



Boletim de Ciências Geodésicas

ISSN: 1982-2170

Universidade Federal do Paraná

Antunes, Rodrigo Rodrigues; Bias, Edilson de Sousa; Costa,  
Gilson Alexandre Ostwald Pedro da; Brites, Ricardo Seixas  
**OBJECT-BASED ANALYSIS FOR URBAN LAND COVER MAPPING USING  
THE INTERIMAGE AND THE SIPINA FREE SOFTWARE PACKAGES**  
Boletim de Ciências Geodésicas, vol. 24, no. 1, 2018, January-March, pp. 01-17  
Universidade Federal do Paraná

DOI: 10.1590/S1982-21702018000100001

Available in: <http://www.redalyc.org/articulo.oa?id=393956630001>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in redalyc.org

UABM [redalyc.org](http://www.redalyc.org)

Scientific Information System Redalyc

Network of Scientific Journals from Latin America and the Caribbean, Spain and  
Portugal

Project academic non-profit, developed under the open access initiative

## OBJECT-BASED ANALYSIS FOR URBAN LAND COVER MAPPING USING THE INTERIMAGE AND THE SIPINA FREE SOFTWARE PACKAGES

*Análise baseada em objeto para mapeamento do uso do solo urbano utilizando os pacotes de software InterIMAGE e SIPINA*

Rodrigo Rodrigues Antunes<sup>1</sup> - ORCID: 0000-0001-9368-5077

Edilson de Souza Bias<sup>1</sup> - ORCID: 0000-0003-0334-0476

Gilson Alexandre Ostwald Pedro da Costa<sup>2</sup> - ORCID: 0000-0001-7341-9118

Ricardo Seixas Brites<sup>1</sup> - ORCID: 0000-0002-8471-8823

<sup>1</sup> Universidade de Brasília, Instituto de Geociências, Brasília, Distrito Federal, Brasil.  
[rodrigorantunes@hotmail.com](mailto:rodrigorantunes@hotmail.com); [edbias@unb.br](mailto:edbias@unb.br); [brites@unb.br](mailto:brites@unb.br)

<sup>2</sup> Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.  
[gilson.costa@ime.uerj.br](mailto:gilson.costa@ime.uerj.br)

Received in October 7<sup>th</sup>, 2016

Accepted in November 4<sup>th</sup>, 2017

### Abstract:

In this work we introduce an object-based method, applied to urban land cover mapping. The method is implemented with two open-source tools: SIPINA, a data mining software package; and InterIMAGE, an object-based image analysis system. Initially, segmentation, feature extraction and sample selection procedures are performed with InterIMAGE. In order to reduce the time and subjectivity involved to develop the decision rules in InterIMAGE, a data mining step is then carried out with SIPINA. In sequence, the decision trees delivered by SIPINA are analysed and encoded into InterIMAGE decision rules for the final classification step. Experiments were conducted using a subset of a GeoEye image, acquired in January 01, 2013, covering the urban portion of the municipality of Goianésia, Brazil. Five decision tree induction algorithms, available in SIPINA, were tested: ID3, C45, GID3, Assistant86 and CHAID. The TAU and Kappa coefficients were used to evaluate the results. The TAU values obtained were in the range of 0.66 and 0.70, while those for Kappa varied from 0.65 to 0.69.

**Keywords:** Object-Based Image Analysis, Data Mining, InterIMAGE, SIPINA.

### Resumo:

Apresentamos neste trabalho um método para o mapeamento do uso do solo urbano, implementado com duas ferramentas de código aberto: SIPINA, um pacote de software de

**How to cite this article:** Antunes, R.R. et al. Object-based analysis for urban land cover mapping using the InterIMAGE and the SIPINA free software packages. *Bulletin of Geodetic Sciences*, Vol. 24, issue 1, 1-17, Jan-Mar, 2018.



