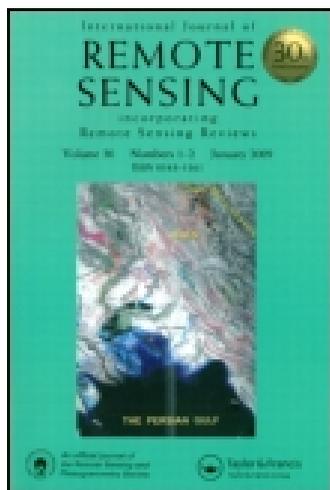


This article was downloaded by: [Tufts University]

On: 29 October 2014, At: 14:51

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tres20>

Information fusion of aerial images and LIDAR data in urban areas: vector-stacking, re-classification and post-processing approaches

Xin Huang^a, Liangpei Zhang^a & Wei Gong^a

^a The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, People's Republic of China

Published online: 06 Feb 2011.

To cite this article: Xin Huang, Liangpei Zhang & Wei Gong (2011) Information fusion of aerial images and LIDAR data in urban areas: vector-stacking, re-classification and post-processing approaches, *International Journal of Remote Sensing*, 32:1, 69-84, DOI: [10.1080/01431160903439882](http://dx.doi.org/10.1080/01431160903439882)

To link to this article: <http://dx.doi.org/10.1080/01431160903439882>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms &

Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Information fusion of aerial images and LIDAR data in urban areas: vector-stacking, re-classification and post-processing approaches

XIN HUANG*, LIANGPEI ZHANG and WEI GONG

The State Key Laboratory of Information Engineering in Surveying, Mapping
and Remote Sensing, Wuhan University, Wuhan, People's Republic of China

(Received 2 June 2008; in final form 27 February 2009)

This research investigates information fusion approaches of high-resolution aerial images and elevation data from Light Detection and Ranging (LIDAR) for urban-environment mapping. Three feature fusion methods are proposed and compared: (1) the vector-stacking approach that combines spectral and LIDAR features in one classifier; (2) the re-classification approach that firstly processes spectral signals in a classifier and then integrates its output with LIDAR features to obtain the final result and (3) the post-processing approach that uses the LIDAR data to refine the results of spectral classification. The height features used in the above three algorithms are extracted from the LIDAR digital surface model (DSM) image; these include elevation difference, maximum and minimum values, variance and the grey-level co-occurrence matrix (GLCM) textures. In addition, the average height from object-based segmentation is also computed. In the experiments, support vector machines (SVMs) are used as classifiers for all fusion schemes due to their capability and robustness for many classification problems. The three algorithms are evaluated using a 40-cm spatial resolution digital orthophoto and the corresponding LIDAR data of Odense, Denmark. In the experiments, the vector-stacking method with the Maximum–Minimum (Max–Min) feature, the re-classification method with the Max–Min feature and the post-processing approach obtain promising results (94.7%, 95.0% and 94.6%, respectively), which are significantly higher than the spectral-only classification (82.5%).

1. Introduction

In recent years, the processing techniques for very-high-resolution (VHR) imagery have received much attention since this new data can provide a large amount of detailed ground information. However, the availability of this type of data poses challenges to image information extraction and classification (Huang *et al.* 2007a). It increases the internal spectral variability (intra-class variability) of each land-cover class, and decreases the spectral variability between different classes (inter-class variability), which leads to a reduction in the statistical separability of the different land-cover classes in the spectral domain (Bruzzone and Carlin 2006). The classification of VHR data based on spectral features is insufficient due to the fact that some different objects have similar spectral reflectances (e.g. buildings and streets, grass and trees, water and shadow). Therefore, spatial features have been used to counter the inadequacy of spectral signals, such as the grey-level co-occurrence matrix (GLCM)

*Corresponding author. Email: huang_wlu@163.com

(Puissant *et al.* 2005, Dell'Acqua and Gamba 2006, Huang *et al.* 2007b), shape and structural features (Benediktsson *et al.* 2005, Huang *et al.* 2007a) and object-based analysis (Gao *et al.* 2006, Huang and Zhang 2008).

The objective of this study is to consider Light Detection and Ranging (LIDAR) data as a complementary source of spectral signals to increase the accuracy of land-cover mapping in urban environments. Some studies have been reported to fuse LIDAR and high-resolution imagery for urban-feature classification. Gamba and Houshmand (2002) used the Fuzzy C-means approach for joint classification of aerial photo and LIDAR data with four classes (buildings, vegetation, roads and open areas). The results showed that the joint classification gave much higher accuracy than the purely spectral method (the overall accuracy increased from 68.9% to 79.5%). By adding the LIDAR image, the commission accuracies for roads, open areas and buildings were significantly improved, while the accuracy for the vegetation decreased slightly. Hodgson *et al.* (2003) mapped and evaluated the imperviousness of land parcels using high-spatial-resolution colour orthophotography and surface-cover height extracted from LIDAR data. Experiments revealed that the maximum-likelihood per-pixel classification yielded a low standard error (6.62%), while a per-segment approach with a rule-based classification resulted in slightly better errors (5.85%). Nguyen *et al.* (2005) considered LIDAR data as an additional source of information in the Hopfield neural network (HNN). A height function based on the Gaussian distribution was added to the energy function of the HNN for super-resolution mapping. Secord and Zakhor (2007) presented an approach for tree detection in registered aerial and range data obtained via LIDAR. A region-growing algorithm was used to segment the image and some features extracted from the LIDAR image (e.g. height value, local height variation) were employed for classification. Geerling *et al.* (2007) integrated the Compact Airborne Spectrographic Imager (CASI) and LIDAR to classify different vegetation classes. The spectral information from the CASI channels was stacked with the texture bands extracted from the LIDAR image using the maximum-likelihood classifier. Results showed that fusion of CASI and LIDAR data could improve the classification significantly more than using spectral or LIDAR information alone.

