

URBAN LAND COVER CLASSIFICATION USING OPTICAL VHR DATA AND THE KNOWLEDGE-BASED SYSTEM INTERIMAGE

H. J. H. Kux^a, T. Novack^a, R. Ferreira^b, D. A. Oliveira^b

^a INPE, Remote Sensing Division, 12.227-010 São José dos Campos, Brazil - (hermann, tessio)@dsr.inpe.br

^b PUC-RIO, Dept. of Electric Engineering, Rio de Janeiro, Brazil (rodrigo, dario)@ele.puc-rio.br

Commission VI, WG IV/4

KEY WORDS: InterIMAGE System, Knowledge-based Image Classification, Urban Land Cover.

ABSTRACT:

This study uses the InterIMAGE system and imagery from the QuickBird II optical sensor for the classification of the land cover of two test-sites in the metropolis of São Paulo, Brazil. InterIMAGE is an open source and free access framework for knowledge-based image classification. Within InterIMAGE human knowledge is represented as a semantic net and by user-defined rules based on the paradigms of object-oriented image analysis. The Segmentation Parameter Tuner system was used for the search of appropriate segmentation parameters. This system uses a genetic algorithm for the optimization of the parameters. For the description of the land cover classes in terms of features and thresholds, a strategy combining a hierarchical structure of classes and the decision tree algorithm C4.5 was elaborated. The classification model was then formatted according to InterIMAGE's image analysis strategy. The proposed methodology has shown to be efficient for the automatic mapping of the land cover in complex urban areas such as those found in the metropolis of São Paulo. The final classification achieved an overall accuracy of 73% and a *Kappa* Accuracy Index of 0.69. This study has explored the main functionalities of InterIMAGE, presenting its potentialities for object- and knowledge-based image classification.

1. INTRODUCTION

According to the United Nations Organization (UNO), presently more than half of the world population lives in urban areas and by 2050 this percentage will increase to 70% (UN-HABITAT, 2009). It foresees an absolute and a relative increase of urban population especially in developing countries. The chaotic growth of cities in these countries causes different social problems, such as violence, traffic jams, visual and sound pollution, decrease of environmental quality, increase of diseases related to air pollution and depreciation of public spaces among others. The mitigation and solution of these problems involves effective urban planning policies based on up-to-date and reliable spatial information.

In this context, remote sensing datasets can provide valuable information about the urban land cover. According to Jensen (2007), for urban land cover and land use classification at the 3rd and 4th scale levels of the United States Geological Survey, remote sensing multispectral imagery with a temporal resolution of at least 3 to 5 years and spatial resolution higher than 5 meters (< 5m) is required. In the last decade, several orbital sensors with these characteristics were launched. The higher availability of data demands now adequate (i.e. accurate, standardized and fast) methodologies for information extraction. The object and knowledge-based image classification approach is presently the most advantageous for the analysis of high spatial resolution imagery for urban planning.

Despite the several potentialities of this approach, the costs of licenses for the use of commercial software that perform object-based and knowledge-based image classification are still very high. Another disadvantage is that these systems are not open source, which means that one cannot have access to the code of the algorithms nor customize the system according to one's

own needs. The free access to these systems would allow planning agencies and research institutes, especially in developing countries, to make full use of this technology.

In that context, the objective of this work is to present an urban land cover classification model using the InterIMAGE system (InterIMAGE 2010) and imagery from the QuickBird sensor (DigitalGlobe 2008). InterIMAGE is a free access and open-source knowledge-based classification system whose image analysis conception is highly flexible, offering great potential for automatic interpretation of remote sensing imagery (InterIMAGE 2010). The system is an on-going project whose main long term objectives are to spread its use among government planning agencies and to develop a powerful tool for image analysis.

1.1 The InterIMAGE System

The InterIMAGE system is a knowledge-based framework for automatic image interpretation which is currently implemented for LINUX and Windows operating systems. The image analysis concept of InterIMAGE is based on an earlier system called GeoAIDA developed at Leibniz University in Hannover (Bückner et al. 2001; Pahl 2003). The classification strategy implemented by InterIMAGE is based on a knowledge model structured as a semantic net defined by the user. The classification strategy has two steps: the Top-Down (TD) and the Bottom-Up (BU).

