BUILDING AND ROAD EXTRACTION ON URBAN VHR IMAGES USING SVM COMBINATIONS AND MEAN SHIFT SEGMENTATION

Christophe Simler and Charles Beumier
Signal & Image Centre, Royal Military Academy, 30 Av. de la Renaissance, 1000 Brussels, Belgium

Keywords: Support vector machine, Mean shift, Classifier combination, Very high spatial resolution image.

Abstract: A method is proposed for building and road detection on very high spatial resolution multispectral aerial image of dense urban areas. First, objects are extracted with a segmentation algorithm in order to use both spectral and spatial information. Second, a spectral-spatial object-level pattern is formed, and then classification is performed using a 3-class SVM classifier, followed by a post-processing using contextual information to handle conflicts. However, in the particular case where many building roofs are grey like the roads and have similar geometry, classification accuracy is inevitably limited. In order to overcome this limitation, different classifiers are combined and different patterns used, improving the accuracy of 10%.

1 INTRODUCTION

The accurate classification of remote sensing images is an important task for applications such as development planning, emergency response or earth survey. Many investigations are currently done in order to provide both efficient and (semi-)automated classification algorithms. Our study deals with building and road extraction on very high spatial resolution (VHR) aerial images of dense urban areas. The sensor is multispectral and covers a range of three optic spectral channels (RGB), having a spatial resolution of 0.5m per pixel.

Most of the remote sensing classification applications work at the pixel level, and use only spectral information. The first step consists generally in extracting pixel spectral features, then patterns are classified usually with the classical Gaussian maximum likelihood (ML) supervised classifier (Bishop, 2006). However, when only spectral information is used classified data often manifest a salt-and-pepper appearance (Lilesand and Kieffer, 1994). In addition, VHR images of urban areas 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.

Approaches involving Markov random fields (MRF) use the contextual information (Jackson, 2002).

A faster and more recent technique intensively used in hyperspectral imaging consists in building morphological profiles (MPs) from the original data to obtain (local) spatial information about size and shape (Palmason et al., 2005), (Fauvel, 2008), (Tuia, 2009). Once geometrical features are extracted (the MPs for example) they can be concatenated with the spectral pattern to form a composite pattern, which is then classified (Fauvel, 2008). Another solution is to classify separately the spectral and spatial patterns with two (or more) different classifiers and to perform a decision fusion for the final class attribution. The decision fusion processing and interpretation can be performed using fuzzy (Fauvel et al., 2006) or probabilistic (Benediktsson, 1999), (Benediktsson et al., 2007) framework. The difficulty with these approaches is to find adequate source weights reflecting source reliabilities. Fusion can also be performed on the final decision of each classifier. In this case, conflicts between classes are handled by another classifier (Benediktsson, 1999), randomly or with additional information. Overviews of multiple classifier system (MCS) are presented in (Bishop, 2006), (Benediktsson et al., 2007). When geometrical features are considered (the MPs for example), the class distributions can generally not be assumed to be Gaussian and nonparametric supervised classifiers such as decision trees, K-neirest-neighbors, neural networks (Fauvel et al., 2006) (Benediktsson, 1999) or Support Vector Machines (SVMs) (Fauvel, 2008) (Tuia, 2009) are

Simler C. and Beumier C. (2010).
BUILDING AND ROAD EXTRACTION ON URBAN VHR IMAGES USING SVM COMBINATIONS AND MEAN SHIFT SEGMENTATION.
In *Proceedings of the International Conference on Computer Vision Theory and Applications*, pages 451-457
DOI: 10.5220/0002851104510457

generally used. These different classifiers are presented in (Bishop, 2006). In the context of multispectral images, SVMs are generally more effective in terms of classification accuracy than most of the other methods (Melgani, 2004), (Foody, 2004). Also, the "geometric" nature of the SVMs enables them to handle small ratio between the number of available training samples and the number of features. Thus, even in the hyperspectral context they generally do not need a feature reduction pre-processing step (Melgani, 2004). However, SVMs have the drawback to be originally developed to solve binary classification problems, and multi-class SVMs are generally handled by the "one-against-all" or the "one-against-one" strategy (Bishop, 2006), (Melgani, 2004).

Another approach to exploit spatial information is suggested in (Tarabalka et al., 2009). First, a classical pixel wise spectral classification is performed, and a segmentation algorithm is applied independently. Then, spatial information is included by merging the segmentation and the classification maps by assigning to a segmented area the predominant pixel class within it.

