COMPARATIVE ANALYSIS OF AUTOMATIC APPROACHES TO BUILDING DETECTION FROM MULTI-SOURCE AERIAL DATA

E. Frontoni [a], K. Khoshelham [b], C. Nardinocchi [b, c], S. Nedkov [b], P. Zingaretti [a, *]

[a] DIIGA, Polytechnic University of Marche, 60131 Ancona, Italy, - (frontoni, zinga)@diiga.univpm.it
[b] Optical and Laser Remote Sensing, Delft University of Technology, 2629 HS Delft, The Netherlands - (k.khoshelham, C.Nardinocchi)@tudelft.nl
[c] DITS, University of Rome "La Sapienza", 00184 Roma, Italy, - carla.nardinocchi@uniroma1.it

KEY WORDS: Building detection, comparative analysis, pixel-based and object-based algorithms, LIDAR, multispectral images

ABSTRACT:

Automatic building detection has been a hot topic since the early 1990's. Early approaches were based on a single aerial image. Detecting buildings is a difficult task so it can be more effective when multiple sources of information are obtained and fused. The objective of this paper is to provide a comparative analysis of automatic approaches to building detection from multi-source aerial images. We analysed data related to both urban and suburban areas and took into consideration both object-based and pixel-based methods. Although many of these methods perform full data classification, we focused only on the detection of building regions. Three measures were used for the evaluation of the performance of each method: number of detected buildings to their total number (detection rate), number of objects wrongly detected as buildings (false positive) and number of missed buildings (false negative) to the number of detected buildings. The data sets we used were RGB and colour infrared (CIR) orthoimages and Digital Surface Models (DSMs) obtained by an airborne laser scanner, which provides a first pulse DSM and a last pulse DSM. In addition, we derived from these data and used other four sources of information: a Digital Terrain Model (DTM) obtained from a filtered version of the last pulse DSM, the height difference between the last pulse and the DTM, the height difference between the first and the last pulse and the Normalized Difference Vegetation Index (NVDI) derived from the red and infrared channels.We analysed results coming from three classification algorithms, namely Bayesian, Dempster-Shafer and AdaBoost, applied to the features extracted both at pixel level and at object level. To obtain a very realistic comparison we used the same training set for all methods, either pixel-based or object-based. Results obtained are interesting and can be synthesised in the need of fusing (the results of) more approaches to yield the best results.

1. INTRODUCTION

Large-scale cadastral maps that contain building boundaries are an important source of information for governments. These maps are mainly used for valuing and taxing properties and creating databases of land ownership. Because of the rapid changes of urban areas, an updating of the cadastral maps must be carried out on a regular basis (i.e., every 5~10 years) to keep the databases up to date. Map updating is traditionally performed manually by an operator who is responsible for the detection of changed buildings by comparing the map with a recent aerial image (or stereo pair). For large cities, this process is very time-consuming and costly. In most cases, a large proportion of buildings, about 95%, remains unchanged, while only a small number of them needs to be updated. Nevertheless, the operator has to inspect the entire scene carefully in order to locate those few buildings that have changed. Automated approaches to building detection are of great importance in map updating, because they can reduce the amount of manual work, and consequently lead to a reduction of time and cost of the map updating process.