In this research, we focus on feature-extraction algorithms from LIDAR data and spectral-LIDAR information fusion approaches. Support vector machines (SVMs) are used as classifiers of the fusion because they are non-parametric and fast-learning algorithms and are not constrained to prior assumptions about the distribution of input data; hence, they are well suited for multisource data. In this paper, three fusion methods are designed as follows:

1. Vector stacking: the statistical and textural features are extracted from the LIDAR digital surface model (DSM) image and then combined with spectral features in an SVM classifier. In this study, the texture measures of GLCM, height difference, height variance and Maximum–Minimum (Max–Min) values within a local area are extracted from the LIDAR image. In addition, the average height resulting from an object-based segmentation is also calculated.
2. Re-classification: the spectral features from the aerial image are firstly classified using an SVM, and its output is then integrated with the textural and statistical features in a second SVM.
3. Post-processing: the height information is used to refine the results of spectral classification and separate the classes with similar spectral responses.

The experiments were conducted on a 40-cm aerial orthophoto and the corresponding LIDAR DSM image of Odense in Denmark.

2. SVMs

SVMs discriminate between two classes by fitting an optimal separating hyperplane (OSH) to the training samples in a multidimensional feature space (Cortes and Vapnik 1995). SVM classifiers of the form $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$ learn from the data $\{(x_i, y_i) | x_i \in \mathbb{R}^d, y_i \in \{-1; +1\}\}$, where x_i represents the i th training sample in a d -dimensional feature space, \mathbb{R}^d , and y_i is the corresponding class label. $f(\mathbf{x})$ is the discriminant function associated with the hyperplane and is defined by a weight vector \mathbf{w} and a bias term b , with $|b|/\|\mathbf{w}\|$ representing the distance between the OSH and the origin. The support vectors lie on two hyperplanes $\mathbf{w} \cdot \mathbf{x} + b = \pm 1$ that are parallel to the OSH. The OSH is calculated by maximizing the margin of the two hyperplanes and minimizing the error:

$$\min_{\mathbf{w}, b, \xi_i} \left\{ \frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^T \xi_i \right\}, \quad (1)$$

where T is the number of training samples, and the slack variables ξ_i and the regularization parameter C are introduced to take into account non-separable data. The constant C is used as a penalty for the samples that are located on the wrong side of the hyperplane, and it controls the shape of the discriminant function. The minimization problem in equation (1) can be solved through a Lagrange dual optimization, and the final hyperplane decision function can be defined using the kernel methods:

$$f(\mathbf{x}) = \left(\sum_{i \in \mathcal{S}} \alpha_i y_i \Phi(x_i, x_j) + b \right), \quad (2)$$

where $\Phi(\cdot)$ is a kernel function and α_i are Lagrange multipliers. \mathcal{S} is the set of support vectors, which is the subset of training samples corresponding to the non-zero Lagrange multipliers. The kernel function is introduced into the SVM so that the original input space can be transformed non-linearly into a higher dimensional feature space where linear methods may be applied. A Gaussian radial basis function (RBF) is used in this study since it has been proved effective in many classification problems (Bruzzone and Carlin 2006):

$$\Phi(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad (3)$$

where γ is the RBF kernel parameter. A specific application for the SVM needs to handle two issues: the parameter optimization and the multiclass problem.

1. Parameter optimization: the kernel-based implementation of the SVM involves problems pertaining to the selection of multiple parameters, that is, the kernel parameter γ and the regularization parameter C . In this study, these parameters were selected automatically based on the leave-one-out model selection (LOOMS) algorithm (Lee and Lin 2000). This approach is based on the idea of estimating the parameters so that the estimate of the expected generalization error is minimized. Optimization is carried out using a gradient descent search

over the space of the parameters. The search was conducted between 0.1 and 1000 for C and from 0.01 to 100 for γ .

2. **Multiclass problem:** the SVM was originally designed for binary classification; however, most remote-sensing applications involve multiple classes. Two strategies are commonly used for extending SVMs to multiclass classifications: one-against-one (OAO) and one-against-all (OAA) (Foody and Mathur 2004). The OAO strategy applies a series of classifiers to each pair of classes, with the most commonly computed class reserved for the final label of each pixel. For the OAA algorithm, the K -class problem is decomposed into K binary classifiers, each focused on the recognition of one class against all the others. The final class label is determined using the maximum decision function value (i.e. the distance to the hyperplane).