In the TD step, the system descends the semantic net triggering the so-called holistic operators. Holistic operators are image processing operators, external to the system's core, specialized in the detection of a certain class. Every node of the semantic net may or may not contain a holistic operator. These can be developed by any user who has programming skills.

InterIMAGE (2010) also offers an online repository of operators, developed already by the project team. In theory, the holistic operators can process any type of image, enabling InterIMAGE to perform multi-sensor analysis.

For the detection of objects from a certain class, holistic operators usually perform three procedures in the following order: (1) segmentation (or import GIS data), (2) attribute extraction and (3) classification. The geographic regions detected by a holistic operator inserted into a given node are transmitted as masks (hypothesis) to its child nodes on the lower level of the semantic net, where its own holistic operators will work. Presently, the user can elaborate a holistic operator considering: (1) one of the two segmentation algorithms available in the system, namely those ones developed by Baatz and Schäpe (2000) and Bins et al. (1996), (2) spectral, textural and geometric features and (3) classification rules using different boolean operators structured, if necessary, hierarchically.

In the BU step, the system ascends the semantic network solving spatial conflicts between hypotheses based on user-defined rules inserted in every node that is not a leaf node. Doing so, the system either partially or totally discards the hypotheses or turns them into instances (i.e. validates the hypotheses). These user-defined rules may or may not involve additional logical selections. If after the discard of hypotheses by these additional logical selections spatial conflicts still remain, they are solved either by the supervised definition of priority for the classes or by the competition of membership values given by user-defined fuzzy membership functions. BU rules are customized in a friendly interface that allows the development of complex class description and hypothesis judgment criteria.

2. TEST-SITES AND METHODS

The test sites for the application of the methodology are two sections of the São Paulo (Brazil) municipality. One is a high social-economic standard residential area with two types of land use, namely: residential horizontal of high standard and vertical residential of high standard. The most common targets in that area are vegetation cover of trees and grass, ceramic tile roofs, swimming pools and concrete buildings painted in white. The other test site is an industrial and residential area with bare soil area, ceramic tile roofs, different kind of asbestos roofs and pavement types. Both test-sites have a modest size of about 810.000 square meters. This limitation of image size is actually a limitation on the number of hypotheses (i.e. segments) that InterIMAGE could process at the time of the images processing phase of this work.

As explained below, the land cover classification in the two test-sites was done combining semantic nets for human knowledge representation and machine learning algorithms for segmentation parameters search and classification rules definition. This combination has the intent to minimize the intervention of the user to the tasks of collecting samples for every land cover class and organizing the semantic nets. The proposed methodology is to be understood as an image analysis processing chain that in the near future will be fully integrated into the InterIMAGE system.

2.1 Segmentation Parameters Search and Definition

The algorithm developed by Baatz & Schäpe (2000) was used for the segmentation of the images. Two parameter sets were

applied, one for the segmentation of vegetated areas (named PS 1) and another for medium and small sized roofs (named PS 2) (Table 1). Since the relationship between the parameter sets and the segments generated by the algorithm is not intuitive, a process of calibration of the parameters is always necessary. Instead of a time-consuming trial-and-error process, the system Segmentation Parameter Tuner (SPT 2008; Costa et al. 2008) was used for this purpose. The SPT uses a genetic algorithm to search for the optimal parameter set based on the fitness between samples drawn by the user over the image and the segments generated by the algorithm. The set of parameters that generates the segments most fitted to the samples is recommended by SPT as the best set when the searching process ends. The fitness measure used was the Reference Bounded Segment Booster (RBSB). Costa et al. (2008) give a detailed explanation of how the calibration of parameters using the SPT system works as well as on the calculation of the RBSB fitness measure. The SPT system can be downloaded from the Internet and in the near future it will be integrated into the InterIMAGE system.

<i>Parameter sets (PS)</i>		
Parameters	PS 1	PS 2
Scale	21	41
Shape	0.25	0.84
Compactness	0.61	0.8
Band weights (B1, B2, B3, B4)	0.24, 0.07, 0.42, 0.26	1,1,1,1

Table 1 – Parameter sets applied for the segmentation of the land cover objects.