Early approaches to automated building detection relied mostly on a single source of data. Huertas et al., (1993) and Nevatia et al., (1997) developed methods for automated building detection in monocular aerial images based on shadows as evidence. The methods of Fischer et al., (1998) and Fradkin et al., (2001) were based on the processing of multiple-overlap aerial images. Weidner and Forstner (1995), and Vosselman (1999) used height data in the form of a digital surface model for building detection. Today we can register approaches based on multiple sources of information, and not only from aerial images but also, for example, using already available urban maps (with the aim to update them) or LADAR data acquired by unmanned ground vehicles. Recently, with the availability of airborne laser data and imagery in multiple spectral bands, the application of data fusion methods to building detection has attracted more attention. Khoshelham et al., (2005) developed a method to fit planar surfaces to height data within regions of a segmented aerial image for the detection of building roofs. Walter (2004) applied a Bayesian maximum likelihood method to object-based classification of multi-spectral aerial data. Bartels and Wei (2006) performed pixel-based classification of aerial imagery and laser range data using the Bayesian maximum likelihood approach. Rottensteiner et al., (2004) and Lu et al., (2006) developed methods to extract buildings from aerial imagery and laser range data based on Dempster-Shafer evidence theory. Zingaretti et al., (2007) adopted an AdaBoost algorithm for the automatic identification of rules for the classification of raw LIDAR data mainly as buildings, ground and vegetation.

* Corresponding author.

While relatively successful applications of the fusion methods to the problem of automated building detection in multi-source aerial data have been reported, a comparison of the performance of these methods is not available. The objective of this paper is to provide a comparative evaluation of three common data fusion and classification methods, namely Bayesian, Dempster-Shafer and AdaBoost, as applied to the detection of buildings in multi-source aerial data. We present results of both pixel-based and object-based implementations of the methods, and compare the performance of the methods on the basis of ground truth information obtained by manual extraction of buildings.

The paper proceeds with a brief overview of the Bayesian decision theory, the Dempster-Shafer evidence theory, and the Adaboost classification algorithm in Section 2. Section 3 describes the experimental setup, including a description of the data and the extraction of pixel-based and object-based features. In Section 4, the results of the experimental evaluation of the methods are presented, and a discussion on the various factors affecting the performance of the methods is provided. The paper concludes in Section 5.

2. AN OVERVIEW OF THE METHODS

In a typical data fusion and classification method, first a set of features are extracted from the data, and a number of class hypotheses are defined. In the next step, a decision is made for each feature as to what class of objects it belongs to. The principle of decision making varies across different classification methods. In the following, a brief description of the decision-making principle in three classification methods, Bayesian, Dempster-Shafer and Adaboost method is presented.

2.1 Bayesian method

In the Bayesian method, a decision is made based on maximizing the likelihood that a feature vector x belongs to a class w_j. Formally, this can be expressed as (Duda et al., 2001):

$$d_j(x) = p(x/w_j).P(w_j) \tag{1}$$

where $p(x/w_j)$ is the conditional probability of x in the probability distribution function of class w_j, $P(w_j)$ is the prior probability of class w_j, and $d_j(x)$ is a decision function that is evaluated for each feature x and class w_j, and is to be maximized in order to make a decision. Often, it can be assumed that the classes have Gaussian probability distribution functions. In this case, the *maximum likelihood* decision function can be expressed as:

$$d_j(x) = -\frac{1}{2}(x-\mu_j)^T \sum\nolimits_j^{-1}(x-\mu_j) - \frac{1}{2}\log\left|\sum\nolimits_j\right| + \log P(w_j) \tag{2}$$

where parameters μ_j and Σ_j are respectively the mean and covariance matrix of the multi-dimensional Gaussian probability distribution function of the class w_j.

A simplification of the maximum likelihood method can be achieved if an assumption can be made that the features in all classes are independent and have the same variance. Further, if it can be assumed that the prior probabilities of all classes are the same, the decision function in Eq. (2) will reduce to:

$$d_j(x) = -\frac{1}{2}(x-\mu_j)^T(x-\mu_j) \tag{3}$$

A classifier based on the decision function given in Eq. 3 is referred to as a *minimum distance* classifier. The principle of the minimum distance classification is that a decision on the class of a feature can be made by minimizing the distance of the feature to the means of the hypothesized classes.

