavelet constructs smooth scaling functions of compact support with orthonormal shifts. The DWT method can then be used to obtain smooth orthogonal wavelets. Many examples of textured image analysis, comparison, and segmentation have shown that Daubechies' wavelet is a successful technique<sup>[25,26]</sup>. Daubechies' wavelet transform (WT) based texture analysis produces the best result in invariant texture classification<sup>[27,28]</sup>.

For texture extraction, principal component analysis (PCA) was first conducted. The first four principal components (PCs) were chosen for wavelet texture extraction because they contain most of the information (about 90%) from the original data. Window size of the wavelet texture was also significant. We tested  $3 \times 3$ ,  $8 \times 8$ , and  $16 \times 16$  window sizes, finally adopting the  $8 \times 8$  window size. Figure 2 shows the original remote sensing image and the two-dimensional (2D) wavelet texture images, which have seven different texture images as the seven vectors. To obtain the wavelet texture, one-dimensional (1D) DWT and 2D DWT were applied. As the extension

of 1D DWT, 2D DWT can be carried out by the tensor product of two 1D wavelet base functions in the horizontal and vertical directions. This produces HH, HL, LH, and LL image sub-bands. Using the above wavelet theory, we obtained four different dimensions data for both 1D DWT and 2D DWT. These four dimensions were named wavelet texture 1 (WT1) and wavelet texture 2 (WT2).

The data used in the experiment were the Operational Modular Imaging Spectrometer II (OMIS II) hyperspectral image of Changping Area, Beijing City, China. The image has 512 rows, 512 columns, and 64 bands ranging from 0.46 to 1.1  $\mu\text{m}$ . It has a spatial resolution of 3 m. Preprocessing was conducted by the data provider. Figure 2(a) is the RGB composite image of the hyperspectral image (R: Band 36 with a wavelength of 0.81  $\mu\text{m}$ ; G: Band 23 with a wavelength of 0.68  $\mu\text{m}$ ; and B: Band 11 with a wavelength of 0.56  $\mu\text{m}$ ). In Fig. 2(a), the black region represents the fish pond area, the yellow (light gray in Print Edition) region represents the yellow grass area, and the white region represents the inhabited area. 16 bands of noise data were removed.

We chose spectral data, PCA data, and wavelet texture for multi-kernel SVM classification. After we applied PCA, the first four PCs were chosen<sup>[3]</sup>. In the following experiments, sole-kernel, multi-kernel, and cross-information kernel were constructed. Figure 3 shows the flow chart of the algorithm. For simplicity, 0.5 was assigned as  $\beta_k$  in both features. As the penalization and RBF parameters were difficult to choose, we used an improved method in all cases. First, we set the penalization range  $C = \{10^{-1}, 1, 10, \dots, 10^3\}$ , then we searched each RBF parameter  $\delta$  in the range  $\delta = [10^{-1}, \dots, 10^2]$ . Although time-consuming, the method worked well. We used one-against-one multi-classification for SVM.

After normalizing all the data, training and test samples were selected. Taking into account all spectral and texture features, the pixel purity index (PPI) of the data was computed. One thousand purity pixels were

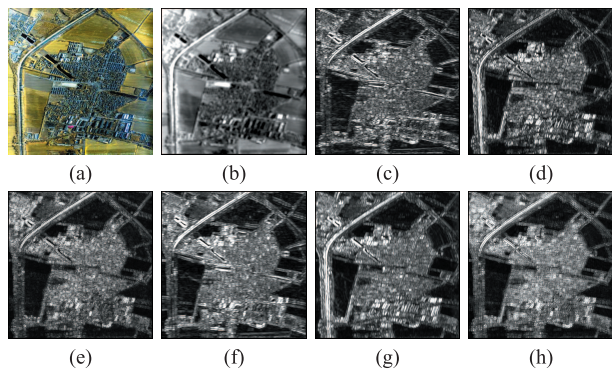


Fig. 2. (a) Original remote sensing image and (b)–(h) 2D wavelet texture images.

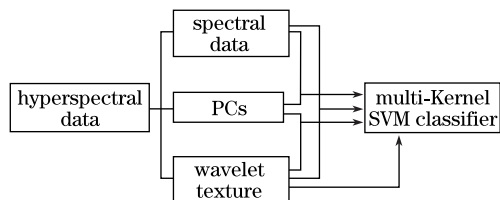


Fig. 3. Flow chart of the experiment.

obtained and chosen as samples. In the experiments, these purity pixels, together with our ground truth, were taken as training and test samples.