This content is licensed under a Creative Commons Attribution 4.0 International License.

mineração de dados; e o InterIMAGE, um sistema de análise de imagens de sensoriamento remoto baseado em objetos. Inicialmente procedimentos de segmentação, extração de atributos e seleção de amostras são realizados com o InterIMAGE. Com o objetivo de reduzir o tempo e a subjetividade envolvidos na definição de regras de decisão no InterIMAGE, um procedimento de mineração de dados é então realizado com a SIPINA. Na sequência, as árvores de decisão geradas através do SIPINA são analisadas e codificadas em regras de decisão do InterIMAGE para o procedimento final de classificação. Experimentos foram realizados sobre uma imagem GeoEye, recobrindo uma paisagem urbana do município de Goianésia, Brasil. Foram testados cinco algoritmos de indução de árvores de decisão disponíveis no SIPINA: ID3, C45, GID3, Assistant86 e CHAID. Os resultados foram avaliados através dos índices TAU e Kappa. Os valores de TAU obtidos variaram entre 0.66 e 0.70, e os valores de Kappa variaram entre 0.65 e 0.69.

**Palavras-chave:** Análise de Imagem Baseada em Objetos; Mineração de Dados; InterIMAGE, SIPINA.

## 1. Introduction

A great challenge is posed to urban planners by the fact that intra-urban occupation has been advancing lately, notably in developing countries (Antunes et al. 2016). Small, medium or large size cities need up-to-date data and automated tools to monitor and regulate urban expansion in order to ensure quick and consistent solutions towards efficient urban planning.

According to Cerqueira and Alves (2010), the number of remote sensing applications for urban environments has increased over the last decade, resulting in advances in large scale mapping, which is an extremely useful tool for urban planning and to manage the unregulated growth of urban areas, noticeably in developing countries.

Advanced high spatial and spectral resolution sensors and the use of Object-Based Image Analysis (OBIA) provide important means for the identification of urban targets (Blaschke 2010). OBIA can be regarded as an improvement of traditional pixel-based analysis techniques, especially when applied to high resolution spatial imagery. It defines image segments as analysis units, which can be characterized by a large number of spectral, morphological and topological features (Blaschke and Tomljenović 2012). According to Francisco and Almeida (2012), a pixel does not meet the conceptual requirements of an “object” according to the OBIA paradigm, as does the segment, which can be characterized in such a way that it can conform to an interpretation model.

A considerable number of OBIA techniques has been successfully applied to urban planning, as demonstrated in the following examples.

Chen and Chen (2014) used WorldView-2 images for detecting changes in urban monitoring. Their results indicated that the object-based methodology significantly improved change detection accuracy as compared to pixel-based techniques. The global accuracy was close to 0.89, and the Kappa coefficient reached 0.65. Bias et al. (2014) used OBIA to evaluate the urban cadastre of Goianésia, Brazil. The authors used the InterIMAGE system together with the WEKA data mining package. The accuracy of the interpretation resulted in a Tau coefficient of 82.6. Orlando and La Rosa (2014) devised an object-based classification method to detect and analyze multi-temporal remote sensing data from Scopello, Italy. Using the eCognition software, the method reached an accuracy of 0.94 (kappa) in the detection of some of the classes of interest.

Due to the large variety of available features, especially in urban environments, object-based classification models, however, tend to be fairly complex and difficult to be designed solely based on empirical evidence or prior knowledge. According to Fayyad, Piatetsky-Shapiro and Smyth (1996), Knowledge Discovery in Databases (KDD) is the global process of discovering knowledge from data, and data mining is a specific step in the identification of patterns in the available data. Data mining techniques can, thus, be very helpful in the definition of interpretation models, making it possible to exploit the vast unequal potential of object features and to gain knowledge about specific characteristics of classes of objects.

The objective of this paper is to jointly use SIPINA and InterIMAGE, both free and open-source software packages, for the urban land cover object-based classification of remotely sensed high spatial resolution data. The SIPINA contains implementations of various supervised learning algorithms, enabling interactive and visual construction of decision trees (Rakotomalala 2008).

In this study we investigated the algorithms currently available in SIPINA (ID3, C45, GID3, ASSISTANT 86 and CHAID) and employed them in the design of classification models in InterIMAGE, an open-source, knowledge-based framework for automatic image interpretation.

## 2. Study Area, Materials and Methods

The study area is the municipality of Goianésia (Figure 1) in the State of Goiás, Brazil, located 168 km from Goiânia, the State's capital.