2.2 Dempster-Shafer method

The Dempster-Shafer method performs a classification of data into different classes on the basis of the evidence that each feature provides for each class hypotheses (Gordon and Shortliffe, 1990). Hypotheses include not only all classes but also any union of the classes. When all the available evidences for the class hypotheses are gathered from different features, they are combined using a combination rule, and the sum of the combined evidences assigned to all subsets of a class hypothesis defines the amount of belief in that hypothesis:

$$Bel(A) = \sum_{B\subseteq A} m(B) \tag{4}$$

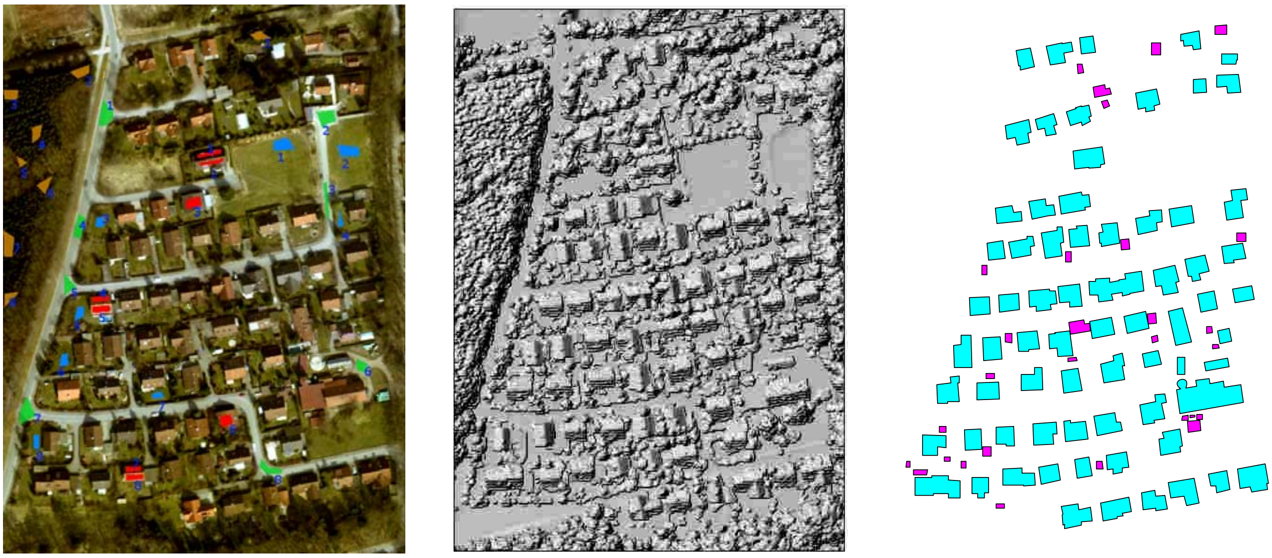
The decision on the class of a feature is made based on a maximum belief decision rule, which assigns a feature to a class A if the total amount of belief supporting A is larger than that supporting its negation:

$$Bel(A) \geq Bel(\overline{A}) \tag{5}$$

Khoshelham et al., (2008) provide a detailed description of evidence gathering, combination, and belief computation using features extracted from aerial imagery and laser range data.

2.3 Adaboost algorithm

Boosting (Sutton, 2005) is a method of combining classifiers, which are iteratively created from weighted versions of the learning sample, with the weights adaptively adjusted at each step to give increased weight to the cases which were misclassified on the previous step. The final predictions are obtained by weighting the results of the iteratively produced predictors. Boosting was originally developed for classification, and is typically applied for creating an accurate strong classifier by combining a set of weak classifiers. A weak classifier is only required to be better than chance, and thus can be very simple and computationally inexpensive. However, combining many of these simple and inexpensive classifiers results in a strong classifier, which often outperforms most “monolithic” strong classifiers such as Support Vector Machines and Neural Networks. In 1990, Schapire (1990) developed the predecessor to later boosting algorithms developed by him and others. AdaBoost (a short for “adaptive boosting”) is now the most popular boosting algorithm (Freund, 1997). Boosting uses a weighted average of results obtained from applying a prediction method to various samples. Also, with boosting, the samples used at each step are not all drawn in the same way from the same population, but rather the incorrectly predicted cases from a given step are given increased weight during the next step. Thus, boosting is an iterative procedure, incorporating weights, as opposed to being based on a simple averaging of predictions, as is the case with bagging (Sutton, 2005).