The classification problem involved the identification of six land cover types for the OMIS II data set: C1: inhabited area (89 training pixels), C2: crop land (92 training pixels), C3: plant (88 training pixels), C4: road (84 training pixels), C5: vegetable (74 training pixels), and C6: water (101 training pixels). The training set was 0.201% of the original dataset. In addition, the test samples, which were obtained in the same manner, were about 0.168% of the dataset: C1: inhabited area (71 pixels), C2: crop land (69 pixels), C3: plant (76 pixels), C4: road (72 pixels), C5: vegetable (73 pixels), and C6: water (79 pixels).

For the sole-kernel SVM classifier, only spectral features were used. In all experiments, we used the RBF kernel for both spectral and texture features because of its superiority over other kernel functions<sup>[28]</sup>. Finally, different features were added in one dataset: spectral dimensions, the first four PCs of PCA, WT1, and WT2. Table 1 shows the classification accuracy and Kappa coefficients<sup>[3,8]</sup>.

With multi-kernel classification, we have different combinations, such as spectral data and wavelet data, four PCs, wavelet data, etc. However, we used the RBF kernel in each dataset. Table 2 shows the classification accuracy and Kappa coefficients.

Finally, the cross-information kernel was applied. However, different data should have the same dimensions; hence, the four PCs and wavelet textures were adopted. Table 3 shows the classification accuracy and Kappa coefficients. Figure 4 shows that the classification results of different SVM kernels in each experiment are the most accurate.

Table 1. Classification Accuracy of the Sole-Kernel SVM

Dataset	Accuracy (%)	Kappa	$C, \delta$
Spectral Data	93.63	0.9236	100, 1
Four PCs	93.64	0.9236	100, 1.07
WT1	79.77	0.7574	120, 0.84
WT2	75.68	0.7081	100, 0.15
Spectral Data + WT1	96.36	0.9563	100, 0.125
Spectral Data + WT2	95.45	0.9454	100, 0.18
Four PCs + WT1	94.77	0.9372	80, 0.019
Four PCs + WT2	92.04	0.9045	100, 0.91
WT1 + WT2	80.68	0.7683	100, 0.125
All Datasets	93.19	0.9181	100, 1.1

Table 2. Classification Accuracy of the Multi-Kernel SVM

Dataset	Accuracy (%)	Kappa	$C, \delta_1, \delta_2$
Spectral Data + WT1	94.55	0.9345	50, 0.1, 0.9
Spectral Data + WT2	93.18	0.9181	80, 0.25, 0.9
Four PCs + WT1	97.05	0.9427	100, 0.85, 0.8
Four PCs + WT2	93.64	0.9236	150, 1.8, 0.8
WT1 + WT2	86.14	0.7738	100, 0.12, 0.8

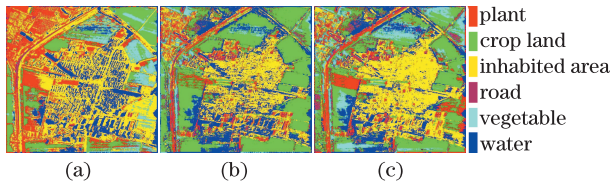


Fig. 4. Classification image. (a) Classification image of sole-kernel SVM using spectral data and WT1, (b) classification image of multi-kernel SVM using the first four principal components and WT1, (c) classification image of cross-information kernel SVM using the first four PCs and WT1.

**Table 3. Classification Accuracy of the Cross-Information Kernel SVM**

Dataset	Accuracy (%)	Kappa	$C, \delta 1, \delta 2$
Four PCs + WT1	94.77	0.9372	80, 0.6, 1.0
Four PCs + WT2	91.82	0.9018	170, 0.5, 1.0
WT1 + WT2	80.23	0.7628	160, 0.5, 0.95

From Tables 1–3, all multi-kernel SVMs obtained better accuracy than sole-kernel SVM and cross-information kernel. The best result was produced by the summation kernels, which used the first four PCs of PCA and WT1 for classification, yielding a total accuracy of 97.06% and a Kappa coefficient of 0.9472. We also compared the computing time for different kernel functions and ran each algorithm 10 times. The average time was chosen as the final training time of each algorithm. The training time of the proposed algorithm was 196.9 s. When cross-information kernel was used, the combination of the four PCs and WT1 also obtained the best results in its group, but the training time was 440.3 s. For sole-kernel SVM, the best classification accuracy was 96.36% when spectral data and WT1 data were applied, and the training time was 164.5 s. In short, good classification results will be ensured if the appropriate multi-kernels are selected. Specifically, summation kernels can reach the best accuracy and consume shorter training times.