A pansharpened GeoEye-1 image acquired in 2013, covering the urban area of Goianésia, was used in this study. The image has a spatial resolution of 0.41 m in the panchromatic band, and of 1.65 m in the multispectral bands (blue, green, red and near infrared).

The following open-source software packages were used in the study: QuantumGIS, version 2.10.1; SIPINA, version 3.12; and InterIMAGE, version 1.43.

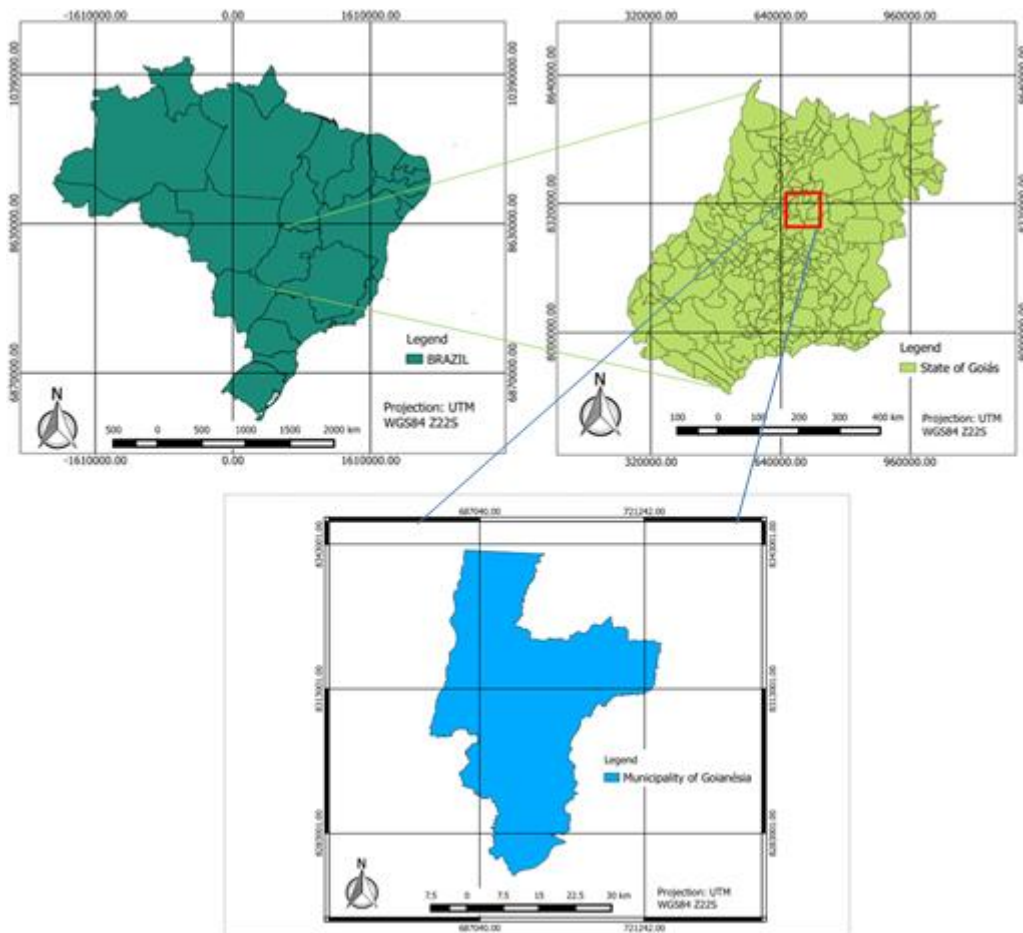
QuantumGIS is a general purpose geoprocessing software, which contains tools for handling georeferenced images and vector data (QGIS Brasil 2015).

InterIMAGE was developed by researchers from the Catholic University of Rio de Janeiro (PUC-Rio) and from the Brazilian Space Research Institute (INPE), and encompass a set of methods for the design and implementation of object-based interpretation models (Costa et al. 2010).

SIPINA was developed at University of Lyon, France, and contains a set of specialized Classification Trees induction algorithms. The first version was distributed in 1995 (Kaur and Singh 2013). It contains implementations of various supervised learning methods, enabling interactive and visual construction of classification trees (Rakotomalala 2008).