3. Methods for aerial images and LIDAR information fusion

SVM-based multisource information fusion can be implemented in three levels: the feature level (vector stacking), the multiclass output level (re-classification) and the decision level (post-processing). The flow charts for the three algorithms are shown in figure 1.

3.1 Feature-level fusion: vector stacking

Vector stacking is a straightforward fusion method. Spectral signals from the aerial image are concatenated with elevation features from the LIDAR data in feature space, and the hybrid feature vectors are then classified using a multiclass SVM classifier. The processing flow is shown in figure 1(a). Let F^{spe} and F^{ele} be the feature sets of spectral and elevation information extracted from aerial and LIDAR data, respectively. The vector-stacking approach can be described as:

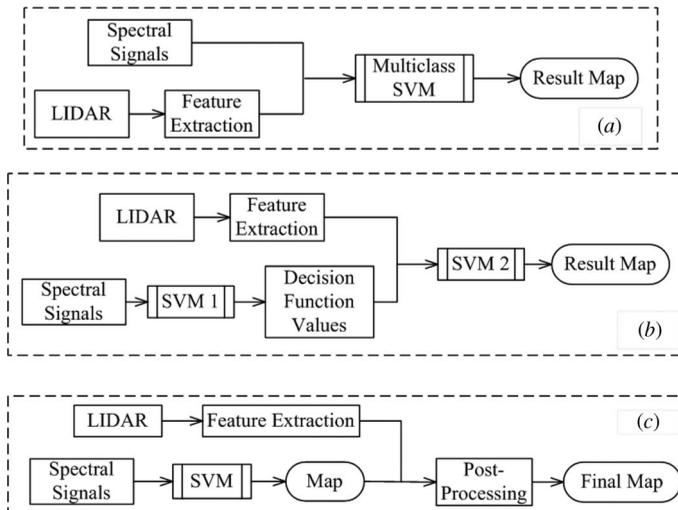


Figure 1. Flow charts for the three fusion algorithms (a) the vector-stacking, (b) re-classification and (c) post-processing approaches.

$$\mathbf{x} \in k \Leftrightarrow k = \text{Cla}(F(\mathbf{x})), \quad \text{with } F(\mathbf{x}) = (F^{\text{spe}}(\mathbf{x}) \cup F^{\text{ele}}(\mathbf{x})), \quad (4)$$

where $\text{Cla}(F(\mathbf{x}))$ denotes the decision label of input vector \mathbf{x} , with the feature set $F(\mathbf{x})$, and k is the class label ($1 \leq k \leq K$). There are two important issues for the vector-stacking fusion. The first one is how to exploit the information from the LIDAR data effectively. Due to the fact that anything can occur at any height, it is not reasonable to assume that features with the same spectral property have the same or different heights. Therefore, in this study, some statistical and textural features are extracted from the LIDAR DSM image. The other issue is the classifier for the hybrid and multisource information. The SVM classifier is a fast and robust machine-learning algorithm and is insensitive to the prior-distribution assumption. Furthermore, it can always find the optimal solution and enable the weighting of the different sources. Therefore, it is worthwhile testing its performance for aerial-LIDAR data fusion.

In this research, the statistical and textural features extracted from the LIDAR image include elevation difference, elevation variance, Max–Min values and GLCM measures. In addition, the average height resulting from the object-based segmentation is also considered.

3.1.1 Statistical features. Elevation difference (Diff) and elevation variance (Var) describe the local height variation. Diff is defined as the difference between the maximum and minimum height value in a local region, while Var is the variance of height values in the region. Due to the complex structures and the multiple returns of laser pulse, the height variation features are expected to discriminate trees with other solid or homogeneous object, such as roofs and grass. In addition, the maximum and minimum values in the local area are calculated. The Max–Min feature is capable of reducing local height variation, representing main characteristics in this area and delineating the shape and structures of objects.

3.1.2 Textural features. Textural attributes are extracted using the GLCM (Puissant *et al.* 2005, Huang *et al.* 2007b). The GLCM is a tabulation of how often different combinations of pixel grey levels occur in a local area. In this research, GLCM texture statistics are calculated based on the LIDAR DSM image with an inter-pixel distance of 1 and with different window sizes. The directional effects are removed by averaging the extracted features over four directions. Four measures, homogeneity (HOM), angular second moment (ASM), entropy (ENT) and dissimilarity (DIS), are chosen to describe the texture information of the LIDAR image:

$$\begin{aligned} \text{HOM} &= \sum_i \sum_j \frac{P(i,j)}{1 + (i-j)^2}, & \text{ASM} &= \sum_i \sum_j (P(i,j))^2, \\ \text{ENT} &= - \sum_i \sum_j P(i,j) \ln(P(i,j)) & \text{and} & \text{DIS} = - \sum_i \sum_j P(i,j) |i-j|, \end{aligned} \quad (5)$$

where (i, j) is the coordinate in the co-occurrence matrix space and $P(i, j)$ is the co-occurrence matrix value at (i, j) . Homogeneity is a measure of lack of variability or the amount of local similarity, and angular second moment is also a measure of local homogeneity. Correspondingly, entropy and dissimilarity are heterogeneity indices. Entropy is a measure of the degree of disorder in an image, and dissimilarity represents the degree of spread of the grey levels or the average grey-level difference between neighbouring pixels. Dissimilarity and homogeneity are inversely correlated. It is interesting to test the performance of GLCM values on LIDAR data since there is

a variation of height for tree objects but no variation for other classes. The information of height variation is potential to distinguish buildings and grasses, which have homogeneous height, from trees, which have textured height information.