In this paper another strategy is suggested, exploiting the fact that in VHR images our two classes of interest (road and building) are objects with specific geometry. The idea consists in building object (global) geometric features. First, extraction of interest objects is performed with a segmentation algorithm applied to the image. Second, a composite object pattern is formed with geometrical and spectral features, then this pattern is classified into class "road", building" or "other". The method provides good classification accuracy with most rural, peri-urban and urban areas. However, in this paper we focus on the difficult case of dense urban areas containing many building roofs of similar spectral signature (and geometry) than the roads. Whatever the classifier used, this leads to a systematic problem of false alarms for the class "road" and of bad detections for the class "building" (many buildings are classified as roads). In this case the building heights would have been useful to discriminate the classes, but with a single image we have no access to this information. In order to compensate this lack of information about the features, we suggest combining several classifiers in a way to exploit simultaneously the ability of all of them to recognize buildings. The paper is organized as follows. The classification technique with a single classifier is presented in part 2. Part 3 deals with the suggested classifier combinations. Finally, part 4 is the conclusion.

2 CLASSIFICATION TECHNIQUE

2.1 Data Specifications

Figure 1 shows a part of multispectral VHR aerial image of dense urban area (2833x2618 pixels). In can be seen that with this example the spectral characteristics of many buildings are similar to the one of the roads. Also, many roads are very narrow, often because partially occluded by houses.

Figure 1: Part of a color aerial image (Brussels center, Belgium) with a spatial resolution of 0.5m.

2.2 Object Extraction with Segmentation

The aim of this part is to extract coherent regions corresponding to actual image objects such as roads, buildings or others. The literature provides many segmentation algorithms divided into three categories: edge-based, region-based and clustering. Some of them are Graph-Cut, region growing, watersheds (Debeir et al., 09), K-means (Bishop, 2006), EM (Bishop, 2006) and mean shift (Comaniciu and Meer, 02). The mean shift algorithm was chosen for the following reasons:

- it was designed for vector processing and thus is adapted to process multispectral images,
- the integration of the spatial coherence property is straightforward,
- it is robust with respect to spectral noise because based on a smoothing process,
- no assumption has to be done about the feature space (number of clusters, underlying distribution).
- There is only one parameter to tune: the segmentation resolution.

The limitation is that if the dimension is high it can suffer from the curse of dimensionality. In our application, a 5-dimensional spectral-spatial feature space is built. In this space, a pixel corresponds to a vector whose components are its three spectral values, and its two spatial coordinates. The mean shift operates on this space by estimating in an iterative way the local maxima of the underlying nonparametric spectral distribution. At each iteration, the components of each vector are replaced by the means of the components of all the vectors situated in a spectral-spatial neighborhood (we use a flat kernel). The convergence toward the local maxima is ensured (Comaniciu and Meer, 02). Then, the pixels having converged toward the same maximum are grouped together to form an object. After observation of the image of figure 1, it has been noticed that our actual objects of interest have always an area upper than one hundred pixels. Thus, an additional step merges (with the other objects) all the mean shift objects smaller than this threshold. The parameter of the mean shift is (in case of a flat kernel) the radius of the spectral-spatial neighborhood. Because this radius is different in spectral than in spatial (the neighborhood is a hyper-ellipsoid), there are in fact two regularization parameters. However, results are not very sensitive with respect to the spatial radius (Comaniciu and Meer, 02), and it can be fixed to an a priori suitable value. In our application it is fixed to seven pixels. The spectral radius is manually tuned in order to find a good compromise between under and over segmentation. With the image of figure 1, a value of 20 is visually optimal. It can be seen in the figure 2 that roads and buildings are generally precisely extracted. Also, there are very few under and over segmentation. It is an advantage with respect to the watershed algorithm, which generally suffer from important over segmentation (Debeir et al., 2009).

Figure 2: Mean shift segmentation results on a zoom of the image of figure 1.