2.2 Land Cover Class Description

Similarly to the calibration of the segmentation parameters, the description of the classes in terms of features and thresholds was obtained automatically. First, the segments generated on the SPT system were imported into the Definiens Developer system (Definiens 2007) and forty-five samples of each land cover class were collected and then exported along with ninety-six spectral, geometrical and textural features. The samples were then organized according to the hierarchical net shown on Figure 1.

The hierarchical net served as a strategy to obtain the description of the land cover classes through a series of sub-classifications on a top-down process. Initially, all the samples were grouped as belonging either to class ‘Shadow’ or ‘Not-Shadow’, then the samples of ‘Not-Shadow’ were divided into groups ‘Vegetation’ and ‘Not-Vegetation’ and so on. Each one of these sub-classifications was then derived using the C4.5 decision tree algorithm (Quinlan, 1993). This algorithm provides, in a hierarchical structure, the features and thresholds that best separates the samples from different classes based on the heuristic of entropy reduction. A decision tree can also be considered as a rule set or simply as a tool for threshold calibration.

Simultaneously to this task, the Fast Correlation-Based Filter (FCBF) (Yu and Liu, 2003), a feature selection algorithm, was used to extract features that best separate the samples belonging to the leaf nodes of the hierarchical net (i.e. the classes of interest). The frequency distributions of the samples concerning the two features that were best ranked by the FCBF algorithm were considered for the shape definition of the membership functions associated to these features. The FCBF algorithm was also used for the selection of the features for the decision tree

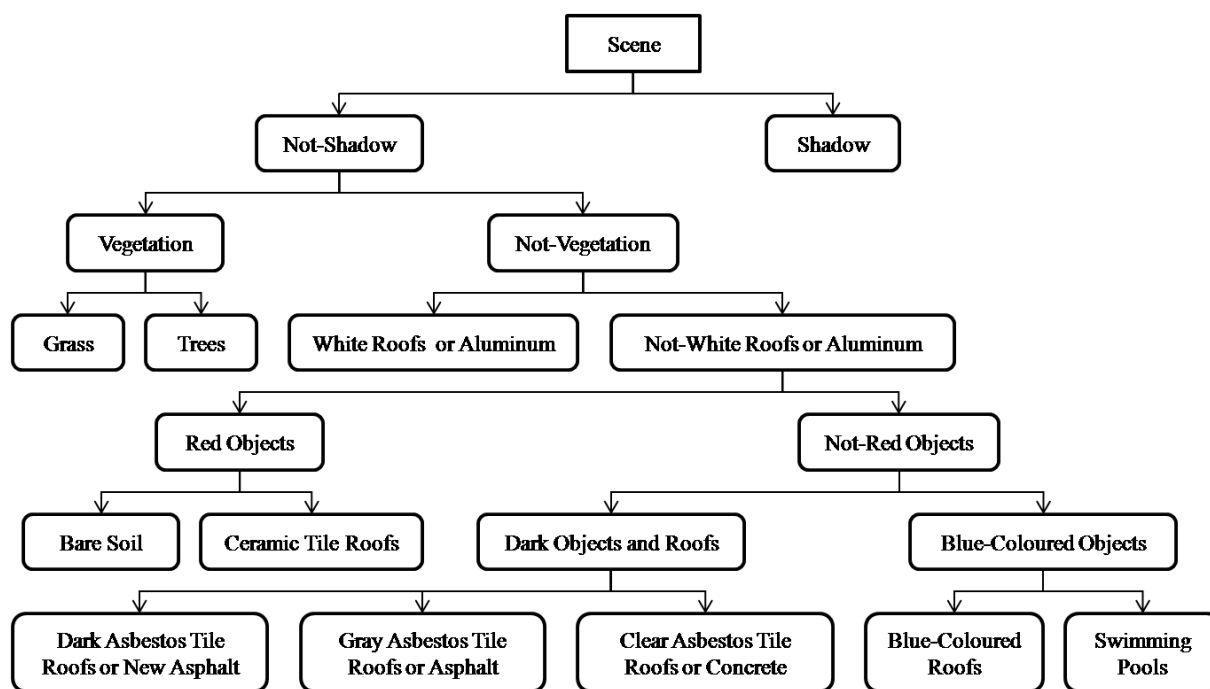


Figure 1 – Hierarchical net used as a strategy to obtain the description of the land cover classes.

induction. In Novack and Kux (2009) we observed that the reduction of the dimensionality causes an increase of the accuracy of the decision tree and also reduces its size.