3.1.3 Object-based approach. The object-based analysis aims to group the spatially adjacent pixels into homogeneous objects and it has been successfully applied to the VHR image processing. In this study, the object-based analysis is used to segment the image first, and the average height within each image segment is then extracted. The object-based approach is able to reduce local elevation variation and generalize elevation information in a spatial neighbour. In experiments, the fractal net evolution approach (FNEA) (Hay *et al.* 2003) is employed for segmentation. It uses fuzzy-set theory to extract the objects of interest, at the scale of interest, segmenting images simultaneously at both fine and coarse scales (Hay *et al.* 2003). The FNEA is a bottom-up region merging technique starting from a single pixel. In an iterative way, at each subsequent step, image objects are merged into larger ones. The region merging decision is made with local homogeneity criteria. When a possible merge of a pair of image objects is examined, the fusion heterogeneity value between the two objects is calculated and compared to a pre-defined scale parameter.

3.2 SVM output-level fusion: re-classification

The re-classification method integrates the elevation features extracted from LIDAR data and the output of the SVM with only an aerial image (SVM1) in a second SVM (SVM2). The flow chart is shown in figure 1(b). The re-classification method can be defined as:

$$x \in k \Leftrightarrow k = \text{Cla}(F(\mathbf{x})) \text{ with } \text{Cla} = \text{SVM2} \text{ and } F(\mathbf{x}) = (O^{\text{spe}}(\mathbf{x}) \cup F^{\text{ele}}(\mathbf{x})), \quad (6)$$

where $O^{\text{spe}}(\mathbf{x})$ denotes the output of SVM1 and $F^{\text{ele}}(\mathbf{x})$ represents the elevation features extracted from the LIDAR DSM image. In terms of the outputs of SVM1, the re-classification method can be divided into two schemes, namely crisp and soft re-classifications. The crisp method fuses the elevation features and the class label of each pixel output by SVM1, while the soft approach combines the elevation features and the multiclass output of SVM1, which are the discriminant function values defined in equation (2) (i.e. the distance between the sample point and the hyperplane).

The re-classification can be considered as a second classification by integrating the height features and the results of pre-classification. The soft output of SVM1 can be regarded as a class-specific data transformation, by which, the image data are transformed into a new feature space that is made up of the distance values of the individual SVM rule images. The multiclass output results are better comparable than the original feature space and are more appropriate to represent the inter-class difference (Waske and van der Linden 2008). Therefore, it is worth testing whether the SVM re-classification approach can integrate spectral and LIDAR features effectively.

3.3 Decision-level fusion: post-processing

The basic idea of decision fusion is to exploit the elevation features from LIDAR data to improve the spectral classification (figure 1(c)). The spectral classification is firstly performed by implementing the SVM classifier on the aerial image, and then its result

is refined using the conditional probability of height distribution. In this study, the average height features within each segment are used for classification refinement, and the elevation feature of each land-cover class is modelled as a Gaussian distribution:

$$P(k|h(x)) = \frac{1}{\sqrt{2\pi} \delta_k} \exp\left(-\frac{(h(x) - \mu_k)^2}{2 \delta_k^2}\right), \quad (7)$$

where $P(\cdot)$ denotes probability, $h(x)$ is the elevation information of point x and μ_k and δ_k are the mean and standard deviation of class k , respectively. The parameters in equation (7) are estimated based on training samples. The probability functions are used to discriminate the classes with similar spectral response but with different height characteristics; consequently, the post-processing method can be described as:

$$x \in k \Leftrightarrow k = \arg \max_{c_h^c} \langle P(c_h|h(x)) \rangle, \quad \text{for } x \in (c = [c_h]_{h=1}^H), \quad (8)$$

where c represents the set of spectrally similar classes ($c_1, \dots, c_h, \dots, c_H$).

The post-processing approach is based on the assumption that some classes (e.g. trees and buildings) are higher than their surroundings and hence the height distribution is used to discriminate some spectrally similar classes and improve the results of spectral classification.

4. Experiments and analysis

Experiments were conducted on a 40-cm spatial resolution digital orthophoto of Odense provided in 1999 by COWI, Denmark (figure 2(a)) and the co-registered DSM image acquired by LIDAR TopoSys in 2001 (figure 2(b)). Three visible channels were available for the aerial orthophoto. The LIDAR data were produced by the second return of the laser pulse. The accuracy of the DSM was 15 cm (vertical) and 50 cm (planimetric). The image is used as demonstration data for the eCognition software (Trimble Navigation Ltd, Sunnyvale, CA, USA). This dataset shows a typical