(a) (b) (c)

Figure 1. Dataset used in the experiments (a) RGB orthoimage with superimposed the 8 samples for each class, constituting the training set; (b) First pulse airborne laser range image; (c) Reference building map manually extracted from the image and laser data.

3. EXPERIMENTAL SETUP

3.1 Description of the data

The study area of the experiments is a small suburban neighbourhood of about 1.2 km^2 in the city of Memmingen, south of Germany. About seventy buildings with dimensions ranging from around 100 to 300 m^2 and with vegetation between them are comprised in the area. Moreover, a large number of garages and garden sheds are present in the vicinity of buildings.

The multi-source data available for the experiments include an aerial orthorectified image in four spectral channels, red, green, blue, and near infrared, and laser range data in both first and last pulse recorded by an airborne laser scanner. Figure 1a depicts the RGB image, while Figure 1b shows the DSM (Digital Surface Model) corresponding to the first pulse laser range data. In addition, a digital elevation model (DEM) of the terrain as a filtered version of the last pulse laser scanner data is available. Radiometric data have a resolution of 0.5 meters, while the laser data are provided at a density of 1 point/m^2.

The whole dataset was used in both qualitative and quantitative comparisons. The comparisons were carried out on the basis of a reference dataset generated by manual extraction of the buildings in the image and the laser data. No in-situ information for the recognition of the buildings was available. All points in the reference dataset were labelled as either Building (B) or Not-Building (NB). Figure 1c shows the reference building data in blue, while garages and garden sheds are depicted in pink. In particular, only those with a dimension of 15~30 m^2 and a height of at least 2.5 meters were included in the reference data. In spite of focusing only on buildings at an early stage, a classification of the data in the following four classes was first performed by all the methods: building, tree, bare land and grass. Building regions were then detected from the classification results.

To allow a very realistic comparison a strong assumption was to use the same training set for all methods, either pixel-based or object-based. In particular, we selected eight sets of pixels, totalling more or less an equal number of samples (from 2 to 3 thousand pixels), for each class. Consequently, in the case of pixel-based approaches the training set represents about 1.8% of the total pixels, while in the object based approach the regions corresponding to that pixels represent about the 2.1% of the total regions of the image.

3.2 Pixel based features

In the pixel-based classification data fusion was carried out at a pixel level. Each pixel of the image is visited once, its features extracted and then passed on to the classification methods. All methods work with the following three features: Δh, the height difference between the last echo and the DTM; Δp, the height difference between first and last echoes; NVDI, the Normalized Difference Vegetation Index obtained from the red and near-infrared channels. Since the Adaboost algorithm was expected to perform better with a larger number of features, it was tested with additional features from all the channels of the radiometric data. This allowed the algorithm to be tested with five (Δh, Δp, NVDI, G,B) and seven features (Δh, Δp, NVDI, R, G,B, NIR).

3.3 Object based features

To perform object-based classification with features at a region level, a preliminary segmentation process was applied to the image data. For each region in the segmented image the average, minimum, maximum and root mean square value was calculated for the first three features described in the pixel-based classification. In addition, the number of points belonging to each region, the average, minimum, maximum and root mean square values of the multispectral intensities and of the first and last pulse, the kurtosis (relative peakedness or flatness of a distribution compared to the normal distribution) and skewness (the degree of asymmetry of a distribution around its mean) were considered. As done with pixel based methods, first a comparison was carried out using only the average value of Δh, Δp and *NVDI*. Later, all the additional features were included.