In this way both the set of logical selections (given by the decision trees) and the membership functions that describe the land cover classes of interest were obtained. For instance, the description of class 'Grass' is comprised by the rules given by the decision trees for the classification of 'Not-Shadow' and 'Vegetation' as well as the two membership functions associated to the two features that were best ranked by the FCBF algorithm for the separation of 'Grass' and 'Trees'. Therefore, the class 'Trees' has at its description the same rules and features as 'Grass', but different membership functions although associated to the same features. The same process was applied to obtain the description of classes 'Bare Soil', 'Ceramic Tile Roofs', 'Blue-Coloured Roofs', 'Swimming Pools', 'Dark Asbestos Tile Roof or New Asphalt', 'Gray Asbestos Tile Roofs or Asphalt' and 'Clear Asbestos Tile Roofs or Concrete'. At the description of all these classes, the operator 'Minimum' was applied for the aggregation of the membership values returned by the functions associated with the two features best ranked by the FCBF algorithm.

All the features used in the class descriptions of the land cover classes were then implemented in InterIMAGE in order to apply the model.

2.3 Formatting of the Land Cover Classification Model

As the InterIMAGE system uses a two-step analysis strategy (i.e. the TD and the BU steps) a new semantic net was elaborated (Figure 5) where the expressions of the class descriptions were organized and structured for the generation of hypothesis in the TD step and the resolution of spatial conflicts between these hypothesis in the BU step. Our objective was to explore the potential of the two-step analysis strategy in a more semantically coherent and computationally less expensive net.

In the TD step the operator *Dummy Top-Down* was inserted into the nodes 'Service and Residential Roofs', 'Red Objects' and 'Vegetation'. This operator only passes to the child nodes the masks it received from the parent node, which in this case is the whole image. The node 'Swimming Pools' only segments the images using the parameter set PS 2 (Table 1) and the node 'Shadow' selects the pixels of a brightness image (sum of the four spectral bands divided by four) with digital numbers below 200. On the third level of the semantic net the nodes 'Grass' and 'Trees' perform the segmentation of the image applying the parameter set PS 1 (Table 1) as well as the logic selection given by the C4.5 algorithm for the separation of classes 'Vegetation' and 'Not-Vegetation'. The same procedures are done on nodes 'Bare Soil' and 'Ceramic Tile Roofs', except that the segmentation parameter set is the PS 2 shown in Table 1 and the logic selection is given by the C4.5 algorithm for the separation of 'Red Objects' and 'Not-Red Objects'. For the child nodes of 'Service and Residential Roofs' the PS 2 segmentation parameters in Table 1.0 were applied. As for the logic selections, we applied for class 'White Roofs or Aluminium' those ones given by the C4.5 for the separation of 'White Roofs or Aluminium' and 'Not-White Roofs or Aluminium' on the hierarchical net shown on Figure 2. At the nodes 'Clear Asbestos Tile Roofs or Concrete', 'Gray Asbestos Tile Roofs or Asphalt' and 'Dark Asbestos Tile Roofs or New Asphalt', the logic selections used for the separation of 'Dark Objects and Roofs' and 'Blue-Coloured Objects' were applied. The node 'Blue-Coloured Roofs' performs the segmentation of the image using the PS 2 shown in Table 1 as well as the logical selections that are the opposite to those ones applied for the classification of classes 'White Roofs or Aluminium', 'Service and Residential Roofs' and 'Swimming Pools' in the hierarchical net shown in Figure 2.

