

# PERFORMANCE EVALUATION OF ROAD AND BUILDING CLASSIFIERS ON VHR IMAGES

C. Simler, C. Beumier

Signal & Image Centre, Royal Military Academy, 30 Av. de la Renaissance, 1000 Brussels, Belgium –  
chris\_simler@yahoo.fr

**KEY WORDS:** Support vector machine, mean shift, classifier combination, very high spatial resolution image.

## ABSTRACT:

A method is proposed for building and road detection on VHR multispectral aerial images of dense urban areas. Spatial and spectral features of segmented areas are classified using a 3-class SVM integrating some a priori and contextual information to handle unclassified patterns and conflicts. Geometrical object features and additional information improve the classification accuracy in the difficult case where many building roofs are grey like the roads and have similar geometry. Also, road network regularization is suggested to improve the classification accuracy.

## 1. INTRODUCTION

Classification accuracy on remote sensing images is important for development planning, emergency response or earth survey. A current challenge is to provide both accurate and automatic classification algorithms. We aim at extracting roads and buildings in urban area aerial images. Our images have a very high spatial resolution (VHR) of 0.5m per pixels and have three optic spectral bands (RGB multispectral channels).

Classification applications in remote sensing usually work at the pixel level, using only spectral information. Also, extracted pixel spectral features are usually classified with the Gaussian maximum likelihood (ML) supervised classifier (Bi, 06). However, VHR urban images contain a significant amount of spatial information, which should be used to make possible the precise identification of small structures such as houses or narrow roads. Contextual information can be used by means of Markov random fields (MRF) (Ja, 02), morphological profiles (MP) (Pa, 05; Fa, 08; Tu, 09) or image segmentation (Ta, 09; Li, 04; Ta, 10; Si, 10) approaches.

The MPs techniques are fast and intensively used in hyperspectral imaging. In that case, a pixel is often described by both a spectral and a geometrical pattern (the MPs), which can be concatenated to form a composite pattern before classification (Fa, 08). Geometrical features generally invalid the Gaussian assumption for class distributions and nonparametric supervised classifiers such as decision trees, K-nearest-neighbours (Bi, 06), neural networks (Fa, 2006; Be, 99) or kernel methods such as Support Vector Machines (SVMs) (Bi, 06; Fa, 08; Tu, 09) are generally used. With multispectral images, SVMs often provide a better classification accuracy than other methods (Me, 04; Fo, 04), because they handle small ratio between the number of available training samples and the number of features. However, SVMs were designed to solve binary classification problems, and multi-class SVMs are generally handled by the “one-against-all” or the “one-against-one” strategy (Bi, 06; Me, 04).

The use of image segmentation approaches overcomes the scale selection problem of the MPs methods (Ta, 10). One way to use segmentation consists in merging the segmentation map with the results of a classical pixel wise spectral classification by assigning to a segmented area the predominant pixel class within it (majority vote) (Ta, 09, Ta, 10). Other works suggest computing spectral pattern over segmented area (Ta, 10) and then (object) patterns are classified.

In the part 2 of this paper the two previous methods using segmentation are compared with the classical pixel wise spectral classification procedure, in the case where each method uses a 3-class SVM classifier handled by the “one-against-all” strategy. Unfortunately, due to the presence of class overlaps, problems of bad detection for the class “building” and of false alarm for the class “road” occur. Because we are in the context of dense urban VHR images with class “road”, “building” and “other”, class geometries are more characteristic and discriminative with respect to each other than with most of the previous works. Thus a solution is presented in part 3 consisting to perform a better exploitation of the available spatial information (provided by segmentation) by computing object geometrical features over segmented areas, and then classifying the composite spectral-spatial (object) patterns as previously with a 3-class SVM classifier. The fact to concatenate spectral and spatial patterns improves class separability. However, another limitation is that the final class attribution is performed by applying the “winner-take-all” rule to the binary SVM classifier discriminant function values (Me, 04), because this information can suffer from a lack of reliability. In order to overcome this problem, a solution using contextual information and a priori knowledge is suggested in part 3 to handle conflicts and non-assigned patterns. In part 4, a road network refinement is suggested, filling the gaps in the roads and smoothing road borders, on the basis of straight segment detection. Part 5 is the conclusion.