As known, overgrown and undergrown regions are inevitable in the segmented image. The classification methods can cope with undergrown regions by assigning them to a same class; however, in overgrown regions features of two or more different objects are present, and their merger would certainly influence the classification results. For this reason, the segmentation algorithm was applied with parameter settings that produced oversegmented results. To study the influence of the parameter setting, two segmentations were obtained: a slightly oversegmented image with smoothing parameter 10, and a largely oversegmented image with smoothing parameter 5. The object-based classifications were applied to both segmented images.

4. RESULTS AND DISCUSSION

The classification methods were applied to the features extracted both at pixel level and at object level. Then the building class were extracted from the classification results. A cleaning operation based on morphological opening and reconstruction was applied to the detected buildings in order to remove regions that were smaller than a threshold (Khoshelham et al., 2008). The comparisons were carried out both qualitatively, by visual inspection of the results, and quantitatively, by deriving a number of performance measures using the reference data. To this aim we have defined the following quantities: *a*, the number of points correctly classified as buildings; *b*, the number of buildings points classified as other objects (False Negative - FN); *c*, the number of other objects classified as buildings (False Positive - FP); *d*, the number of other objects correctly classified. Then, the detection rate is expressed as a/(a+b), the percentage of FN and FP as b/(a+b) and c/(c+d), respectively.

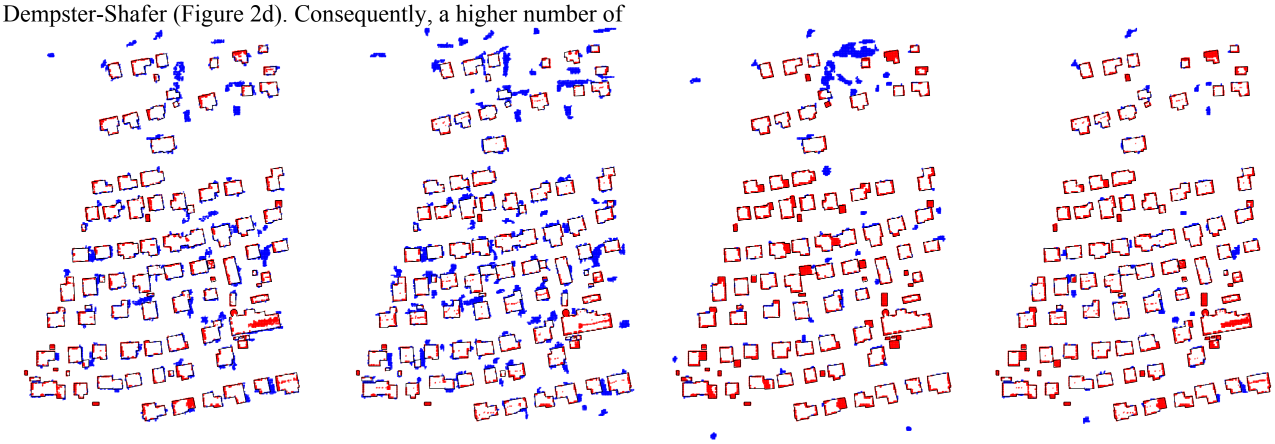
4.1 Qualitative comparison of approaches

The qualitative comparison is performed visually comparing the results by taking error location into particular attention. To this aim, in Figure 2 FP (blue) and FN (red) for the pixel based classification realized with only three features are superimposed to the reference data. The images highlight the different contribution of the two kind of errors very well: a predominance of FP in the first two images, namely those corresponding to the Adaboost (a) and the Bayesian (b), but a predominance of FN in the Minimum Distance (Figure 2c) and Dempster-Shafer (Figure 2d). Consequently, a higher number of missed elements (particularly garages or garden sheds due to their small dimensions) is present in the last two methods. To sum up, the not detected garages or garden sheds were: 8, 9 and 16 for the Adaboost working respectively with 3, 5 and 7 features; 5 and 7 for the Bayesian working respectively with 3 and 5 features; 28 with the Minimum Distance and 19 with the Dempster-Shafer. Moreover both these last two methods do not detect 1 building.

