Urban detection using Decision Tree classifier: a case study

Sara Steffenino[1,2], Irene Angeluccetti[1] and Franca Disabato[2,3]

[1]*Politecnico di Torino, DIST, Torino, Italy; sara.steffenino@polito.it, irene.angeluccetti@polito.it*
[2]*ITHACA (Information Technology for Humanitarian Assistance, Cooperation and Action), Torino, Italy; franca.disabato@ithaca.polito.it*
[3]*WFP(World Food Programme), Roma, Italy*

Abstract. This work constitutes a first step towards the definition of a methodology for automatic urban extraction from medium spatial resolution Landsat data. Decision Tree is investigated as classification technique due to its ability in establishing which is the most relevant information to be used for the classification process and its capability of extracting rules that can be further applied to other inputs. The attention was focused on the evaluation of parameters that better define the training set to be used for the learning phase of the classifier since its definition affects all the next steps of the process. Different training sets were created by combining different features, such as different level of radiometric pre-processing applied to the input images, the number of classes considered to train the classifier, the temporal extent of the training set and the use of different attributes (bands or spectral indexes). Different post-processing techniques were also evaluated. Classifiers, obtained by the generated training sets, were evaluated in two different areas of Piedmont Region, where the official regional cartography at scale 1:10000 was used for validation. Accuracies round 81% in the Torino case study and around 96%-97% in Asti case study were reached, thanks to the use of indexes such as NDVI and NDBBBI and the use of post-processing such as majority filtering that allowed enhancing classifier performances.

Keywords. Urban detection, Decision Tree classifier, Landsat ETM+ data, medium resolution

1. Introduction

Rapid mapping activities based on the use of satellite images can be a crucial support for emergency management. Data at suitable spectral resolution, with global coverage, often permit to gather information about inaccessible areas and to have a large archive of historical data available. Information derived from post-event satellite images can be effectively used only when other spatial data (e.g. administrative boundaries, infrastructure details, settlements etc.) are integrated; when this kind of information is missing, its derivation from different data sources is needed, such as archive satellite imagery. The possibility of having algorithms for automatic features extraction could be very helpful in order to avoid time-consuming visual interpretation and related digitalization operations, which are the most common used approach in an operational context to date [1].

Among all necessary information, those related to the extent of urban areas are particularly relevant for supporting crisis management because the knowledge of where buildings and population are located is essential in order to provide information about the extent of damaged areas and population potentially affected: the whole crisis management cycle - including damage assessment, recovery, reconstruction and planning - should benefit from an improved and globally-consistent description of human settlements [2].

Information about built-up areas can be differently extracted according to the desired output scale; if a very small area is investigated and large scale maps are required, information about single buildings is necessary; otherwise, when large areas are investigated and small scale maps are required, necessary information can merely be the extent of the built-up area. The herein presented work mainly focuses on the extraction of this kind of information: therefore, medium spatial resolution Landsat data were used, considering both the scope of applying the methodology at a global extent, and the usefulness of deriving information at such a scale.

In this context, decision tree is investigated as classification technique, due to its ability in establishing which is the most relevant information to be used for the classification process and its capability of extracting rules that can be further applied to other inputs.

In particular this work aims at obtaining preliminary rules for the extraction of urban areas, evaluating the effectiveness of decision tree classifiers and identifying which steps to perform in order to obtain a suitable classification procedure. The main focus is on the level of processing that can be applied to images (different kind of pre and post processing), on the choice of training set features such as the number of classes with whom is created, on the temporal extent of the training set and on the used input attributes.

2. Methods

2.1. Decision Tree classifier: an overview

The main aim of this work is the automation of procedures for features extraction. The chosen approach is related to the extraction of rules that allow identifying the desired land cover type. Rules are based on the definition of the attributes to include in with the choice of decision thresholds; if, on one side, thresholds could be obtained using the knowledge provided by experts, on the other side it is advisable to explore data mining approaches for the identification of suitable bands for classification as well as determining the decision thresholds. In fact, experts may disagree on the decision boundaries, while with the use of data mining techniques, reliable, transferable and reproducible decision thresholds can be obtained [3].