## 2. TRADITIONAL PIXEL AND OBJECT CLASSIFICATIONS WITH SVM

Among the numerous existing supervised nonparametric classification methods, the compact kernel SVM classifier was chosen because of its superiority in terms of classification accuracy in the context of remote sensing images, and its ability to handle the curse of dimensionality (Bi, 06; Fa, 08; Me, 04; Fo, 04; Si, 10). The Gaussian kernel provides often the best results, and is used in this paper. In this case, the SVM algorithm has two parameters (for each class): the misclassification penalty term and the Gaussian width.

### 2.1 Classification schemes

In this part, three SVM-based classification methods are explained. These methods are already used in the literature, and are the followings:

- 1- The classical pixel wise spectral classification with SVM. A pixel is first described by the 3-d RGB colour vector, and then patterns are classified with SVM.
- 2- The object spectral classification with SVM. A segmentation algorithm is applied first to the image; second the 3-d RGB mean colour vector describes segmented areas, and third (object) patterns are classified with SVM.
- 3- The resulting pixel classification map of method 1 is merged with a segmentation map established independently. The predominant pixel class within a segmented area is assigned to the whole area (majority vote).

Note that methods 2 and 3 exploit geometrical information thanks to segmentation. In this paper, the mean shift segmentation algorithm is used (Ch, 95; Co, 02; Si, 10). Details about the mean shift can be found in previous works (Si, 10), and a result is shown on figure 1.



Figure 1. Mean shift segmentation results on a part of a colour VHR aerial image.

In this paper we focus on building and road extraction. In order to achieve comparisons, both pixel and object SVM classifiers use the same 3-class training set, composed of four hundred “building”, four hundred “road” and two hundred “other” elements. It was built at the object level by manually assigning to a class some mean shift areas situated outside the classification area, and computing over each one the 3-d mean colour vector. Because the SVM algorithm is a 2-class classifier, our 3-classes are handled by the “one-against-all” multiclass SVM strategy. It consists in using three binary SVM classifiers independently, one for each class. For each class, first

a 2-class training set is built by opposing the elements of the training set of the considered class to the elements of the two other classes. Second, the two SVM parameter values have to be set. We optimize them using cross-validation, by minimizing the false classification rate over a 2D-grid of ten thousand couples of values for the two tuned parameters. This is costly but ensures to find the global minimum. In order to have a very high precision, this procedure is repeated three times in a coarse to fine scheme. Finally, the optimal values are used to learn the classifier on the entire 2-class training set. When a new pattern  $x$  is presented, each binary classifier first computes its linear discriminant function value (the SVM decision boundary is a hyperplane linear model in the final feature space (Bi, 06), which is a signed measure of the distance of  $x$  with respect to the hyperplane,  $d(x)$ ). Second, the final decision (class label) is established by looking to which side of the hyperplane  $x$  belongs; i.e. if  $d(x) > 0$ ,  $x$  is attributed to the considered class, else  $x$  is not attributed to the class. Finally, when all the patterns are classified, we have three binary-labeled images (independent binary SVM classifier final decisions), one for each class. There are eight possibilities to handle, comprising four conflict situations (multiple assignments) and the non-attribution case. A 3-class classifier is generally built by taking the final decision with the “winner-take-all” rule on the discriminant function values of the binary classifiers. This strategy enables to handle easily and automatically conflicts and unclassified patterns.

### 2.2 Experimental results

The three classification methods of part 2.1 have been evaluated on the part of a multispectral VHR aerial image of dense urban area of figure 2.a. This image of Brussels centre (Belgium) contains 2393x1804 pixels, with a spatial resolution of 0.5m per pixel, and is composed of three optic spectral bands (RGB). Figure 2.b is the ground truth built by visual interpretation. The red on the ground truth corresponds to areas where it was visually difficult to discriminate roads and buildings, and thus road or building detections on these areas are considered as exact. The classification maps obtained with the three methods of part 2.1 are shown on figure 3. Some corresponding descriptive measures of classification accuracies are reported in table 1. They were computed from the 3x3 confusion matrix (in terms of pixels) between the considered 3-class classified image (examples are on figure 3) and the 3-class ground truth image (figure 2.b).