As mentioned before, in this study we investigated a set of methods available in SIPINA: ID3, C45, GID3, ASSISTANT 86, and CHAID.

The ID3 algorithm (Iterative DiChaudomiser 3) was originally developed by Quinlan (1986) at the University of Sydney, Australia. The algorithm selects classification attributes for a decision tree based on entropy information and information gain. Entropy from Information Theory (the impurity of the attribute) is used to measure the information gain of an attribute. The information gain refers to the type of impurity. The lower the entropy value, the less uncertainty and more utility the pre-classified product has (Wilges et al. 2010).



**Figure 1:** Study area: Municipality of Goianésia, Goiás.

According to Hssina et al. (2014), the C4.5 algorithm was proposed in 1993, also by Quinlan, to overcome the limitations of ID3, such as the sensitivity of resources in face of a high number of feature values.

The algorithm GID3 is a generalization of ID3 and C4.5, in which some leaves of the separation process may mix together. The idea is to highlight the more interesting leaves and merge the others into a “standard page” (Fayyad 1994).

The Assistant 86 is another enhancement of ID3. In it there is a criteria set for improving the information gain and the various parameters which can control the size of the tree (Cestnik et al. 1987).

CHAID is an enhanced version of Morgan and Sonquist’s (1963) AID algorithm. Kass (1980) explains the particularities of CHAID, such as the use of Chi-square statistics for division criteria, and fusing some pages into a single node.

At this point, it is important to note that the InterIMAGE package also includes a supervised classification operator that implements the C4.5 algorithm. Assuming that the implementation of the algorithm in InterIMAGE and in SPINA are similar, thus producing equivalent results, and bearing in mind that this particular study investigates the integration of the two open-source packages, we decided not to include it in a comparative evaluation, considering the C4.5 implementation built in InterIMAGE.

### 3. Methodology

The devised methodology has five steps: pre-processing; pre-classification; data mining; classification; and results analysis (Figure 2). The following subsections describe those steps.