2.3 Spectral-spatial Object Pattern

The aim of this part is to establish an object pattern able to separate the “building”, “road” and “other” classes. It exists many geometrical, spectral or textural features to characterize an area. In our application, the features were selected by observing the mean shift areas in figures such as the figure 2. An idea to discriminate our classes was to use the specific polygonal geometries assumed for roads and buildings. However, the boundaries of mean shift areas are often too jagged to fit suitably polygonal models. The area (size) and eccentricity (shape) descriptors are retained because of their abilities to discriminate roads. The eccentricity (the ratio of the lengths of the two main inertia axis of the area) is estimated by computing the ratio of the two eigenvalues of the (spatial) pixel vector covariance matrix. Textural features are not retained, because the mean shift objects are generally low textured, especially in urban area. Some man-made objects have sometimes some kinds of texture, but it is exception (chimneys on a roof) or perturbations (cars on a road). In addition, the image is filtered before feature computation in order to limit the noise effect and other small perturbations or occlusions, decreasing also the texture. The spectral features retained are the means of the multispectral (RGB) vector computed on the area. It discriminates the classes “road” and “other”. The buildings are generally grey or brown-red. Also a building is often divided into two parts, the sun part and the shadow part. Another color space is tested, the (L*, a*, b*), because it corresponds better to human visual perception. In summary, we suggest testing two spectral-spatial mean shift object patterns: {area, eccentricity, mean of the RGB vector}, and {area, eccentricity, mean of the L*a*b* vector}. Each component is normalized to work in Euclidean space. The main limitation is that some buildings have both similar geometry and color than roads. It exists many automatic feature reduction techniques intensively used in hyperspectral imagery before applying the classification (Fauvel, 2008) (Tuia, 2009) (Melgani, 2004). They are not considered here for the following reasons: feature selection has been done above by observing mean shift areas, five dimensions is a low dimensional problem and redundancy is low with these features, feature reduction is seldom justified with the SVM classifier used in part 2.4 (Melgani, 2004).

2.4 Object Classification 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 (Bishop, 2006), (Fauvel, 2008), (Melgani, 2004), (Foody, 2004). The SVM algorithm is a 2-class classifier. We consider the general case of a training set of two overlapping classes. First, a nonlinear kernel function is applied on the input space in order to obtain a higher dimensional feature space having a better class separability. The Gaussian kernel provides often the best results, and is used in this paper. Second, the parameters of the hyperplan linear model are estimated according to the maximal margin criterion and by penalizing the classification errors. The SVM algorithm with Gaussian kernel has two regularization parameters: the misclassification penalty term and the Gaussian width. In this paper, these parameters are optimized 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 training set. In our application we have three classes ("road", "building" and "other"), and the "one-against-all" multiclass SVM strategy is used. It consists in using three binary SVM classifiers independently, one for each class. During the learning of one class, the elements of the training set of the considered class are opposed to the elements of the two other classes. This technique can provide unbalanced training sets. However, in our application this phenomenon is limited because we have only three classes, and the training set is composed of four hundred buildings, four hundred roads and two hundred others. This training set was built by manually assigning to a class some mean shift areas situated outside the classification part (outside the image of figure 1). It can be noted that such a training set is designed only to classify parts of the considered aerial image, and not parts of other images with different illuminations. In fact, in our application for each aerial image a training set is built on parts of it, and the other parts are classified.

With the "one-against-all" approach, the final decision can be taken by applying the "winner-take-all" to the binary classifier probabilities (Melgani, 2004). Another possibility is to consider the final binary classifier decisions (binary word) (Bishop, 2006). In that case, for three classes there are eight possibilities and the five conflict situations (multiple assignments) are generally handled by choosing randomly one of the classes. In our application, it is possible to handle conflicts by using a priori knowledge and contextual information. In dense urban area, the classes "building" and "road" are largely predominant and have priority in case of conflicts with the class "other". Also, it has been noticed by visualizing the "building" and "road" conflict areas 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). It would be interesting in further work to compare this approach with the one using the binary classifier probabilities (Melgani, 2004). Also, combining contextual information with probabilities would certainly be optimal.

2.5 Classification Accuracy

Figure 3 shows the SVM classification results for the image of figure 1, with the pattern {area, eccentricity, mean of the RGB vector}. On the top: superimposition of the binary SVM results. Detected roads, buildings, and others are respectively drawn in yellow, green and black. The "building" and "road" conflict areas are shows in red. It can be noticed that there are few conflicts. On the middle: 3-class SVM results after handling the previous conflicts with contextual information. Conflicts were generally well handled. Bottom: 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 road or building detection on these areas are considered as exact. Computing the 3x3 confusion matrix (in terms of pixels) between the classified image (an example is in the middle of figure 3) and the ground truth assesses classification accuracy. Some descriptive measures computed from the confusion matrix are in table 1.

Table 1: Classification accuracy measures.

	3-class SVM, pattern {a,e,R,G,B}
Overall accuracy	0.60
Producer's accuracy road	0.66
Producer's accuracy building	0.58
Producer's accuracy other	0.59
User's accuracy road	0.47
User's accuracy building	0.86
User's accuracy other	0.35

As expected with a dense urban area containing similar spectral and spatial "road" and "building" objects, results are poor. It can be seen in table 1 that the false alarm rate is high for both the classes "road" and "other", in the detriment of the class "building" having a high bad detection rate. In fact, many grey building roofs are classified roads. This is because the class "road" contains only grey elements (pure class) and the class "building" contains both grey and brown-red elements. Some other classifiers were tested and without surprise the same problem occurs, with even worse results than the SVM classifier. In order to overcome this problem, a finer class definition could be used. For example the class "building" can be divided into two sub-classes, one for the grey buildings and the other for the brown-red ones. Another idea consists in increasing the number of features. However, it is not ensured at all that much better classification accuracy will be obtained. In this paper, another strategy is suggested. It consists in combining the decisions of different classifiers and is the topic of part 3.