Decision Tree classifier was chosen for this study for the following reasons: first, it’s able to extract rules in an automatic way, that can be later applied to different images; second, the model built by the algorithm is easily interpreted, and can therefore be modified according to the user needs; third, the choice of which attributes to include in the classifier is performed by the algorithm, on the basis of the criteria used by the algorithm itself. Finally, decision tree algorithms don’t require additional information besides those already contained in the training data (e.g. domain knowledge or prior knowledge of distributions on the data or classes) and generally produce good classification accuracies if compared to other techniques.

The tree classifier used to perform the analysis is J48, an open source Java implementation of C4.5 data mining algorithm developed by Quinlan [4], one of the most popular decision tree algorithms due to its well-known efficiency [4].

2.2. Data and case study

The Piedmont region, located in the north-west part of Italy, was chosen as study area: it presents a good temporal coverage of Landsat images and have an official regional cartography (Cartografia Tecnica Regionale, CTRN) at scale 1:10:000 available, freely accessible and suitable for validation purposes. Particularly, two different areas were chosen: Asti province and Torino municipality. In these areas, the temporal coherence between data used for the classification process and update of the regional cartography used for validation was verified.

Landsat ETM+ images, acquired in modality SLC-on (i.e. before 31/05/2003) ,were used for the training phase of the classifier, while Landsat images from GLS2005 collection (that provides one clear image during leaf-on conditions for every location of the globe between 2004 and 2007) were used for the classification phase. Acquisition dates of data used in input and for validation purposes are summarized in Table 1.

Table 1. Data used for the analysis

Case study	Data used for training sets generation (Landsat ETM+) [acquisition date – path/row]	Data used for the classification (GLS2005) [acquisition date – path/row]	Data used for validation (CTRN) [update – n° of sections]
Torino	06.10.1999 (194/29)	02.07.2005 (194/29)	CTRN 2005 (5 sections)
	24.08.2001 (194/29)		
	28.09.2002 (194/29)		
	30.11.1999 (195/29)		
	05.06.2000 (195/29)		
	30.07.2001 (195/29)		
	02.07.2005 (194/29)		
Asti	06.10.1999 (194/29)	02.07.2005 (194/29)	CTRN 2004 (61 sections)
	24.08.2001 (194/29)		
	28.09.2002 (194/29)		
	23.11.1999 (194/29)		
	01.05.2000 (194/29)		
	21.06.2001 (194/29)		
	02.07.2005 (194/29)		

2.3. Applied methodology

The process of image classification may include the following major steps [5]: determination of a suitable classification system, selection of training samples, image pre-processing, features extraction, selection of suitable classification approaches, post-classification processing and accuracy assessment.

In this work some of these steps are considered variables; which pre-processing does return higher accuracy? Which features can be used in the classification? How can the results be improved with post-classification?

Different kind of pre-processing and post-processing were therefore tested, and the following features considered: attributes in input to the classifier, use of multitemporal data and the selection of 2 or more classes for the training phase of the classifier.

The procedure applied is the following: first, sample areas, to be used to create the training set to learn the classifier, are detected on satellite images with the help of high resolution. Different training sets on the detected sample areas are then built taking into account different variables: the radiometric pre-processing applied to images, the attributes used as input, the number of classes used to learn the classifier and, finally, the temporal extension of the dataset. The combination of all these variables led to the creation of 36 different training sets; each training set is then used as input to the classifier. The application of the same algorithm to the above mentioned training sets led to the definition of different classifiers, each with given attributes and decision thresholds. Each classifier is then applied to one common image; it derives as many classifications as classifiers are. Finally, different kinds of post processing are applied, thus resulting in a further increase in the number of thematic map produced. At the end, an accuracy assessment is performed for each thematic map. A schema of the procedure is provided in Figure 1.