Figure 3.a shows that the pixel wise spectral classification manifests a salt-and-paper appearance, as it is usually the case when only spectral information is used. It can be seen in table 1 that with this method, all classes suffer from a bad detection problem (especially the “building” and “other” classes). In addition, a false alarm problem occurs for the class “other”, and especially for the class “road”. The method 2 provides also poor results, with important bad detection problem for the “building” and “other” classes, and false alarm problem for the “road” class.

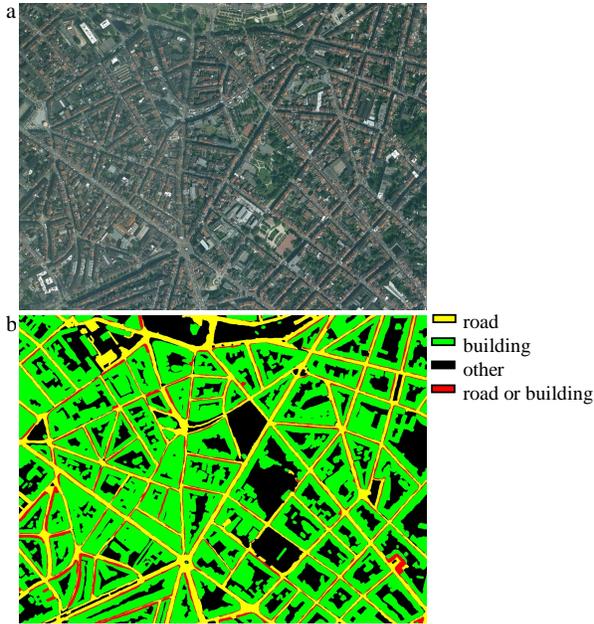


Figure 2. Brussels centre: (a) part of a colour aerial image with a spatial resolution of 0.5m, (b) ground truth.

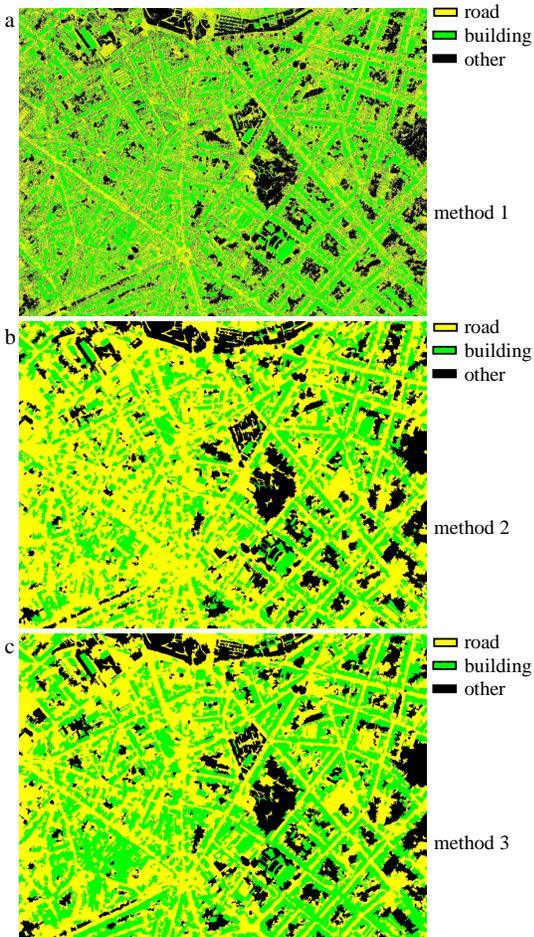


Figure 3. 3-class SVM classification map obtained by: (a) pixel spectral classification (method 1), (b) object spectral classification (method 2), (c) majority vote over areas (method 3). Classification accuracies are reported in table 1.