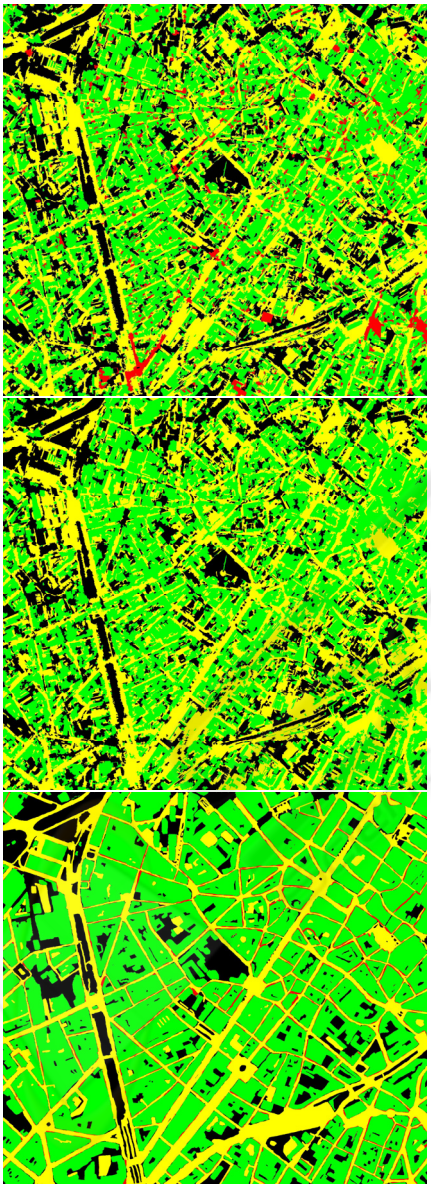

Figure 3: Top: superimposition of the binary SVM results, with the pattern {area, ecc., meanRGB}. Yellow: roads; green: buildings; black: other; red: building and road conflicts. Middle: 3-class SVM results. Bot.: ground truth.

3 CLASSIFIER COMBINAISONS

Our aim is to solve the false alarm problem for the classes "road" and "other", and the bad detection problem for the class "building". The solution consists in covering some road and other areas by building areas. However, if some roads are automatically eliminated, some correct roads will also be lost and it is not conceivable because they are already too divided up and cut. On the contrary, some other areas can be eliminated with fewer risks. In a dense urban area they are not numerous, and in addition the class "other" is by far the less important. Several classifiers were tested and the bad detection problem for the class "building" is recurrent to all of them. However, by observing the visual results it was noticed that they are complementary in the sense that each of them detects correctly a little part of grey buildings, which are classified roads by the others. In fact, a union of all the detected building will certainly leads to a much lower bad detection rate, without significant increase of the false alarms. We suggest adding the building areas detected by another classifier to the building areas of the SVM of part 2, while preserving its road network. This second classifier is a single class SVM (Tax and Duin, 2001). The difference between a single class and a two class SVM can be illustrated as follows: while the two class SVM attempts to separate two classes with a linear hyperplan, the single class SVM attempts to encircle the target class with a hypersphere to isolate it from the rest. Also, only target samples are needed for the training. In this paper, it uses the same training set and has the same parameters than the classical 2-class SVM. Of course the cross-validation scheme is slightly different and the cost function minimized in the one suggested in (Tax and Duin, 2001), with equal weighting. The results can be seen in table 2.

Table 2: Classification accuracy measures.