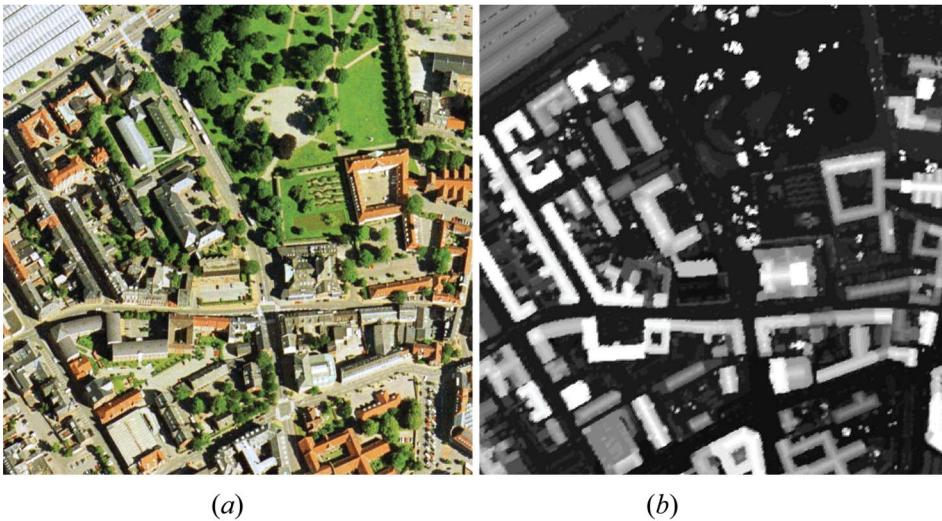


Figure 2. (a) 1021×1021 pixel digital orthophoto and (b) LIDAR DSM image.

Table 1. Number of training and test samples.

Class	Number of training samples	Number of test samples
Ground	100	2252
Grass	93	2206
Shadow	107	1751
Buildings	101	2198
Trees	101	1832
Total	502	10 239

urban environment. Five information classes are defined in this area: low-lying impervious area (ground), high-lying impervious area (buildings), grass, trees and shadow, where ground and buildings and grass and trees have similar spectral attributes. The available training and test samples are provided in table 1. The samples have a uniform distribution all around the scene, and they are converted to regions of interest on the co-registered aerial and LIDAR images. In this work, the commercial software eCognition was used to implement the object-based segmentation and obtain the average height information from LIDAR image. The crisp and soft outputs of the SVM were implemented based on ENVI 4.4 (ITT Visual Information Solutions, Boulder, CO, USA).

The objectives of the experiment are to evaluate the three fusion methods with different parameters and find an appropriate way to fuse the LIDAR and aerial images. In order to validate the information-fusion algorithms, the spectral classification is used as a benchmark, and its classification map and confusion matrix are shown in figure 3 and table 2(a), respectively. By analysing the results, it can be seen



Figure 3. Map of spectral-only classification obtained by performing SVMs on the aerial image (overall accuracy = 82.5%).

Table 2. Confusion matrices for: (a) the spectral-only classification approach (figure 3); (b) the soft re-classification approach with Max-Min feature (figure 6); and (c) the post-processing approach (figure 7(a)).

(a) Class	Ground	Grass	Shadow	Buildings	Trees	Total
Ground	1756	1	0	623	5	2385
Grass	0	2136	0	1	476	2613
Shadow	0	0	1742	2	4	1748
Buildings	496	0	9	1468	0	1973
Trees	0	69	0	104	1347	1520
Total	2252	2206	1751	2198	1832	10 239

Overall accuracy = 82.5%, Kappa coefficient = 0.781.

(b) Class	Ground	Grass	Shadow	Buildings	Trees	Total
Ground	2231	0	0	87	1	2319
Grass	0	2201	0	0	232	2433
Shadow	0	0	1745	1	1	1747
Buildings	21	0	6	2098	146	2271
Trees	0	5	0	12	1452	1469
Total	2252	2206	1751	2198	1832	10 239

Overall accuracy = 95.0%, Kappa coefficient = 0.937.

(c) Class	Ground	Grass	Shadow	Buildings	Trees	Total
Ground	2246	1	9	57	5	2318
Grass	0	2136	0	0	297	2433
Shadow	0	0	1742	2	2	1746
Buildings	6	0	0	2031	0	2037
Trees	0	69	0	108	1528	1705
Total	2252	2206	1751	2198	1832	10 239

Overall accuracy = 94.6%, Kappa coefficient = 0.932.

that misclassifications mainly occur in ground and buildings and trees and grass, as they show similar spectral reflectance.

4.1 Experiments of vector-stacking fusion

This section aims to evaluate the performance of different features with different parameters for the vector-stacking fusion. In order to analyse the window-size effects, we computed the statistical and textural features with different sizes of moving windows (from 3×3 to 21×21). The overall accuracy (OA) based on the confusion matrix was used to assess the algorithms. Figure 4 shows the accuracies of the