	Pixel spectral SVM classification (method 1)	Object spectral SVM classification (method 2)	SVM + majority vote over area (method 3)
Overall accuracy	0.56	0.54	0.60
Producer's accuracy road	0.66	0.92	0.87
Producer's accuracy building	0.55	0.43	0.55
Producer's accuracy other	0.50	0.48	0.50
User's accuracy road	0.30	0.31	0.34
User's accuracy building	0.78	0.88	0.85
User's accuracy other	0.65	0.82	0.81

Table 1. 3-class classification accuracies (in percentage). The SVM final decisions are taken with the “winner-take-all” rule on the discriminant function values of the binary classifiers.

A significant improvement is obtained with the method 3. However, results are still not satisfactory because the overall accuracy is only equal to 60%. Also, the problems of bad detection for the “building” and “other” classes and of false alarm for the “road” class remain. These problems are mainly due to the fact that in our case many building roofs and “other” areas have similar grey spectral signature as road pixels (class overlaps). In our context (urban VHR images, with class “road” and “building”), a solution to improve the class separability consists in exploiting more deeply the available geometrical information provided by segmentation. This can be performed by describing segmented areas with geometrical features and is the topic of part 3.

### 3. SPECTRAL-SPATIAL OBJECT CLASSIFICATION WITH SVM

In order to overcome the class overlap problems of the methods of part 2, we suggest computing some spatial object features. We consider the method 2 of part 2.1, and in addition with the mean colour vector we compute the area and the eccentricity of the segmented area. The eccentricity computation is described in (Si, 10). Finally, the 3-class SVM classifier operates now on 5-d composite spectral-spatial object patterns. This suggested method is called the method 4. The classification maps obtained with it is on figure 4.b, and some corresponding descriptive measures of classification accuracies are reported in table 2.

It can be seen in table 2 that the overall accuracy of the method 4 is of 6% upper than the one of the method 3 (which is the best among the previous methods, see table 1). In fact, the “building” bad detection and the “road” false alarm problems are still present but significantly attenuated. It can be noticed however that the “other” bad detection problem remains unchanged; this is because the “other” class has no specific shape and size (geometrical features are not useful for this

class). This significant improvement with respect to previous methods shows that the concatenation of the spectral pattern with area and eccentricity features improves class separability in our context. However, the problems are not totally cancelled and that shows the presence of remaining class overlaps. It seems to be difficult to solve entirely this problem because in our context many building roofs and “other” areas have both similar grey spectral signature and rectangular geometry as road objects.

Up to now, the 3-class SVM final class attribution has been performed by applying the “winner-take-all” rule to the three binary classifier discriminant function values. This is a straightforward procedure enabling to handle easily conflicts between classes (multiple assignments) and unclassified patterns. A conflict occurs when more than one discriminant function values are positives, and a non-assignment occurs when all the three discriminant function values are negatives. Such areas can be observed on figure 4.a, which is the superimposition of the three binary-labelled images obtained with the method 4. It can be noticed that with this example the conflict between “road” and “building” classes (in red) is more present than the other types of conflicts. Also, unclassified areas (in white) are relatively numerous. Discriminant function information can suffer from a lack of reliability. For instance, if all the discriminant function values are negatives, even the “winner” is on the bad side of the hyperplane. Also, in this case the “winner” is the closest from the hyperplane and the closer  $x$  is from the hyperplane, the less reliable the decision is. An idea to overcome this limitation is to use some additional information in the final class attribution process (in case of conflicts or non-assignments), in addition to the discriminant function values. In our context, we can exploit some a priori knowledge and contextual information. Because unclassified areas are often larges, it is difficult to establish by visual observation some pertinent contextual rules to handle them. However, some a priori knowledge can be used. In fact, the detected “building” and “other” bad detection and “road” false alarm phenomenon enable to assume that an unclassified area is probably not a “road” object, but rather a “building” or “other” one. This assumption has been confirmed by visual observation. Thus with unclassified areas we suggest considering only the discriminant function values of class “building” and “other” with the “winner-take-all” rule, preventing the “road” class attribution. Now conflict areas are investigated, with a focus on the road-building one because it is the single being significant. In that case, applying the “winner-take-all” rule on the three discriminant function values is theoretically suitable. However, visual observation has shown that contextual information can be advantageously used. For example, if buildings (or roads) mainly surround a conflict area, most of the time it is a building (or a road). These contextual rules have showed better efficiency than the “winner-take-all” rule on discriminant function values. The drawback is that contextual rules can be difficult to establish in case of more classes are defined, and depend on image content. It can also be noted that context exploitation is a post-classification processing step. This suggested method handling non-assignments with a priori knowledge and conflicts with contextual information is called the method 5. Its classification map is on figure 4.c, and the corresponding classification accuracies are in table 2.