In the BU step the decision rule on node 'Service and Residential Roofs' solves the spatial conflicts among the hypotheses of its child nodes as follows. The hypotheses of the

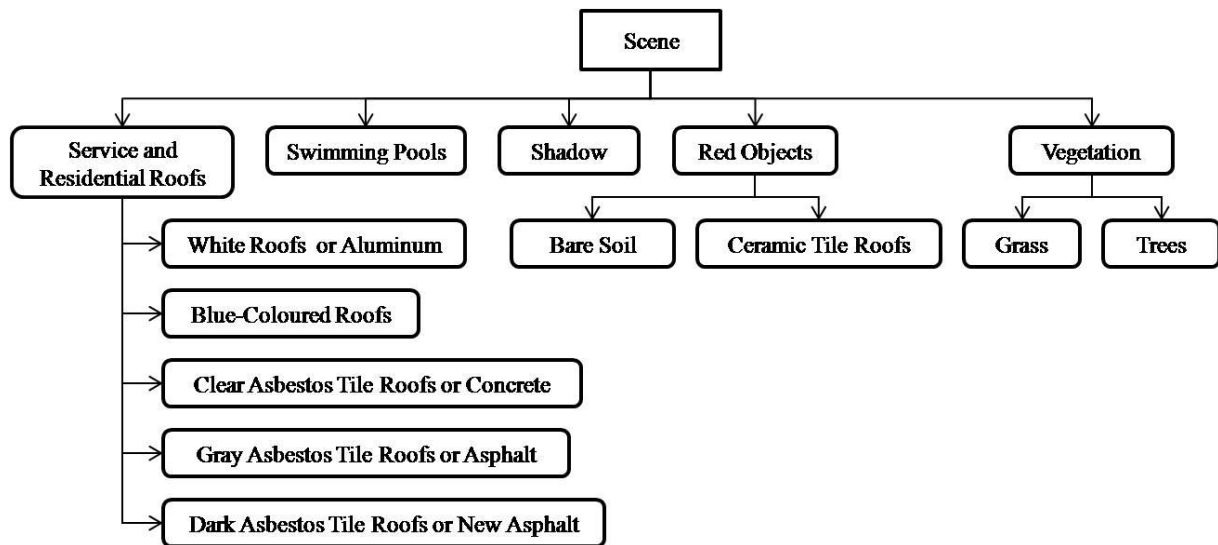


Figure 2 – Semantic net elaborated in the system InterIMAGE for the classification of the land cover.

class ‘White Roofs or Aluminum’ wins all spatial conflicts and the hypotheses of class ‘Blue-Coloured Roofs’ loses all spatial conflicts. The solution of spatial conflicts between the hypotheses of classes ‘Clear Asbestos Tile Roofs or Concrete’, ‘Gray Asbestos Tile Roofs or Asphalt’ and ‘Dark Asbestos Tile Roofs or New Asphalt’ is completed by the competition of the membership values given by the membership functions associated to the two features best ranked by the FCBF algorithm. The BU decision rules on nodes ‘Red Objects’ and ‘Vegetation’ also solves the spatial conflicts between the hypotheses of its child nodes using the membership functions

associated to the features selected by the FCBF algorithm. Finally, the remaining spatial conflicts between hypotheses of classes ‘Service and Residential Roofs’, ‘Swimming Pools’, ‘Shadow’, ‘Red Objects’ and ‘Vegetation’ are solved according to an order of priority. Class ‘Shadow’ wins all spatial conflicts followed by classes ‘Vegetation’, ‘Red Objects’, ‘Service and Residential Roofs’ and ‘Swimming Pools’. Alternatively to the definition of an order of priority for the solution of these spatial conflicts, the use of membership functions was also examined. However, the results were visually evaluated as inferior to the results obtained when defining an order of priority of classes.