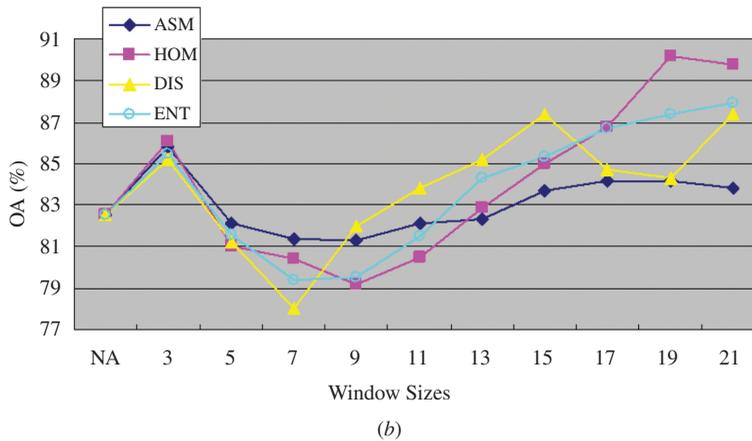
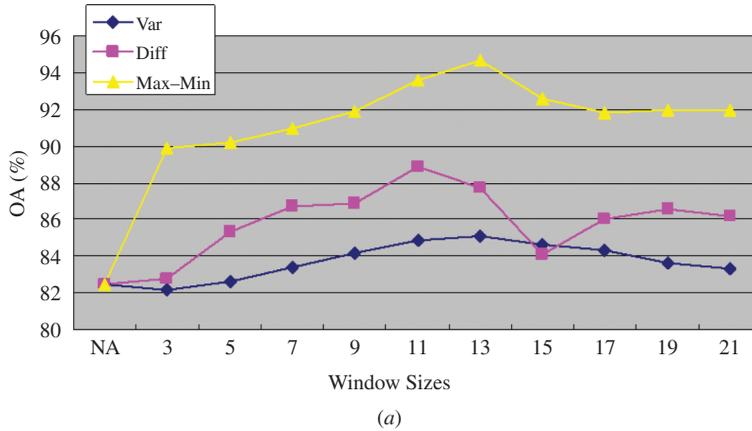


Figure 4. (a) Overall accuracy (OA) of different statistical measures (variance (Var), height difference (Diff) and maximum–minimum values (Max–Min)) for different window sizes. (b) OA of different GLCM textures (ASM, HOM, DIS and ENT) with different windows.

statistical and GLCM textural features with different windows, respectively. In both figures, ‘NA’ on the horizontal axis denotes the spectral-only classification (OA = 82.5%). By analysing figure 4(a), it is interesting to find that the Max–Min statistic obviously outperformed other features such as height difference and variance. The Max–Min operator with a 13×13 window size gave very promising accuracies (OA = 94.7% and Kappa coefficient = 0.934). The reasons may be that it can effectively describe the main characteristics in a local region, and it is able to reduce local height variation and hence improve the classification in homogeneous textures (e.g. buildings, grass). In addition, it helps to reduce the effects resulting from the returns of the laser pulse in trees. On the other hand, Diff with an 11×11 window gave 88.9% for OA and 0.861 for Kappa, which were also satisfactory. As for the height variance (Var), it did not achieve significantly higher accuracies than the spectral classification (82.5%). The best result of Var was 85.1% for OA and 0.813 for Kappa with a 13×13 window. By looking at figure 4(b), it can be found that homogeneity measures (e.g. HOM and ASM) gave comparable results with the

Table 3. Class-specific accuracies of different features for vector-stacking fusion. All values are percentages (%).

Class	Spectral method	HOM (19×19)		Ave-H		Max-Min (13×13)	
		Accuracy	Δ (%)	Accuracy	Δ (%)	Accuracy	Δ (%)
Ground	78.0	72.3	-5.7	99.5	21.5	98.9	20.9
Grass	96.8	99.1	2.3	95.5	-1.3	99.7	2.9
Shadow	99.5	99.7	0.2	99.7	0.2	99.6	0.1
Buildings	66.8	89.9	23.1	94.3	27.5	96.4	29.6
Trees	73.5	86.0	12.5	58.4	-15.1	76.9	3.4
OA (%)	82.5	90.2	7.7	90.2	7.7	94.7	12.2
AA (%)	82.9	89.4	6.5	89.5	6.6	94.3	11.4

heterogeneity measures (e.g. DIS and ENT). The HOM feature with a 19×19 window achieved the highest accuracies: 90.2% for OA and 0.877 for Kappa.

The class-specific accuracies for the spectral classification, Max-Min feature with a 13×13 window, the HOM textural measure with a 19×19 window and the average height information (Ave-H) extracted by the object-based segmentation are compared in table 3. In the table, ' Δ ' denotes the accuracy variation between vector-stacking fusion and the spectral classification and 'AA' is the average accuracy of all the information classes. From table 3, it can be seen that, compared to the spectral classification, the improvements of OA are 7.7%, 7.7% and 12.2% for the HOM, Ave-H and Max-Min features, respectively, and the respective improvements of AA are 6.5%, 6.6% and 11.4%. It is worth noting that only the Max-Min feature can improve the accuracies of all the classes. It increased the accuracies of the impervious areas substantially, by 20.9% and 29.6% for ground and buildings, respectively. Meanwhile, the accuracies of vegetation were increased by 2.9% for grass and 3.4% for trees.