In the same way Figure 3 shows error location of the object based classification based on the segmentation that produces a larger number of regions (threshold 5). The same trend in the distribution of errors can be noticed. With the two different segmentations, threshold 5 and 10, respectively, the Adaboost using only 3 features missed 27 and 12 garages or garden sheds and 1 and 2 buildings. Using 15 features and segmentation threshold 5 it missed 12 garages or garden sheds and 1 building, while with segmentation threshold 10 all buildings were detected, but missing 6 garages or garden sheds. With the two different segmentations, threshold 5 and 10, respectively, only 2 and 3 garages were missed by the Bayesian; 28 garages and 1 building were not detected by the Minimum Distance, and 21, 24 garages and 1, 2 buildings were not detected by the Dempster-Shafer.

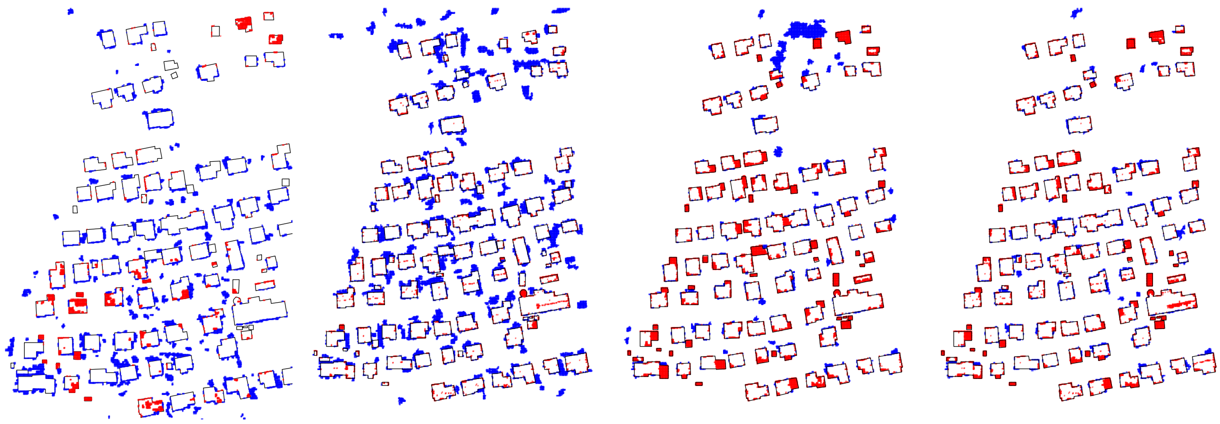
4.2 - Quantitative comparison of approaches

On a quantitative level the detection rate and error rates are furnished for every elaboration.
Classification results expressed as percentage of detection rate or error rates, obtained with the three pixel-based methods using only three features are shown in Table 1 both for building (B) and not-building (NB) objects. Similarly, Table 2 summarizes the detection and error rate percentages for every object-based classification algorithm.
It's important to notice that both Adaboost and Dempster-Shafer furnished a number of not classified (NC) points, which means there are points where the algorithm was not able to decide among the four classes. The percentage of NC building points is shown in a separate row of the table, while Fig. 4 emphasizes how FN and FP are relatively small in the Adaboost algorithm despite of it has the same amount of errors of the Bayesian algorithm due to the contribution of NC points.
Similarly the same results are shown for the object-based classification in Figure 5.



(a) (b) (c) (d)

Figure 2. Results of pixel based classification using three features. False Positive (blue) and False Negative (red) superimposed at the reference map. (a) Adaboost; (b) Bayesian; (c) Minimum Distance; (d) Dempster-Shafer.



(a) (b) (c) (d)

Figure 3. Results of object based classification using three features and the segmentation with threshold 5. False Positive (blue) and False Negative (red) superimposed at the reference map. (a) Adaboost; (b) Bayesian; (c) Minimum Distance; (d) Dempster-Shafer.