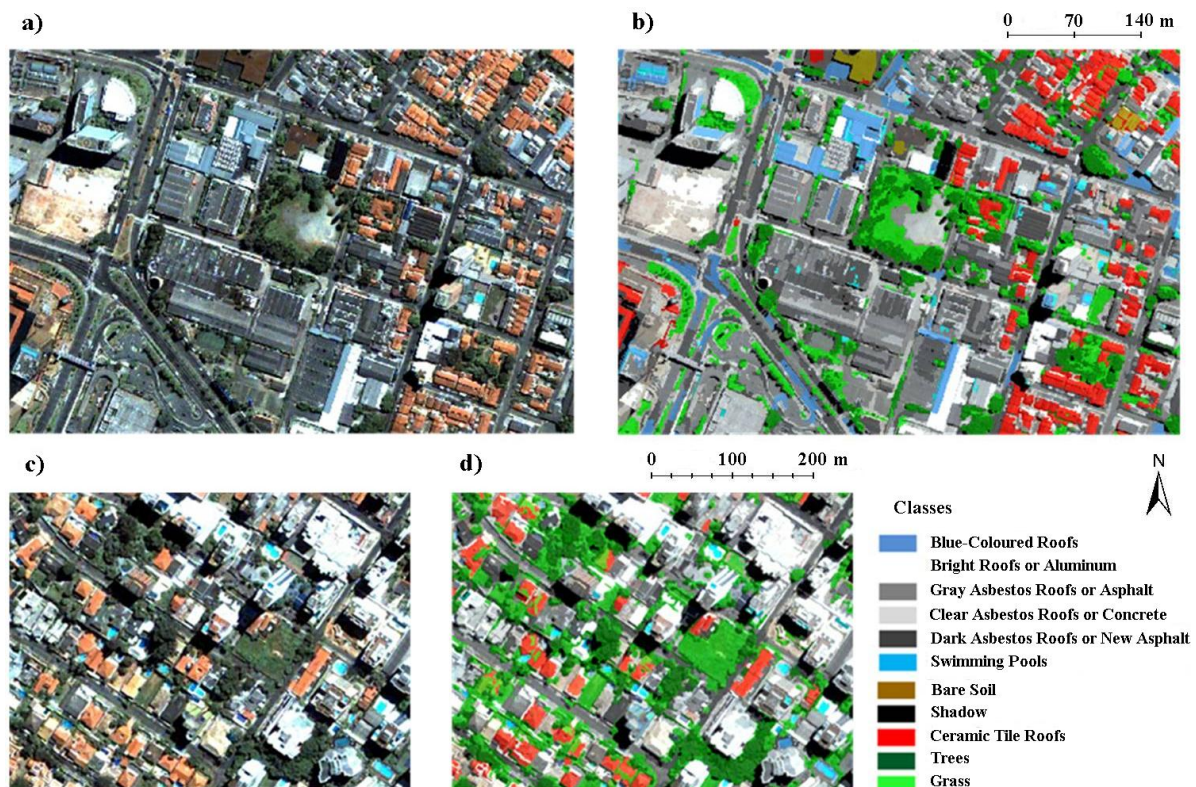


Figure 3 – Colour compositions (a and b) and the respective land cover classifications (c and d).

Naturally, this depends on the shape and thresholds of the membership functions. It is important to stress that this alternative can only be explored on the InterIMAGE system.

3. RESULTS AND DISCUSSIONS

For the evaluation of the accuracy of the land cover classifications, the following procedure was carried out. First, samples of every land cover class were drawn over the image by visual interpretation. Most of the objects whose land cover class could be clearly identified were sampled. Then, the vector layer containing the drawn samples was converted to a raster layer and the resulting reference map was compared pixel by pixel with the classifications obtained. Such evaluation is more realistic than collecting whole segments as samples due to the fact that an area wrongly or correctly classified will be weighted by its area in the accuracy index computation.

Figure 3 shows the land cover classifications at sections of the two test-sites and Table 2 present the confusion matrix of both classifications. A Global Accuracy of 73% and a Kappa Index (Congalton and Green 1999) of 0.69 were obtained. These values are satisfactory and corroborate the made visual

evaluation. With few exceptions, all land cover classes achieved very good User's and Producer's Accuracies. The vegetated areas (the area covered either by 'Trees' or 'Grass') were excellently mapped if we consider that most of the omission errors of class Trees were classified as 'Grass'. On the other hand, 7% of the sampled 'Bare Soil areas were wrongly classified as 'Grass', which has decreased the user's accuracy of this class down to 69%, along with the commission errors over 'Trees mentioned. In fact, there are many areas in the image where the grass is sparse, becoming a transition area to bare soil. As for the confusion among 'Trees and 'Grass', it is related to the fact that, depending on its position in relation to the sun, some parts of the tree canopies appear in a brighter tone of green which is similar to the one of grass covered areas. The classification of 'Bright Roofs' has commission and omission errors almost exclusively with classes 'Bare Soil' and 'Clear Asbestos Roofs or Concrete'. 30% of the samples collected as 'Bright Roofs' were wrongly classified as 'Clear Asbestos Roofs or Concrete'. However, such confusion is acceptable when asbestos roofs or concrete surfaces are new and clean, as they reflect sunlight similarly to bright roofs. Bare