4.2 Experiments of the re-classification approach

In this experiment, the crisp and soft re-classifications were implemented using different statistical measures (including height difference, height variance, Max-Min values), GLCM texture and average height. The statistical measures were calculated using a 13×13 window, and GLCM texture was computed using the homogeneity feature with a 19×19 window. The OA of the crisp and soft

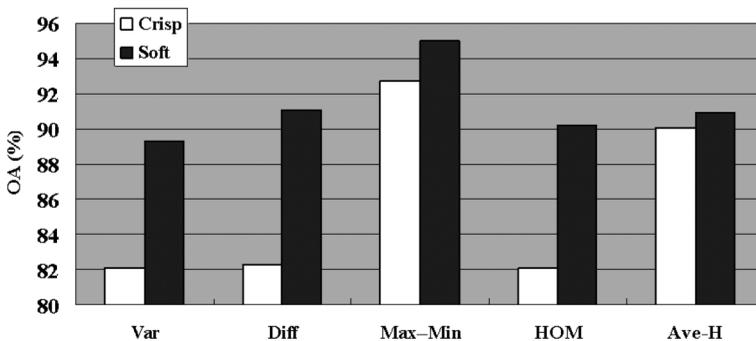


Figure 5. Accuracies of crisp and soft re-classifications for different features.



Figure 6. Classification map for soft re-classification approach with Max-Min feature (overall accuracy = 95.0%).

re-classifications for different features is shown in figure 5. By looking at the figure, we can find two interesting conclusions: (1) the Max-Min feature gives the highest accuracies for both crisp and soft methods, which verifies its effectiveness for feature extraction from LIDAR data and (2) the soft re-classification obviously outperforms the crisp one and the accuracy improvements were 7.2%, 8.8%, 2.3%, 8.1% and 0.8% for Var, Diff, Max-Min, HOM and Ave-H features, respectively. The results reveal that the decision function output of multiclass SVMs is more appropriate for information fusion than the label output by the same SVM. The soft re-classification with the Max-Min feature obtains satisfactory results: 95.0% for OA and 0.937 for the Kappa coefficient, and its classification map and confusion matrix are shown in figure 6 and table 2(b).

4.3 Experiments of the post-processing approach

For the post-processing approach, we should predefine the spectrally similar groups according to equation (8). The image was firstly classified into impervious areas, vegetation and shadow using the aerial image alone. The impervious areas included two spectrally similar classes: ground (low-lying impervious area) and buildings (high-lying impervious area), while the vegetation was then divided into grass and trees. The post-processing method was based on the conditional probabilities (equation (7)) that were estimated based on the training samples. The confusion matrix for the post-processing method is provided in table 2(c) and the classification map is shown in figure 7(a). From the results, it can be stated that the post-processing approach achieved satisfactory classification (OA = 94.6%, Kappa = 0.932). Compared to the spectral-only classification, the improvements of the OA and Kappa coefficient are 12.1% and 0.151, respectively. Figure 7(b) shows where the post-processing method improves the result of the spectral classification. The classes C1, C2 and C3 represent the building, tree and ground areas that were correctly classified using the

Table 4. Comparison of class-specific accuracies for the three fusion schemes. All values are percentages (%).

Class	Spectral method	Vector stacking		Re-classification		Post-processing	
		Accuracy	Δ (%)	Accuracy	Δ (%)	Accuracy	Δ (%)
Ground	78.0	98.9	20.9	99.1	21.1	99.7	21.7
Grass	96.8	99.7	2.9	99.8	3.0	96.8	0
Shadow	99.5	99.6	0.1	99.7	0.2	99.5	0
Buildings	66.8	96.4	29.6	95.5	28.7	92.4	25.6
Trees	73.5	76.9	3.4	79.3	5.8	83.4	9.9
OA (%)	82.5	94.7	12.2	95.0	12.5	94.6	12.1
AA (%)	82.9	94.3	11.4	94.7	11.8	94.4	11.5

5. Conclusions

This study has investigated the aerial and LIDAR data fusion approaches for urban mapping. To this end, three fusion schemes were proposed: the vector-stacking method in feature space, the crisp and soft re-classifications methods and the post-processing method for refinement. These fusion schemes were carried out and compared to a 40-cm spatial resolution digital orthophoto and a co-registered LIDAR DSM image of Odense, Denmark. Based on the experiments, the following conclusions can be drawn.

1. For the vector-stacking fusion, the height difference, average height, GLCM textures and Max–Min features gave much better results than the spectral-only classification. The Max–Min statistic with a 13×13 window achieved very promising results (94.7%) because it is able to describe the main characteristics in a local region and reduce the effects of laser returns in heterogeneous areas. It is interesting to observe that, among all the features, only the Max–Min can increase the accuracies of all the classes at the same time.
2. The re-classification approaches are categorized into crisp and soft ones according to the different outputs of the SVM. Experiments revealed that the soft re-classification method obviously outperformed the crisp one because the

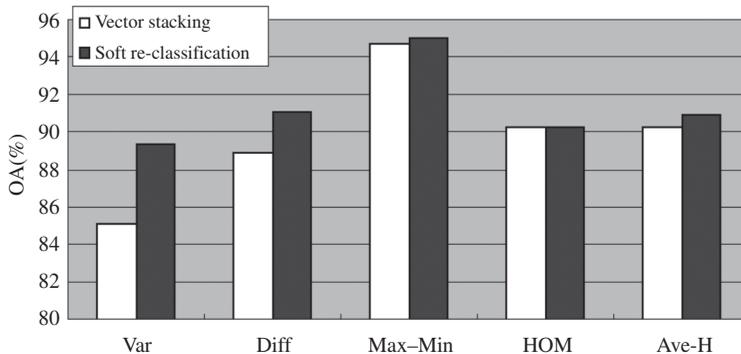


Figure 8. Comparison of vector-stacking and soft re-classification approaches with the same features used.