		Reference	
		B	NB
Adaboost	B	83.44	2.52
	NB	9.54	85.87
	NC	7.02	11.61
Maximum Likelihood	B	85.89	3.91
	NB	14.11	96.09
	NC	-	-
Minimum Distance	B	72.26	1.49
	NB	27.74	98.51
	NC	-	-
Demsptер Shafer	B	76.49	0.89
	NB	23.02	99.11
	NC	0.31	1.09

Table 1. Pixel-based classification results using three features.

		Reference	
		B	NB
Adaboost	B	80.86	3.53
	NB	6.84	67.46
	NC	12.31	29.01
Maximum Likelihood	B	93.42	5.97
	NB	6.58	94.03
	NC	-	-
Minimum Distance	B	76.09	1.83
	NB	23.91	98.17
	NC	-	
Demsptер Shafer	B	80.93	1.20
	NB	18.83	97.60
	NC	0.24	1.20

Table 2. Object-based classification results using three features.

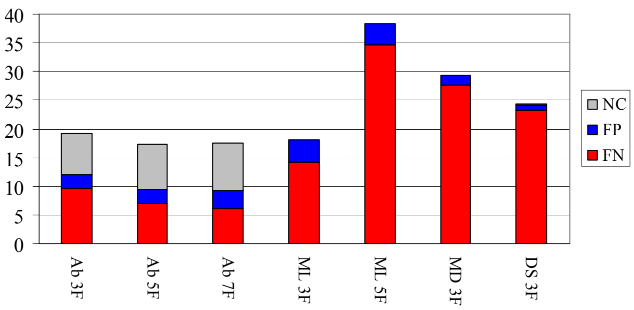


Figure 4. FN, FP and NC for pixel-based classification.

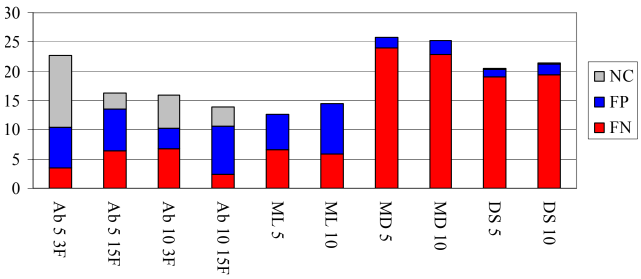


Figure 5. FN, FP and NC for object-based classification

4.3 - Discussion

Generally, object based algorithms perform better than pixel based methods, on average about 5% more. In the case of the Adaboost algorithm this is not evident from Table 2, but using all 15 features and segmentation threshold 10 in the object-based approach Adaboost reached a building detection rate of 94.27% (about the same for ML).

Moreover it was experimented that a larger number of features does not produce a very important improvement in the classification results. On the contrary, often not adequate features can deteriorate the results, as it happened the Maximum Likelihood when working with 5 features. Future works will investigate deeply this field and we are also studying the relevance of the used features looking for redundancies.

Finally, another aspect to further investigate is the influence of segmentation threshold in the object-based classification, in particular in the case of over-segmentation, when many regions become so small to reduce the method very similar to a pixel-based one.

5. CONCLUSIONS

We analysed results coming from three classification algorithms, namely Bayesian, Dempster-Shafer and AdaBoost algorithm, applied to the features extracted both at pixel level and at object level. To obtain a very realistic comparison we used the same training set for all methods, either pixel-based or object-based.
Results obtained are interesting and can be synthesised in the need of fusing (the results of) more approaches to yield the best results.
Some conclusions can be also done on the classification performances: as usually probabilistic classifiers bring to better results due to the fact that they can model noisy data and implicitly filter bad associated results. From this point of view the boosting part of the Adaboost algorithm guarantees optimal performances. The presence of a large number of not classified data is another good quality of this approach; usually in the classification literature this kind of result is considered better than false and positive results and usually not classified classes data are then disambiguated using a second level classifier (different from the first one) mixed with rule based approaches.
Future works will investigate deeply this field and we are also studying the relevance of the used features looking for redundancies. We will finally try to use a cluster based learning approach to train probabilistic classifier using a reduced number of data. This will allow us to compare performance of different complex classifiers (such as SVM, Neural Networks and Particle based approaches) over the huge amount of data used in the proposed experiments.