Classes	Trees (T)	Grass (G)	Bright Roofs (BR)	Grey Asbestos Roofs or Asphalt (GARA)	Clear Asbestos Roofs or Concrete (CARC)	Dark Asbestos Roofs or New Asphalt (DARNA)	Shadow (S)	Swimming Pools (SP)	Ceramic Tile Roofs (CTR)	Blue-Coloured Roofs (BCR)	Bare Soil (BS)	Total
T	2950	28	0	63	0	92	0	0	8	0	0	3141
G	1084	17136	0	16	0	23	0	3	0	0	1820	20082
BR	0	0	2780	0	0	0	0	0	0	38	659	3477
GARA	14	761	26	3064	266	572	0	27	122	6	452	5310
CARC	0	2	1237	94	9554	15	0	71	198	388	12412	23971
DARNA	154	165	0	881	0	6033	96	3	129	133	87	7681
S	11	0	0	9	0	9	7328	0	0	2	0	7359
SP	0	0	158	0	0	0	0	1308	0	221	0	1687
CTR	9	0	0	0	0	3	0	0	9034	0	2899	11945
BCR	0	0	0	0	114	0	0	99	0	9654	0	9867
BS	0	0	0	0	0	0	0	0	1935	0	6167	8102
Total	4222	18092	4201	4127	9934	6747	7424	1511	11426	10442	24496	102622
User's Accuracy	0.94	0.85	0.80	0.58	0.40	0.79	1.00	0.78	0.76	0.98	0.76	
Producer's Accuracy	0.70	0.95	0.66	0.74	0.96	0.89	0.99	0.87	0.79	0.92	0.25	

Table 2 – Confusion matrix and per class User's and Producer's Accuracies.

soils can also be very reflective, which has caused confusion with 'Bright Roofs' mainly with 'Clear Asbestos Roofs or Concrete'. In fact, the class 'Bare Soil' is the one with relatively most omission errors (a Producer's Accuracy of only 25% was achieved).

Classes 'Grey Asbestos Roofs or Asphalt', 'Clear Asbestos Roofs or Concrete' and 'Dark Asbestos Roofs or New Asphalt' presented mainly commission and omission errors among each other and with class 'Bright Roofs'. This can be explained by the fact that a clear and objective criterion of distinction between these classes cannot be defined, especially because of different conditions of preservation and inclination angles of paved areas and roofs. This has lowered the accuracy indexes of these classes. At class 'Shadow' there were only few commission errors with 'Trees' and 'Asbestos Roofs' and a small percentage (1.2%) of the samples collected from this class were wrongly classified as 'Dark Asbestos Roofs or New Asphalt', which is the second less reflective class considered in this work.

Other classification confusion worth mention is that between 'Ceramic Tile Roofs and Bare Soils'. These two classes have very similar chemical compositions and Kux and Araújo (2008) and Novack (2009) have already reported the similar spectral characteristics of these two classes when using multispectral images with high spatial resolution. As shown in Table 2, almost 22% of the samples collected for 'Ceramic Tile Roofs' were wrongly classified as 'Bare Soils' and all commission errors of this class were attributed to class 'Ceramic Tile Roofs'.

4. CONCLUSIONS AND REMARKS

This study demonstrates that the InterIMAGE system is a powerful tool for knowledge-based analysis of high resolution remote sensing imagery. InterIMAGE performed complex land cover classifications at two different and complex urban test-sites using spectral and geometric features calculated both on the TD step and on the BU rules. The functionalities implemented in InterIMAGE in order to make this application possible have certainly improved the system. The land cover classification achieved very good accuracy indexes and was qualitatively evaluated as very good.

One still limitation of InterIMAGE is the modest number of segments it can process, which directly determines the limit of the image's size. The project's development crew is already working to overcome this restriction.

The partial results presented in this study demonstrate that the InterIMAGE system could be used by urban planning agencies for mapping the land cover of cities in a fast and standardized way with no cost. In future work, topological features recently implemented at InterIMAGE will be explored and the model will be tested in more extensive areas for the evaluation of its accuracy stability.