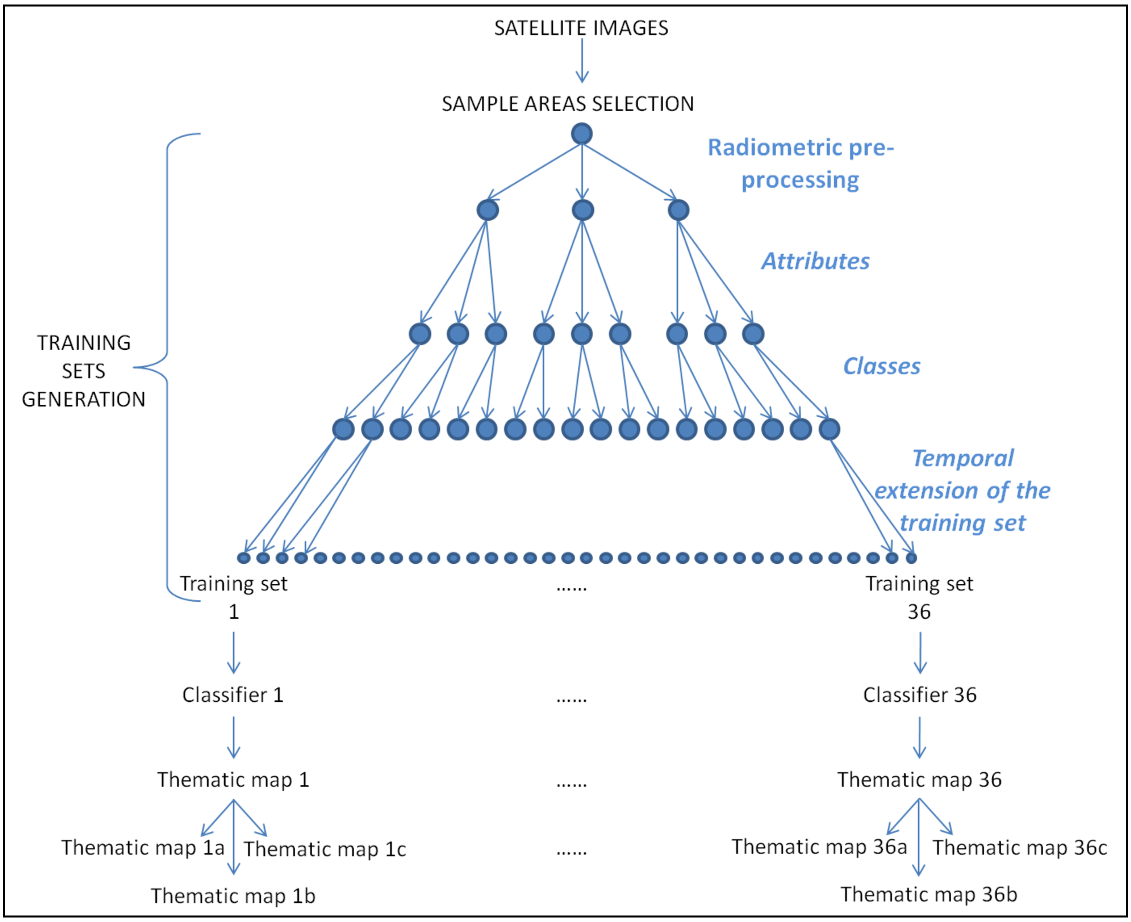

Figure 1. Applied methodology (output achieved at each step)

Pre and post-processing, together with the validation process and evaluated features are described in the following sections.

2.3.1. Radiometric pre-processing

Radiometric pre-processing is performed when scenes acquired at different times and from different sensors need to be compared; the aim is to correct errors due to the noise and distortions generated during the acquisition and transmission phase. Different methods can be used in order to compare scenes acquired in different times and from different sensors; one of the simplest one, which is called *radiometric calibration into reflectance,* allows obtaining apparent reflectance through normalization with respect to the incident radiation. Other methods, on the other hand, take also into account the atmospheric effects, and allow transforming spectral radiance into albedo.