It can be seen that the overall accuracy is of 3% upper than the one of the method 4, and the most significant improvement is for the class “building” (producer’s accuracy of 73%). The “other” bad detection and the “road” false alarm problems are

still present but attenuated. This significant improvement with respect to method 4 shows the importance of additional information.

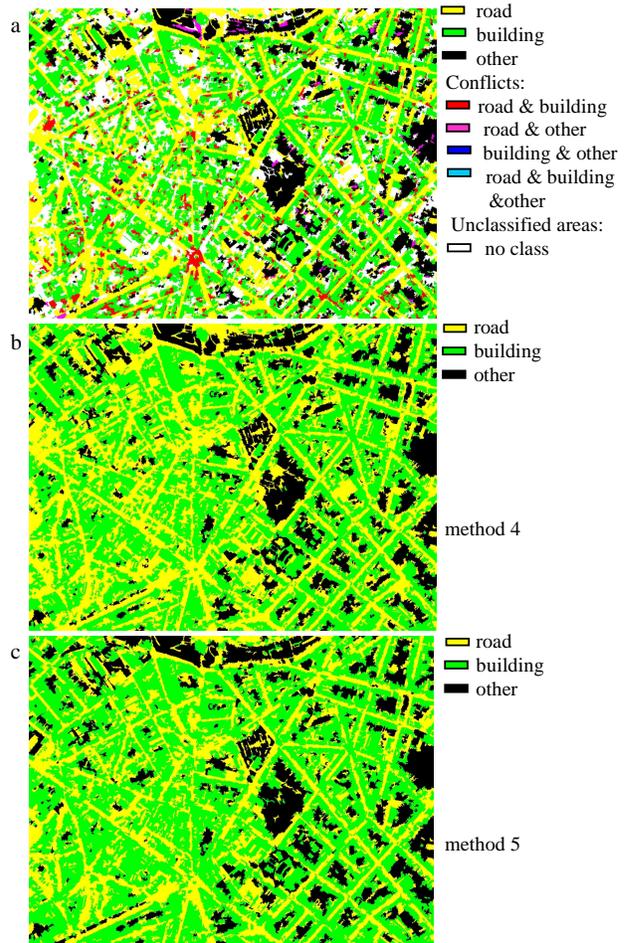


Figure 4. (a) superimposition of the tree SVM binary classifier final decisions. 3-class SVM classification map obtained by: (b) object spectral-spatial classification (method 4), (c) method 4 + a priori knowledge and contextual information (method 5). Classification accuracies are reported in table 2.

	Object spectral-spatial SVM classification with “winner-take-all” (method 4)	method 4 + a priori knowledge and contextual information (method 5)
Overall accuracy	0.66	0.69
Producer’s accuracy road	0.85	0.79
Producer’s accuracy building	0.66	0.73
Producer’s accuracy other	0.50	0.54
User’s accuracy road	0.41	0.46
User’s accuracy building	0.82	0.80
User’s accuracy other	0.82	0.79

Table 2. 3-class SVM classification accuracies (in percentage).

#### 4. POST-CLASSIFICATION REGULARIZATION

This part aims at improving the classification accuracy by performing a refinement of the road network, filling the gaps in the roads and smoothing road borders, on the basis of straight segment detection.

Consider a 3-class labelled image such as the one of figure 4.c. First, a binary image containing the road network (closed and one pixel width) boundaries is formed (the white pixels in the example of figure 5.a). Second, straight segments are detected on this image (the pixels in red in figure 5.a). Third, segments are associated according to two sets of geometrical rules forming some rectangular areas (the red and blue pixels of figure 5.b), which are then set to the class “road” in order to regularize the road network.