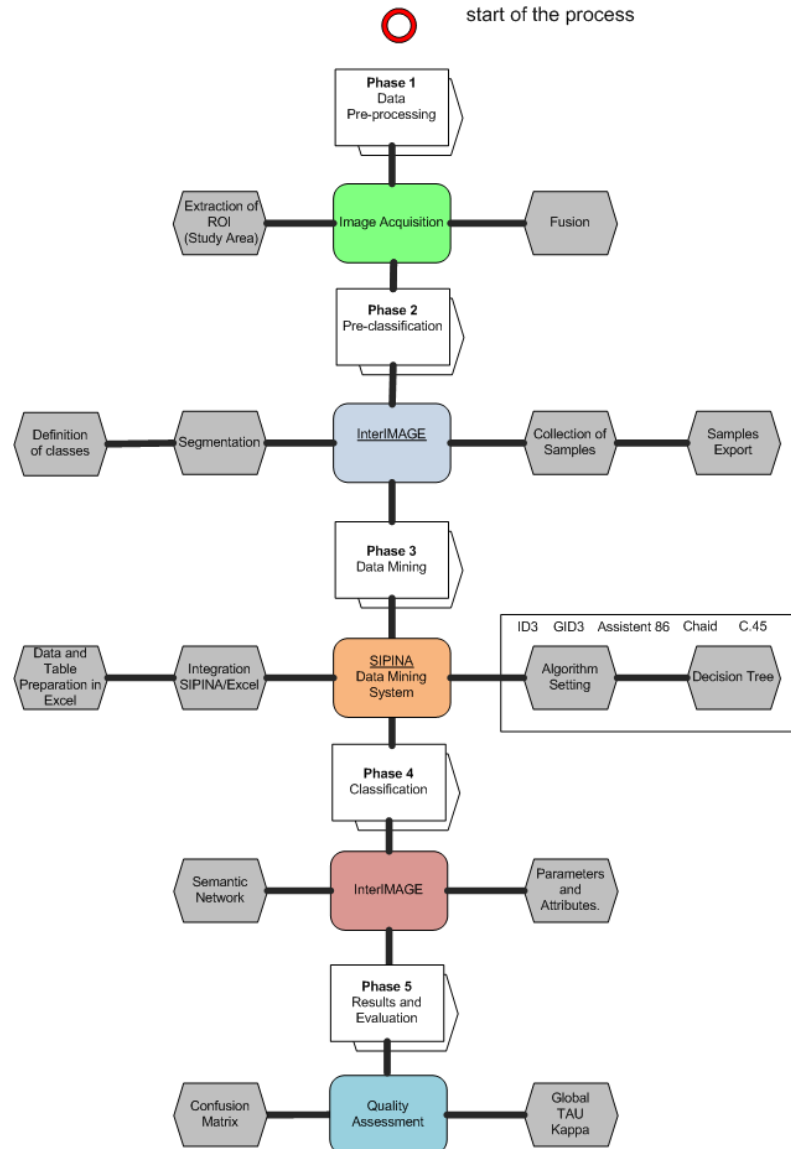


Figure 2: Methodological steps.

#### 3.1 Step 1: Data Pre-processing

In this work, we used a GeoEye-1 sensor image, from 2013, covering the city of Goianésia, Goiás, Brazil. The image was pansharpened and the ROI (Region of Interest) corresponding to the borders of the study area was extracted from the image.

### 3.2 Step 2: Pre-classification

Nine target classes were defined for the interpretation task: metallic roof; asbestos roof; ceramic roof (clear and dark); swimming pools; vegetation; bare soil; concrete pavement; and shadow.

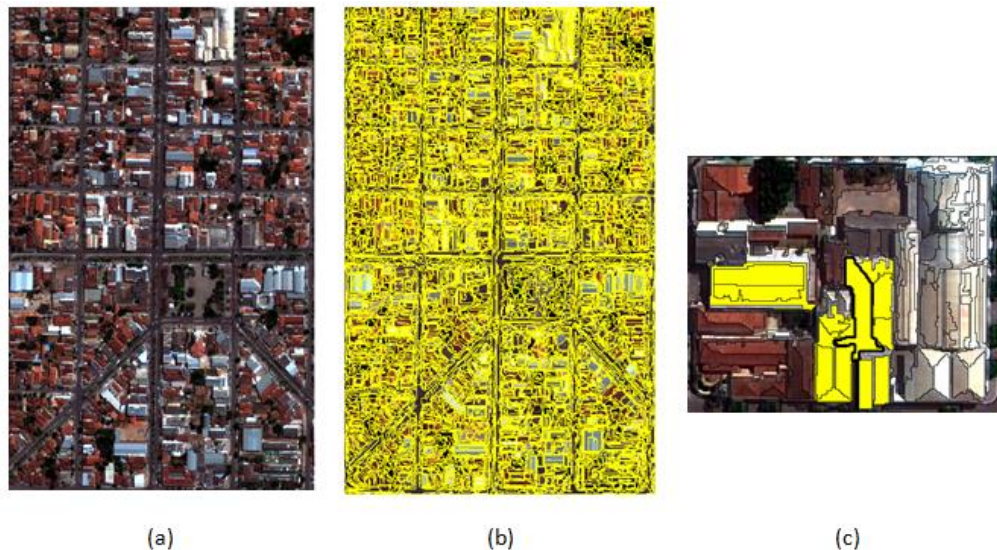
After the definition of the classes of interest, image segmentation was performed in InterIMAGE, using the algorithm proposed by Baatz and Schäpe (2000). According to (Ferreira et al. 2013) the quality of the segmentation produced by that algorithm improves when the heterogeneity criteria that govern the growth of regions (segments) takes into account morphological attributes in addition to the spectral ones. According to the same authors, the quality gain can significantly depend on the characteristic shape of the objects of a particular class. In this work, therefore, the segmentation procedure was specialized for the different classes of interest, thus producing different segmentation outcomes. The segmentation parameters values used for the different classes are detailed in Table 1.