Three different kind of radiometric pre-processing were tested in this case study: *radiometric calibration into reflectance*, atmospheric calibration using a simplified method (*Dark Subtraction [6],[7]*), atmospheric calibration using a rigorous method (*Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes*, FLAASH [8]).

The classifiers were trained with training sets extracted from images upon which these three pre-processing were performed alternatively; the aim is to verify if more complex (and most time-consuming) processing allow a significant increase in the classification accuracies.

2.3.2. Attributes

In this study, two main types of attributes were used in input to the classifier: spectral information and indexes. They were used both alone (only spectral information, only indexes) and combined. The aim is to evaluate if indexes are selected as suitable variables by the algorithm and, in case, which indexes more easily allow to separate classes.

The six bands provided by Landsat ETM+ images (excluded the thermal bands) are used as spectral information, while considered indexes, derived by the literature related to classification of urban areas, are the following:

- NDBI (Normalized Difference Built-up Index), used in previous studies [9] together with NDVI, and giving classification accuracies up to 92%;
- NDBBBI (Normalized Difference Blue Band Built-up Index), used in [10], that takes into account the blue component of urban areas and non cultivated fields;
- NDVI (Normalized Difference Vegetation Index), used for vegetation studies and useful in order to separate water, bare soil and vegetation classes;
- BUI (Built-Up Index), used in [11];
- A test_index, proposed by the authors on the basis of the spectral signature analysis over sample areas. This index shows values slightly above 1 (1.1 ÷ 1.2) in urban areas and higher values in the other classes.

Related formulas are summarized in Table 2; TM1, TM2, TM3, TM4, TM5, TM7 are, respectively, bands 1,3,4,5,7 of the Landsat Thematic Mapper (TM) sensor

Table 2. Indexes used in input to the classifier and related formula

Index	Formula
NDBI	$\frac{TM5 - TM4}{TM5 + TM4}$
NDBBBI	$\frac{TM1 - TM5}{TM1 + TM5 + 0.001}$
NDVI	$\frac{TM4 - TM3}{TM4 + TM3}$
BUI	NDBI - NDVI
Test_index	$TM5 + \frac{TM4}{TM7}$

2.3.3. Classes

In order to access if it's enough to learn the classifier on two macro class of Urban and Non Urban or if it's better to consider a higher number of classes in sample areas detection, thus creating a more complex tree where only rules for urban detection will be considered, sample areas belonging to the following classes were identified over the reference image: Urban (residential buildings), Industrial areas (industrial buildings), Vegetation, Water and Bare soil. Two options were tested during the creation of the training set: in one case input related to residential and industrial buildings were merged in order to create a unique "urban" class, and all the other merged in order to create a unique “non urban” class, in the second case each class is considered separately. Therefore in the following we refer to a 2-class classifier or to a 5-class classifier.

2.3.4. Temporal extension of the training set

The use of information derived from a single scene or from the mean over a temporal extent was tested, given the fact that the scope of the work was to obtain an algorithm able to automatically classify urban areas, potentially derived from different scenes. Therefore, information for training set generation were derived from a single scene in on one case (the one used in the classification process), and as a mean of the chosen multitemporal stack in the other case.

2.3.5. Post-processing

Post-classification processing is an important step for improving the quality of classifications [12][13]; filters can be applied to reduce the noise that come out from traditional per-pixel classifier (“salt and pepper” effect). The application of a majority filter with a kernel size 3x3 was herein tested.

2.3.6. Accuracy assessment

The CTRN of the Piedmont Region, at scale 1:10.000, was used to provide an Accuracy Assessment of obtained thematic maps.

CTRN polygon elements were firstly assigned to the correct class (Urban/Non Urban), then a conversion was performed into raster format in order to obtain a Validation Mask.