ACKNOWLEDGEMENTS

The authors would like to thank the Toposys company for providing the dataset used in the experiments.

References

Bartels, M. and Wei, H., 2006. Maximum likelihood classification of LIDAR data incorporating multiple co-registered bands, *4th Int. Workshop on Pattern Recognition in Remote Sensing in conjunction with the 18th Int. Conf. on Pattern Recognition*, Hong Kong, pp. 17-20.

Duda, R.O., Hart, P.E. and Stork, D.G., 2001. *Pattern classification*, second edition. Wiley, New York, pp. 654.

Fischer, A. et al., 1998. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D, *Computer Vision and Image Understanding*, 72(2), pp. 185-203.

Fradkin, M., Maitre, H. and Roux, M., 2001. Building detection from multiple aerial images in dense urban areas. *Computer Vision and Image Understanding*, 82, pp. 181-207.

Freund Y. and Schapire R.E., 1997, A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, 55(1), pp. 119–139.

Gordon, J. and Shortliffe, E.H., 1990. The Dempster-Shafer theory of evidence. In: G. Shafer and J. Pearl (Editors), *Readings in uncertain reasoning*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 768.

Huertas, A., Lin, C. and Nevatia, R., 1993. Detection of buildings from monocular views of aerial scenes using perceptual organization and shadows, *ARPA image understanding workshop*, Washington, DC, pp. 253-260.

Khoshelham, K., Li, Z.L. and King, B., 2005. A split-and-merge technique for automated reconstruction of roof planes. *Photogrammetric Engineering and Remote Sensing*, 71(7), pp. 855-862.

Khoshelham, K., Nedkov, S. and Nardinocchi, C., 2008. A comparison of Bayesian and evidence-based fusion methods for automated building detection in aerial data, *21st ISPRS Congress, Commission VI, WG VI/4*, Beijing China.

Lu, Y.H., Trinder, J.C. and Kubik, K., 2006. Automatic building detection using the Dempster-Shafer algorithm. *Photogrammetric Engineering and Remote Sensing*, 72(4), pp. 395-403.

Nevatia, R., Lin, C. and Huertas, A., 1997. A system for building detection from aerial images. In: A. Gruen, E. Baltsavias and O. Henricsson (Editors), *Automatic extraction of man-made objects from aerial images (II)*, Birkhauser Verlag, Basel, pp. 77-86.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K. and Lovell, B., 2004. Building detection by Dempster-Shafer fusion of LIDAR data and multispectral aerial imagery, *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, Cambridge, United Kingdom.

Schapire, R., 1990. The strength of weak learnability, *Machine Learning*, 5, pp. 197–227.

Sutton, C.D., 2005. *Classification and regression trees, bagging, and boosting, Handbook of Statistics*, vol. 24, cap. 11. Elsevier, Amsterdam.

Vosselman, G., 1999. Building reconstruction using planar faces in very high density height data, *ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery*, Munich, pp. 87-92.

Walter, V., 2004. Object-based classification of remote sensing data for change detection, *ISPRS Journal of Photogrammetry and Remote Sensing*, 58, pp. 225-238.

Weidner, U. and Forstner, W., 1995. Towards automatic building extraction from high resolution digital elevation models, *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4), pp. 38-49.

Zingaretti P., Frontoni E., Forlani G., Nardinocchi C., 2007. Automatic extraction of LIDAR data classification rules, *14th Int. Conf. on Image Analysis and Processing- ICIAP*, IEEE Computer Society, pp. 273-278.