In conclusion, SVM classifiers using different kernels are applied to hyperspectral data to test their applicability. Extracted features are shown to improve SVM classification, thus presenting an alternative to what Hughes terms “the curse of dimensionality”. In these cases, multi-kernel approaches show results superior to those achieved by sole-kernel approaches. With good results, more sophisticated cross-kernel approaches are applied in our experiments. Note that the parameters are selected using an algorithm. Results from these experiments show that wavelet texture classification is a recommended application for use with hyperspectral data. Future work should consider multi-temporal data and the kernel-based theory.

This letter was completed with the support of the Center for International Earth Science Information Network (CIESIN), Columbia University, USA. This work was supported by the National Natural Science Foundation of China (Nos. 40401038 and 40871195), the National High-Tech Program of China (No. 2007AA12Z162), Jiangsu Provincial Innovative Planning (No. CX08B\_112Z), and the Fundamental Research Funds for the Central Universities (2010QNA18).

## References

- G. Camps-Valls and L. Bruzzone, *IEEE Trans. Geosci. Remote Sensing* **43**, 1351 (2005).
- D. Landgrebe, *IEEE Signal Processing Magazine* **19**, (1) 17 (2002).
- K. Tan and P. Du, *Spectrosc. Spectral Anal.* **28**, 2009 (2008).
- G. F. Hughes, *IEEE Trans. Information Theory* **14**, 55 (1968).
- G. Camps-Valls, L. Gomez-Chova, J. Munoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, *IEEE Geosci. Remote Sensing Lett.* **3**, 93 (2006).
- Y. Bazi and F. Melgani, *IEEE Trans. Geosci. Remote Sensing* **44**, 3374 (2006).
- G. M. Foody and A. Mathur, *Remote Sensing of Environment* **93**, 107 (2004).
- F. Melgani and L. Bruzzone, *IEEE Trans. Geosci. Remote Sensing* **42**, 1778 (2004).
- K. Tan and P. J. Du, *J. Infrared Millim Waves* **27**, 123 (2008).
- G. Camps-Valls, L. Gómez-Chova, J. Munoz-Marí, J. L. Rojo-Alvarez, and M. Martínez-Ramón, *IEEE Trans. Geosci. Remote Sensing* **46**, 1822 (2008).
- S. Arivazhagan, L. Ganesan, and S. P. Priyal, *Pattern Recognition Lett.* **27**, 1976 (2006).
- Y. O. Ouma, J. Tetuko, and R. Tateishi, *Int. J. Remote Sensing* **29**, 3417 (2008).
- S. Arivazhagan and L. Ganesan, *Pattern Recognition Lett.* **24**, 3197 (2003).
- S. R. Garrity, L. A. Vierling, A. M. S. Smith, M. J. Falkowski, and D. B. Hann, *Can. J. Remote Sensing* **34**, S376 (2008).
- X. Huang, L. Zhang, and P. Li, *Photogramm. Eng. Remote Sensing* **74**, 1585 (2008).
- G. Mercier, G. Moser, and S. B. Serpico, *IEEE Trans. Geosci. Remote Sensing* **46**, 1428 (2008).
- L. Gueguen and M. Datcu, *IEEE Trans. Geosci. Remote Sensing* **45**, 827 (2007).
- I. Daubechies, *IEEE Trans. Information Theory* **36**, 961 (1990).
- L. Jiang, Y. Luo, J. Zhao, and T. Zhuang, *Chin. Opt. Lett.* **6**, 495 (2008).
- Y. Wang, T. Liu, and J. Jiang, *Chin. Opt. Lett.* **6**, 657 (2008).
- Z. Du, X. Jin, and Y. Yang, *HVAC&R Res.* **14**, 959 (2008).
- B. Scholkopf, C. Burges, and A. Smola, (eds.) *Advances in Kernel Methods: Support Vector Learning* (MIT Press, Cambridge, 1999).
- V. N. Vapnik, *Statistical Learning Theory* (Springer, New York, 1998).
- S. G. Mallat, *IEEE Trans. Pattern Anal. Machine Intell.* **11**, 674 (1989).
- W. Y. Ma and B. S. Manjunath, in *Proceedings of the IEEE Int. Conf. Image Processing* 256 (1995).
- Z.-Z. Wang and J.-H. Yong, *IEEE Trans. Image Processing* **17**, 1421 (2008).
- S. Borah, E. L. Hines, and M. Bhuyan, *J. Food Eng.* **79**, 629 (2007).
- V. Manian and R. Vasquez, *Pattern Recognition* **31**, 1937 (1998).