**Table 1:** Parameters used in image segmentation.

Classes	Input Band	Relative Band Weight	Compactness	Color	Scale
Metal roofing	0,1,2,3	1,1,1,1	0.8	0.4	90
Asbestos roofing	0,1,2,3	1,1,1,1	0.7	0.5	80
Clear ceramics roofing	0,1,2,3	1,1,1,1	0.5	0.5	80
Dark ceramics roofing	0,1,2,3	1,1,1,1	0.5	0.5	70
Swimming pools	0,1,2,3	1,1,1,1	0.8	0.4	60
Bare soil	0,1,2,3	1,1,1,1	0.5	0.5	60
Concrete pavement	0,1,2,3	1,1,1,1	0.5	0.5	60

After image segmentation, samples from each class were collected using the InterIMAGE's Sample Editor tool. Figure 3 shows the original image (a) and the corresponding segmentation for the metallic roof class (b), respectively. Examples of metallic roofs' sample objects are shown in (c).

In order to properly deal with the large variety of colors corresponding to metallic and asbestos roofs and to select samples for data mining (Step 3), these classes were divided into sub-classes: asbestos; asbestos\_1; metallic; metallic\_1; metallic\_2; and metallic\_3. The sub-classes were regrouped in the classification step (Step 4).



**Figure 3:** Original image (a). Segmentation for metallic roofs (b). Samples of metallic roofs (c).

Following Antunes et al. (2014), image segmentation for the vegetation and shadow classes was performed using a per-pixel thresholding segmentation tool of InterIMAGE. The vegetation class thresholding was carried out using the Normalized Difference Vegetation Index (NDVI) feature. All contiguous pixels with NDVI above 0.65 were enclosed in the corresponding vegetation segments. The segmentation for the shadow class was based on pixel brightness values: brightness above 30 and below 54.

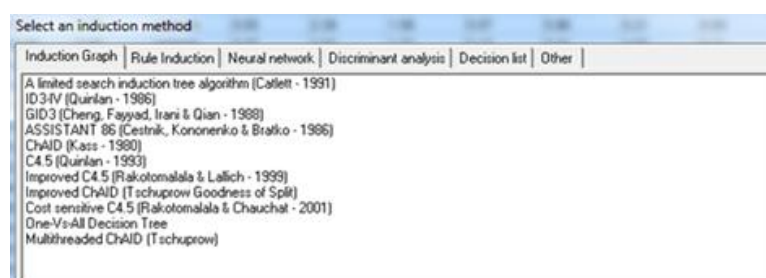
After segmentation, the following features were calculated for each segment (from the different segmentations):

- a) Spectral features: compacity, brightness, entropy, maxpixelvalue, mean, minpixelvalue, ratio and bandmeandiv;
- b) Morphological features: angle, squareness and circleness.

These attributes were used in the data mining process (Step 3).

### 3.3 Step 3: Data Mining

All segments' features were exported by InterIMAGE in shapefile format and then the corresponding data file was opened in Microsoft Excel with the SIPINA add-in (SIPINA.XLA) previously installed (Figure 4).



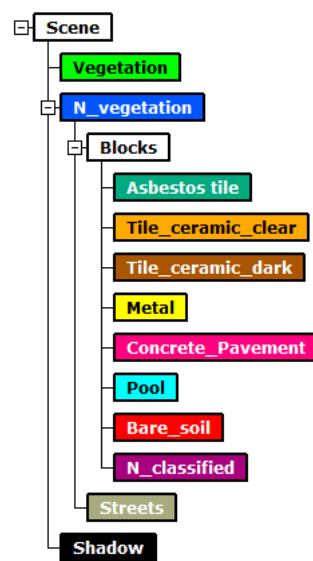
**Figure 4:** Induction algorithms for SIPINA decision trees.



Then, decision trees were created using the methods ID3, C45, GID3, ASSISTANT 86, and CHAID.

### 3.4 Step 4: Classification

InterIMAGE was used for classification. An interpretation model in InterIMAGE contains information used by its control process to interpret a scene. It is represented by a semantic network in which the nodes are associated with classes of objects and are organized in a hierarchical fashion (Costa et al. 2010). The semantic network designed for this work is shown in Figure 5. The operators and respectively used parameters are described in Bias et al. (2014).