	3-class SVM + buildings of the single class SVM while roads preserving, pattern {a,e,R,G,B}
Overall accuracy	0.66
Producer's accuracy road	0.66
Producer's accuracy building	0.69
Producer's accuracy other	0.48
User's accuracy road	0.47
User's accuracy building	0.81
User's accuracy other	0.53

Table 2 shows that such a combination of two classifiers improves of 6% the overall accuracy. As expected, the building bad detection rate is much lower, with just a small increase of the building false alarm rate. In summary, performances are better concerning the buildings, but not excellent. Also, the problems with the roads remain. Up to now, only one of our two patterns has been considered, but the same experiments were done with the pattern using the (L*, a*, b*) color space. There is one significant improvement with respect to the color space (R, G, B): the producer's accuracy road increases of 11%, without adding many more false alarms. An idea to decrease both the road false alarm rate and the building bad detection rate is to combine this last classifier combination with the buildings of the previous classifier combination. In this last fusion, all the buildings are added while no road preservation is done. The risk is thus to loose some correct roads. However, despite this drawback, the classification accuracy is significantly better than all the results obtained previously, as shown in table 3. This four-classifier fusion improves the overall accuracy of 9% with respect to the SVM classifier used individually. The road false alarm rate is lower but remains high. Encouraging results concern mainly the "building" class with a much lower bad detection rate. Visual results are shown on figure 4.

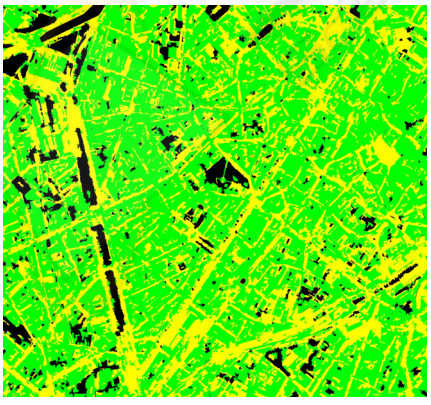

Figure 4: Classification results of the four-classifier combination of table 3 (ground truth on figure 3).

Table 3: Classification accuracy measures.

	(3-class SVM + buildings of the single class SVM while roads preserving, pattern {a,e,L*,a*,b*}) + buildings of the combination of table 2
Overall accuracy	0.69
Producer's accuracy road	0.67
Producer's accuracy building	0.76
Producer's accuracy other	0.42
User's accuracy road	0.49
User's accuracy building	0.80
User's accuracy other	0.75

4 CONCLUSIONS

Building and road detection on VHR 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. Also, contextual information around objects is used to solve the SVM conflicts between roads and buildings. In order to overcome the high road false alarm rate and the high building bad detection rate in the presence of similar road and building objects, some classifier combinations were suggested and different features used. Significant improvements are achieved in terms of accuracy. Our current research aims at filling road gaps and smoothing road borders, on the basis of straight segment detection.

ACKNOWLEDGEMENTS

We thank the financer IRSIB and the data provider IGN (Be).

REFERENCES

Bishop, C., 2006. Pattern recognition and machine learning, Springer.

Lilesand, T., Kieffer, R., 1994. Remote sensing and image interpretation, Third edition, John Wiley & Sons, Inc.

Jackson, Q., 2002. Adaptative Bayesian contextual classification based on Markov random fields, IEEE TGRS, Vol. 40, No. 11, pp. 2454-2463.

Palmason, J. and all, 2005. Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis. Proc. IGARSS, vol. 1, pp. 176-179.

Fauvel, M., 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profile. IEEE TGRS, Vol. 46, No. 11, pp. 3804-3814.

Tuia, D., 2009. Classification of very high spatial resolution imagery using mathematical morphologie and support vector machine. IEEE TGRS, Vol. 47, No. 11, pp. 3866-3879.

Fauvel, M and all, 2006. Decision fusion for the classification of urban remote sensing images. IEEE TGRS, Vol. 44, No. 10, pp. 2828-2838.

Benediktsson, J., 1999. Classification of multisourse and hyperspectral data based on decision fusion. IEEE TGRS, Vol. 37, pp. 1367-1377.

Benediktsson and all, 2007. Multiple classifier systems in remote sensing: from basics to recent developments. MCS, 7th International Workshop, Praque, Tchèque.

Melgani, F., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE TGRS, Vol. 42, No. 8, pp. 1778-1790.

Foody, G., 2004. A relative evaluation of multiclass image classification by support vector machines. IEEE TGRS, Vol. 42, No. 6, pp. 1335-1343.

Tarabalka, Y., and all, 2009. Spectral-spatial classification of hyperspectral Imagery based on partitional clustering techniques. IEEE TGRS, Vol. 47, No. 8, pp. 2973-2987.

Debeir, O., Atoui H., Simler 2009. Weakened Watershed Assembly for Remote Sensing Image Segmentation and Change Detection. VISAPP, Portugal.

Comaniciu, D., Meer, P., 2002. Mean Shift: A Robust Approach Toward Feature Space Analysis. IEEE PAMI, Vol. 24, No. 5.

Tax, D., Duin, R., 2001. Uniform object generation for optimizing one-class classifiers. JMLR 2, pp. 155-173.