**Straight segment detection principle:** the user enters two parameters, which are the number of pixels of a segment and a threshold on the residual (quality of the fit). First the binary image is scanned, and for each white pixel encountered, a neighboring white pixel is searched and so on up to the imposed number of pixels is reached. Then a first test is applied to the segment: the distance between the first and the last pixel must be larger than the number of points multiplied by  $3/4$ . In case of success, the straight line model parameters are estimated by mean squares, and the squared root-mean-square error (residual) is computed in order to assess a quality measure. A second test is then applied: if the residual is lower than a threshold (provided by the user), a straight segment is detected, and parameters are stored in an array as well as the residual. At the end, we have a list of straight segments. This list is then ordered with respect to the residual values, and a non-maxima suppression with respect to the residual is performed in order to avoid aggregates. A segment is eliminated if the distance of one of these two extremities is smaller than 7 pixels from a higher residual segment, and if the extremity projection belongs to this second segment.

**Segment association principle:** two geometrical sets of rules are used to associate segments. The first handle the case of two close and collinear segments. Two segments are associated if the distance between their centers is lower than a threshold given by the user, and if the angular distance is lower than  $3^\circ$ , and if the distance from a center of one segment to the straight line of the other (and vice-versa) is smaller than 5 pixels. In case of association, a resulting straight line is build between the farthest extremities of the two segments and class “road” is attributed to the a 15-pixel width rectangle on the road side on this line if more than half of the pixels on this area are initially of class “road”. The second set of rules handle the case of two close, parallel and non-collinear segments. If the two distances from the centers to the lines belongs to the interval in are comprised between 7 and 25 pixels, and if the angular distance is lower than  $3^\circ$ , and if the distance between the segment centers is lower than a threshold given by the user, and if the mean of the centers of the two segments belongs to the road sides of the two segments, an association is performed. Then class “road” is attributed to the part between the two segments if more than half of the pixels on this area are initially of class “road”.

We have used the segment detector algorithm with a number of pixels equal to 30, and with a threshold on the residual equal to 10. Then with the segment association algorithm the two thresholds (for rule 1 and 2) were equal to 180 and 100 pixels. This four parameter configuration is noted  $\{30, 10, 180, 100\}$ . Two other configurations were tested,  $\{50, 20, 380, 200\}$  and  $\{80, 100, 600, 400\}$ . Results were combined to work with several segment lengths (figure 5.b shows the final

combination). It can be seen in table 3 that the overall accuracy, and most of the other measures, are better than without regularization (compare with method 4 in table 2).

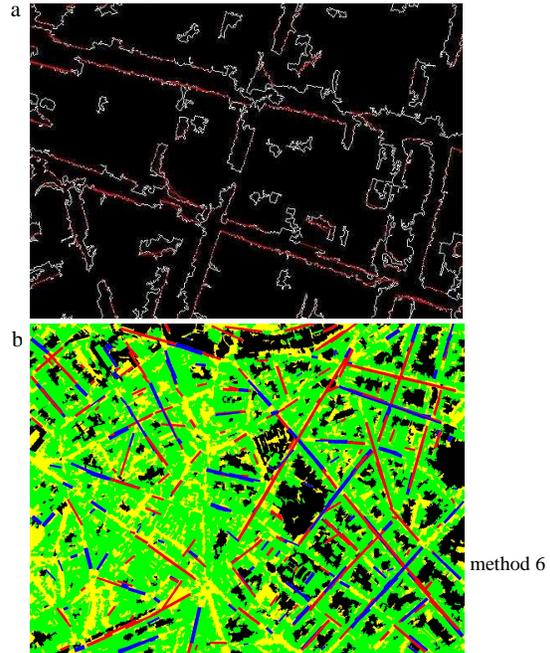


Figure 5. (a) white: road network boundaries on a part on the image of figure 2.a, red: straight line segment detected ; (b) the red and blue rectangle areas (resulting from two different segment associating rules) are set to class “road” for road network regularization (method 6).

	method 5 + segment regularization (length 30, 50 and 80 pixels) (method 6)
Overall accuracy	0.72
Producer's accuracy road	0.81
Producer's accuracy building	0.75
Producer's accuracy other	0.55
User's accuracy road	0.49
User's accuracy building	0.81
User's accuracy other	0.79

Table 3. Classifier of method 5 with regularization classification accuracies (in percentage).