The comparison between thematic maps and the Validation Mask was performed trough the use of the Confusion Matrix (or error matrix) that represents a contingency table in which the diagonal entries represent the agreement between the map and reference data, while off-diagonal entries represent lack of agreement between the map and reference data. Measures of accuracy such as Overall Accuracy (OA), User’s Accuracy (UA), Producer’s Accuracy (PA) and f-measure were extracted from the Confusion Matrix and used for the comparison. The cited parameters are extensively described in literature, for further details see [14][15].

3. Results

Overall, 5 CTRN section (around 300.000 pixels) were used for the validation in Torino case study, while more than 61 CTRN sections (around 2.500.000 pixels) in the Asti case study. A subset of around 200.000 pixels was also considered in Asti case study, in order to validate an area more similar to the previous one (that is, more centered on a built-up area of medium dimension).

From a general point of view (Table 3), best Overall Accuracies were high in both cases: around 81% in the Torino case study, around 96%-97% in the Asti case study. By the analysis of the f-measure, that is representative of the accuracy in the detection of the Urban class, it can be inferred that the Urban class was detected better in Torino case study (best f-measure: 77%) than in Asti case study (best f- measure: 48%). The different behaviors can be attributed to the different urban land cover consistency: very dense in the first case, a medium to low density in the second case. This observation was confirmed by the following analysis: when the dimension of the validated area was reduced, and mainly focused on the greater town depicted in the scene (Asti), f-measure increased from 48% to 66%.

Table 3. Results: general overview

	TORINO	ASTI	ASTI subset
Validated pixel	300.000 (270 km^2)	2.500.000 (2.250 km^2)	200.000 (180 km^2)
Used CTRN sections	5	61	4
Best OA	81%	97%	96%
Best f-measure	77%	48%	66%

For each case study, 36 classifiers were obtained (taking into consideration the whole range of possible combination for training sets generation); the ones that provided better accuracies are reported in the following table (Table 4):

Table 4. Summary of best classifiers

Best classifiers (used features and radiometric pre-processing)	**Case study**	**OA – f measure [%]**
Single scene/ Indexes / Dark Subtraction / 2 classes	TORINO	79.9 % - 78.3%
Single scene/ Indexes / Dark Subtraction / 5 classes	TORINO	80.7% – 77.5%
Single scene/ Spectral info and indexes / FLAASH / 5 classes	ASTI subset	93.4% – 59.4%
Single scene/ Indexes / FLAASH / 2 classes	ASTI	97.7% – 46.5%
Single scene/ Spectral info and indexes / FLAASH / 2 classes	ASTI	97.7% – 48.1%
Single scene/ Spectral info and indexes / Dark Subtraction / 2 classes	ASTI subset	96.1% – 66.6%

It is noticeable how all best classifiers make use of indexes, both alone and together with spectral information; moreover, all considered classifiers exploit information from a single image and are obtained from an atmospherically corrected image.

Since accuracy measures derived from the error matrix provide information on the quality of the map as a whole but cannot be used to characterize distinct areas of the map [16], a quality control was performed in order to identify the most recursive errors in classifiers that provided accurate results. An automatic procedure was created for this aim in order to create an error mask, where the information associated to each single pixel is the membership to the Urban or Non Urban class, together with the number of times in which the considered pixel is correctly classified by the “good classifiers”. Results show that cemeteries, excavation sites and bare soil are areas more frequently confused with the urban class; in small built-up areas, errors are mainly located along the built-up area boundaries, and the same built-up area detection is related to the built-up area density: the more dense it is, the easier is to detect.

In order to evaluate the best variables to use for the classification process, the different obtained classifiers were compared by analyzing one by one each variable.

As long as the radiometric pre-processing is concerned, obtained accuracies show that none of the applied corrections could be considered better than others.

The application of a majority filtering with kernel 3x3 was tested; it generally improved all accuracy measures, and is particularly useful where medium density urban land cover and very sparse urban land cover are considered, otherwise it doesn’t improve the accuracies significantly.