**Figure 5:** Semantic network with defined classes.

The TopDown Decision Rule tool in InterIMAGE supports the creation of a set of expressions called decision rules. These expressions represent structured and specific knowledge used by the system in the interpretation (Costa et al. 2010). In this work, the decision rules associated to each semantic network node were based on the decision trees generated automatically by SIPINA.

SIPINA is a specific data miner software for decision tree classification. The threshold values and the decision rules are defined by decision tree induction algorithms. According to Tedesco et al. (2014), the main objective of decision tree algorithms is to find the smallest possible decision tree, coherent with the training samples, achieving the correct classification with a small number of tests.

The automatically generated decision tree allows the analyst to inspect and study the tree structure (classes, values, attributes, and rules). In this way, it is possible for the analyst to identify in the tree the absence of attributes or classes due to their irrelevance (not useful in the classification process), or even due to operational error (when a given attribute or class is accidentally left apart).

### 3.5 Step 5: Analysis of the results

Five classification processes were conducted, based on the different decision tree induction algorithms in SIPINA.

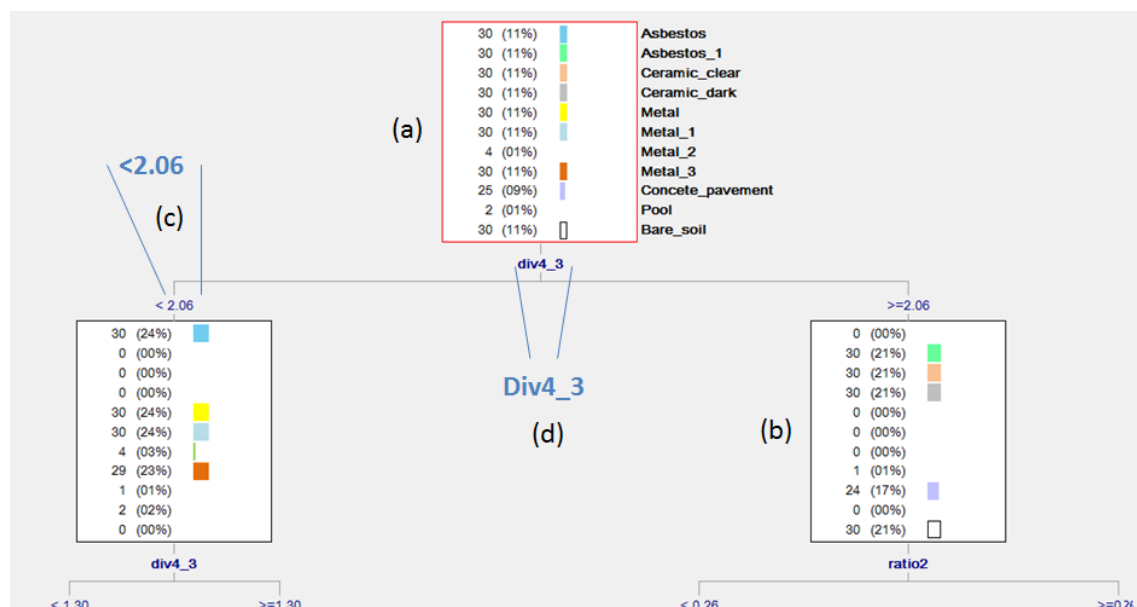
The accuracy analysis was conducted on a test area based on the following evaluations established by Bias et al. (2014): measuring the number of samples; random distribution of check points; visual investigation; composition of the confusion matrix; calculation of the global accuracy, and the TAU and Kappa agreement coefficients. The number of samples was calculated by multinomial distribution. The sample unit for accuracy assessment was a pixel.

The number of samples was determined according to Congalton and Green (1999), which was the same, used for the same area, as in Bias et al. (2014) and Antunes et al. (2014).

Afterwards, randomly generated check points (pixels) were determined for accuracy assessment. Each of the check points was visually inspected and assigned to its correspondent class. The automatic classification result was then compared to the visually assigned class for the construction of confusion matrixes and for calculating the global accuracy, TAU and Kappa agreement coefficients. The results are presented and discussed below.