#### 5. CONCLUSION

Building and road detection on VHR aerial images of dense urban areas has been investigated. The suggested approach contains segmentation and classification algorithms especially well adapted to multispectral data, and both spatial and spectral information are used at the object level. The full exploitation of the geometry improves class separability, attenuating the bad detection and the false alarm problems. However, problems are still present because in our case even geometry is not enough to suppress class overlaps. A second suggestion is to integrate a priori knowledge and contextual information around objects in

the decision process, attenuating again these problems, above all for the “building” class. Finally, a post-classification algorithm has been suggested to regularize the road network, improving the classification accuracy.

#### REFERENCES

C. M. Bishop, Pattern recognition and machine learning, Springer, ISBN: 0-387-31073-8, 2006.

T. Lillesand, R. Kieffer, Remote sensing and image interpretation, 3 edition, John Wiley & So., Inc. 1994

Q. Jackson, D. A. Landgebe, “Adaptative Bayesian contextual classification based on Markov random fields,” IEEE Trans. Geosci. Remote Sensing, Vol. 40, No. 11, pp. 2454-2463, November 2002.

J. A. Palmason, J. A. Benediktsson, J.R. Sveinsson, and J. Chanussot, “Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis,” in Proc. IGARSS, vol. 1, pp. 176-179, July 2005.

M. Fauvel, J.A. Benediktsson, J. Chanussot and J.R. Sveinsson, “Spectral and spatial classification of hyperspectral data using SVMs and morphological profile,” IEEE Trans. Geosci. Remote Sensing, Vol. 46, No. 11, pp. 3804-3814, 2008.

D. Tuia, F. Pacifici, M. Kanevski and W. J. Emery, “Classification of very high spatial resolution imagery using mathematical morphologie and support vector machine,” IEEE Trans. Geosci. Remote Sensing, Vol. 47, No. 11, pp. 3866-3879, November 2009.

M. Fauvel, J. Chanussot and J.A. Benediktsson, “Decision fusion for the classification of urban remote sensing images,” IEEE Trans. Geosci. Remote Sensing, Vol. 44, No. 10, pp. 2828-2838, October 2006.

J.A. Benediktsson, I. Kanellopoulos, “Classification of multisource and hyperspectral data based on decision fusion,” IEEE Trans. Geosci. Remote Sensing, Vol. 37, pp. 1367-1377, May 1999.

F. Melgani, L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines,” IEEE Trans. Geosci. Remote Sensing, Vol. 42, No. 8, pp. 1778-1790, August 2004.

G. M. Foody, A. Mathur, “A relative evaluation of multiclass image classification by support vector machines,” IEEE Trans. Geosci. Remote Sensing, Vol. 42, No. 6, pp. 1335-1343, Juin 2004.

Y. Tarabalka, J.A. Benediktsson and J. Chanussot, “Spectral-spatial classification of hyperspectral imagery based on partitionial clustering techniques,” IEEE Trans. Geosci. Remote Sensing, Vol. 47, No. 8, pp. 2973-2987, August 2009.

Y. Cheng, “Mean Shift, Mode Seeking, and Clustering,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 8, 1995.

D. Comaniciu, P. Meer, “Mean Shift: A Robust Approach Toward Feature Space Analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, 2002.

C. Simler, C. Beumier, “Building and Road Extraction on Urban VHR Images using SVM Combinations and Mean Shift Segmentation,” Proceeding of the International Conference on Computer Vision Theory and Applications (VISAPP 2010), Anger, France, 2010.

Y. Tarabalka, et al., “Segmentation and classification of hyperspectral images using watershed transformation,” Pattern Recognition (2010), doi:10.1016/j.patcog.2010.01.016