The classifiers were then learned using 2 classes or 5 classes, and using a single image or a multitemporal stack. Obtained results diverged considerably: 2-class classifiers were better than others in Asti case study, particularly in post-processed classifications, less good when the validation was performed only on a subset; on the contrary, in Torino case study accuracies were higher when 5-class classifiers were used in nearly 61% of cases (Table 5).

Table 5. Comparison between classifiers learned with 2-class or 5-class training sets in terms of number of times in which one resulted better than the other and of mean reached improvement

	Post processing: YES						Post processing: NO					
	OA [%]			f-measure [%]			OA [%]			f-measure [%]		
	TO	AT	AT subset	TO	AT	AT subset	TO	AT	AT subset	TO	AT	AT subset
Better results with 2-class	39	89	66	39	78	34	39	72	55	33	61	44
mean improvement	2.3	3.0	3.3	1.5	5.8	12.0	1.0	3.9	4.0	1.3	5.1	6.3
Better results with 5-class	61	11	34	61	22	66	61	28	45	67	39	56
mean improvement	5.2	14	3.7	3.1	9.9	5.1	4.6	6.4	3.4	2.6	3.1	4.7

Considering the use of a multitemporal stack, the results were quite similar: in Asti case study single image classifiers provided better results than others in most of the cases, sometimes in a consistent way (the f-measure of post-processed images was, on average, 20% better in single image classifiers with respect to 5-class classifiers), while in Torino case study multitemporal classifiers were better than other in most of the cases and with OA greater, on average, of around 13% (Table 6). Other tests are necessary in order to decide which is the best approach to use.

Table 6. Comparison between classifiers learned with single image or multitemporal stack in terms of number of times in which one resulted better than the other and of mean reached improvement

	Post processing: YES						Post processing: NO					
	OA [%]			f-measure [%]			OA [%]			f-measure [%]		
	TO	AT	AT subset	TO	AT	AT subset	TO	AT	AT subset	TO	AT	AT subset
Better results with single image	11	83	94	17	100	94	11	55	83	17	100	100
mean improvement	0.7	5.0	4.4	2.9	20	20.9	1.7	7.8	5.0	3.5	16	15.8
Better results with multitemporal stack	89	17	6	83	-	6	89	45	17	83	-	-
mean improvement	13	1.3	1.5	5.9	-	1.3	12	0.9	1.0	4.5	-	-

Finally, concerning the use of different attributes results differ in the two considered cases. “Index classifiers” were better in Torino case study, while “spectral” and “spectral and index” classifiers provided higher accuracy in Asti case study and, in most cases, very similar to each other.

On the contrary more evidences can be found analyzing the attributes that were more often chosen by the algorithm to be part of the classifier: some indexes were considered a very few times (BUI), others were widely used in both cases (NDVI and NDBBBI). The test index was often chosen by the classifier. Among considered bands, band 1 was the most used by the algorithm,. This can be due to the high correlation between the three bands: one of the three can be considered representative for the whole set. Band 4, 5 and 7 were considered quite often in both cases.

4. Conclusions

In the automation of feature extraction the presence of mixed pixels, typical of the urban environment, makes more difficult to identify common features, particularly if only spectral information is considered. This work demonstrates the effectiveness of Decision Tree classifiers for the considered aim, the importance of certain features (such as the use of indexes) for classifier generation, the independence from other features (such as the considered number of classes for training set generation), and the improvement given by post-processing. The possibility of having a procedure for the extraction of this thematism, also with relatively coarse accuracies, can be considered a good starting point especially in applications as emergency mapping, where information need to be extracted quickly and where even the partial reduction of manual digitalization procedures, can be considered a breakthrough.

The study proposed in this article should be considered as the first step of a wider work aimed at automatically identifying urban data from medium resolution imagery. In the future, this method will be tested on different areas and the use of other input parameters (e.g. textural information) will be examined, as well as other processing techniques applied to the image (e.g. segmentation techniques) in order to facilitate the theme extraction.