Acknowledgements

This study is part of the MSc-Thesis from the second author. He acknowledges the support given by the National Institute for Space Research (INPE, Brazil) and the Coordination of Improvement of Higher Education Personnel (CAPES, Brazil) for supplying the Masters scholarship. The QuickBird imagery was kindly delivered by the company *Imagem Sensoriamento Remoto*®. The joint work and fruitful discussions with the software development team at PUC-Rio and INPE-DPI is also acknowledged.

REFERENCES

- Baatz, M.; Schäpe, A. 2000. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T. *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT Symposium Salzburg 2000*, Herbert Wichmann Verlag, Karlsruhe.
- Bins, S.A., Fonseca, L. M. G., Erthal, G. J., 1996. Satellite Imagery segmentation: a region growing approach, *Proceedings of the VIII Simpósio Brasileiro de Sensoriamento Remoto (SBSR)*, Salvador, Brazil, pp. 677-680.
- Bückner, J.; Stahlhut, O., Pahl, M., Liedtke., C. E., 2001. GEOAIDA - A knowledge based automatic image data analyzer for remote sensing data. *Proceedings of the II International ICSC Congress on Computational Intelligence: Methods & Applications*, Millet, Alberta: ICSC, Bangor, Wales, United Kingdom.
- Congalton, R.G., Green, K., 1999. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, New York.
- Costa, G. A. O. P.; Feitosa, R. Q.; Cazes, T. B.; Feijó, B., 2008. Genetic adaptation of segmentation parameters. In: Blaschke, T.; Lang, S.; Hay, G.F. (Eds) *Object-Based Image Analysis – Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, Springer Verlag, Berlin, Germany, pp. 679-695.
- Definiens, 2007. Definiens Developer User Guide. <http://www.definiens-imaging.com/down/ecognition>
- DigitalGlobe, 2008. QuickBird Imagery Products – Products Guide. <http://www.digitalglobe.com> (accessed 23 Apr. 2008)
- InterIMAGE, 2010. InterIMAGE – Interpreting Images Freely. <http://www.lvc.ele.puc-rio.br/projects/interimage/> (accessed 3 Jan. 2010)
- Jensen, J.R., 2007. *Remote Sensing of Environment: An Earth Resources Perspective*. Pearson Prentice Hall, New Jersey.
- Kux, H.J.H.; Araújo, E.H.G., 2008. Object-based image analysis using Quickbird satellite images and GIS data, case study Belo Horizonte (Brazil). In: Blaschke, T.; Lang, S.; Hay, G.F. (Eds) *Object-Based Image Analysis – Spatial Concepts for Knowledge-Driven Remote Sensing Applications*, Springer Verlag, Berlin, Germany, pp 571-588.
- Novack, T., Kux, H. J. H., 2009. Classificação da cobertura do solo urbano inserindo árvores de decisão a rede hierárquica. Proceedings of the XIV Simpósio Brasileiro de Sensoriamento Remoto (SBSR), Natal, Brazil. Available at: <http://urlib.net/dpi.inpe.br/sbsr@80/2008/11.17.17.11> (accessed 3 Jan. 2010)
- Novack, T., 2009. *Classificação da cobertura da terra e do uso do solo urbano utilizando o sistema InterIMAGE e imagens do sensor QuickBird*. MSc. Dissertation. National Institute for Space Research (Brazil). <http://mtc->

m18.sid.inpe.br/rep/sid.inpe.br/mtc-m18@80/2009/08.31.21.23.
(accessed 3 Jan. 2010)

Pahl, M., 2003. *Architektur eines wissensbasierten Systems zur Interpretation multisensorieller Fernerkundungsdaten*. PhD-Thesis, Hannover: Leibniz University.

Quinlan, J.R. 1993. *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Francisco.

SPT., 2008. *Segmentation Parameters Tuner*. Computer Vision Lab. (LVC) from PUC-Rio. www.lvc.ele.puc-rio.br. (Accessed 28 Apr. 2008)

UN-HABITAT, 2009. *Global Report on Human Settlements 2009: Planning Sustainable Cities*, Abridged Edition, ISBN: 9781844078998.

Yu, L.; Liu, H., 2003 Feature selection for high-dimensional data: a fast correlation-based filter solution. *Proceedings of the XX International Conference on Machine Learning*, Washington DC, United States.