former exploited the discriminant function values of the multiclass SVM, but the crisp method only considered the classification label. A reasonable explanation for this phenomenon is that the discriminant function output can be regarded as a class-specific data transformation and they are more efficient to represent the inter-class difference. Furthermore, using crisp results actually outweighs the spectral-only classification outcome for the second stage, which then downplays the role of LIDAR data to some extent.

3. The post-processing approach achieved satisfactory accuracy (94.6%) since the conditional probabilities obtained from the LIDAR data were used to refine the pre-classification and discriminate the spectrally similar classes effectively.
4. By comparing the performance of different features, it can be found that the Max–Min statistic is very efficient for feature extraction from the LIDAR data since it gave the best accuracies in both the vector-stacking and re-classification schemes. With respect to the different fusion methods, it can be stated that the vector-stacking and re-classification schemes are related to the features used (e.g. Max–Min), while the post-processing approach is dependent on the prior knowledge for the test areas. With the same features, the soft re-classification method slightly outperformed the vector-stacking method, but the improvements came at the expense of greater computational time for an additional SVM. Therefore, the vector-stacking method is suggested when both time and accuracy are considered, while the re-classification method should be confined to where the performance is of most importance.

The proposed fusion algorithms have potential for not only optical-LIDAR data, but also for information fusion of other multisensors. In future, we plan to use these fusion approaches to integrate the spectral information from the optical sensor and the textural characteristics from the synthetic aperture radar data. Similarly, it is also interesting to fuse the rich spatial feature from the panchromatic image and the spectral information from the multispectral channels. In addition, we also plan to extend our work to rural areas (e.g. farmlands, agriculture).

Acknowledgements

This work was, in part, supported by the Major State Basic Research Development Program (973 Program) under Grant No. 2009CB723905 and by the National Science Foundation of China (40771139 and 40930532). Dr. Xin Huang would like to thank the Editor and referees for their helpful suggestions that significantly improved the manuscript.

References

- 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.
- 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.
- CORTES, C. and VAPNIK, V., 1995, Support vector networks. *Machine Learning*, **20**, pp. 273–297.
- DELL'ACQUA, F. and GAMBA, P., 2006, Discriminating urban environments using multiscale texture and multiple SAR images. *International Journal of Remote Sensing*, **27**, pp. 3797–3812.

- FOODY, G.M. and MATHUR, A., 2004, A relative evaluation of multiclass image classification of support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, pp. 1335–1343.
- GAMBA, P. and HOUSHMAND, B., 2002, Joint analysis of SAR, LIDAR and aerial imagery for simultaneous extraction of land cover, DTM and 3D shape of buildings. *International Journal of Remote Sensing*, **23**, pp. 4439–4450.
- GAO, Y., MAS, J.F., MAATHUIS, B.H.P., ZHANG, X. and VAN DIJK, P.M., 2006, Comparison of pixel-based and object-oriented image classification approaches – a case study in a coal fire area, Wuda, Inner Mongolia, China. *International Journal of Remote Sensing*, **27**, pp. 4039–4055.
- GEERLING, G.W., LABRADOR-GARCIA, M., CLEVERS, J.G.P.W., RAGAS, A.M.J. and SMITS, A.J.M., 2007, Classification of floodplain vegetation by data fusion of spectral (CASI) and LiDAR data. *International Journal of Remote Sensing*, **28**, pp. 4263–4284.
- HAY, G.J., BLASCHKE, T., MARCEAU, D.J. and BOUCHARD, A., 2003, A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, **57**, pp. 327–345.
- HODGSON, M.E., JENSEN, J.R., TULLIS, J.A., RIORDAN, K.D. and ARCHER, C.M., 2003, Synergistic use of Lidar and color aerial photography for mapping urban parcel imperviousness. *Photogrammetric Engineering and Remote Sensing*, **69**, pp. 973–980.
- HUANG, X. and ZHANG, L., 2008, An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. *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.
- LEE, J.H. and LIN, C.J., 2000, Automatic model selection for support vector machines. Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, ROC. Available online at: <http://www.csie.ntu.edu.tw/~cjlin/looms/>.
- NGUYEN, M.Q., ATKINSON, P.M. and LEWIS, H.G., 2005, Superresolution mapping using a hopfield neural network with LIDAR data. *IEEE Geoscience and Remote Sensing Letters*, **2**, pp. 366–370.
- 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*, **4**, pp. 733–745.
- SECORD, J. and ZAKHOR, A., 2007, Tree detection in urban regions using aerial lidar and image data. *IEEE Geoscience and Remote Sensing Letters*, **4**, pp. 196–200.
- WASKE, B. and VAN DER LINDEN, S., 2008, Classifying multilevel imagery from SAR and optical sensors by decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, pp. 1457–1466.