Acknowledgements

The authors would like to thank ESA for the sponsorship provided for the participation to the 33rd EARSeL Symposium, the USGS Earth Resources Observation and Science (EROS) Center and Piedmont Region for the data that was accessed through the internet free of charge, the University of Waikato for creating and providing the open source data mining software WEKA.

References

[1] Boccardo, P. & F. Giulio Tonolo, 2012. *Remote-sensing techniques for natural disaster impact assessment*. Advances in mapping from remote sensor imagery (Taylor & Francis Group) pp. 387–414

[2] Pesaresi, M. & M., Halkia, 2012. *Global Human Settlement Layer / Urban Atlas Integration: Feasibility Report*. Ispra, Italy.

[3] Otukei, J. R. & T. Blaschke, 2010. *Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms*. International Journal of Applied Earth Observation and Geoinformation 12, S27–S31.

[4] Quinlan, J. R., 1993. *C4.5: Programs for Machine Learning* (M. Kaufmann, ed.) Vol. 1, p. 302

[5] Lu, D. & Q., Weng, 2007. *A survey of image classification methods and techniques for improving classification performance* (International Journal of Remote Sensing) 28, pp. 823–870

[6] Chavez, Jr. PS,1988. *An Improved Dark-Object Subtraction Technique for Atmospheric Scattering Correction of Multispectral Data* (Remote Sensing of Environment) 24, pp. 459-479

[7] Chavez, Jr. PS, 1989. *Radiometric Calibration of Landsat Thematic Mapper Multispectral Images* (Photogrammetric Engineering and Remote Sensing) 55, pp. 1285-1294

[8] Adler-Golden, S., Berk, A., Bernstein, L. S. & Richtsmeier, S., 1998. *FlAASH, A MODTRAN4 atmospheric correction package for hyperspectral data retrievals and simulations* (Summaries of the seventh JPLairborne earth science workshop), JPL-Pub., vol. 97-21 (pp. 1 – 9). Pasadena, CA: JPL Pub.

[9] Zha, Y., Gao, J., Ni, S., 2003. *Use of normalized difference built-up index in automatically mapping urban areas from TM imagery* (International Journal of Remote Sensing) 24,3, pp 583-594.

[10] Baraldi, A., V. Puzzolo, P. Blonda, L. Bruzzone, & C., Tarantino, 2006. *Automatic Spectral Rule-Based Preliminary Mapping of Calibrated Landsat TM and ETM+ Images*. (IEEE Transactions on Geoscience and Remote Sensing) 44: 2563–2586.

[11] Lee, J. I. N. A., S. S. Lee, & K. H. Chi, 2010. *Development of an urban classification method using a built-up index* (Selected topics in power systems and remote sensing) pp 39–43

[12] Stefanov, W. L., M. S. Ramsey, & P. R. Christensen, 2001. *Monitoring urban land cover change: an expert system approach to land cover classification of semiarid to arid urban centers*. (Remote Sensing of Environment) 77:173–185.

[13] Lu, D. & Q. Weng, 2004. *Spectral Mixture Analysis of the Urban Landscape in Indianapolis with Landsat ETM + Imagery*. (Photogrammetric Engineering & Remote Sensing) 70:1053–1062.

[14] Congalton, R. G., 1991. *A review of assessing the accuracy of classifications of remotely sensed data* (Remote Sensing of Environment) 37(1): 35–46.

[15] Foody, G. M., 2002. *Status of land cover classification accuracy assessment* (Remote Sensing of Environment) 80(1): 185–201.

[16] Mayaux, H. E., Gallego, J., Strahler, A. H., Herold, M., Agrawal, S., Naumov, S., *et al.*, 2006. *Validation of the global land cover 2000 map* (IEEE Transactions on Geoscience and Remote Sensing) 44:1728–1739.