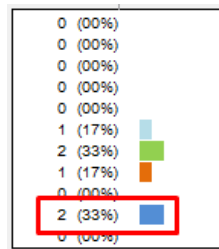
The basic configuration of SIPINA was maintained for the decision tree induction algorithms, without altering any of the standard parameters.

Figure 6 shows a decision tree generated using the algorithm ID3. The tree has 31 nodes, 16 leaves and a maximum depth of 6. Visualizing the tree eased the analysis of each class (Figure 6a). Each tree node shows a confidence percentage and the number of collected samples, seen in Figure 6b. Figures 6c and 6d show the values for thresholds and features, respectively. The threshold values show the separation between two classes and the features on the tree are the ones SIPINA determined as having the best values.



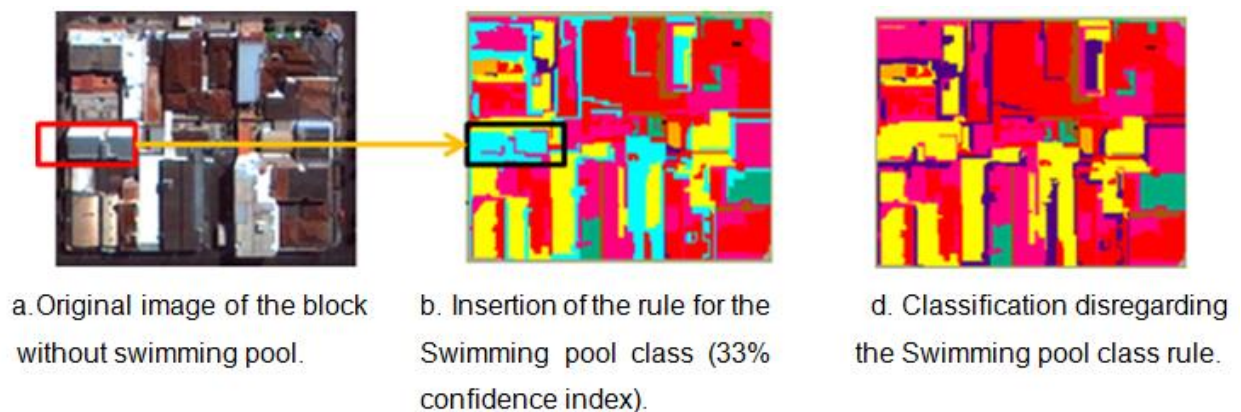
**Figure 6:** Decision tree with decision rules generated by SIPINA using algorithm ID3. (a) Tree nodes with classes. (b) Confidence percentages and class samples. (c) Threshold value. (d) Attribute.

A confidence index of 50% was established for the rules contained in the leaves and chosen for insertion into InterIMAGE, as the rules with a low confidence index tend to reduce classification accuracy. According to Goldschmidt and Passos (2005), the measure of confidence expresses the quality of a rule. The example here is for the swimming pool class: the confidence index of the rule in the leaf reached 33%, as seen in Figure 7. The figure also shows not-null confidence indices for other classes (metallic\_1, metallic\_2 and metallic\_3). The rules associated with this leaf were ignored (not inserted in InterIMAGE) due to their low confidence index.



**Figure 7:** Four classes (Metallic\_1, Metallic\_2, Metallic\_3 and Swimming pool) with the same rule in the leaf and confidence ranging from 17% to 33%.

The swimming pool class rule, with a 33% confidence index, was inserted in InterIMAGE only to illustrate the confusion process generated in the classification. Figure 8 shows the questionable swimming pool classification generated by this rule. In (a) the original image of the block is shown without any Swimming pools. In (b) one can see the inaccurate classification of some objects as swimming pool (in cyan), with roofs classified as such. (c). Lastly, (d) shows the same classification with the swimming pool class rule discarded.



**Figure 8:** Result of a classification with a 33% confidence rule defined in SIPINA and executed in InterIMAGE.

Table 2 shows an analysis of the decision tree based on the best rules using algorithm ID3. The rule for identifying the metallic\_3 class was the most complex one, depending on three different criteria. There were also rejected rules for this class (two), which is related to those presenting a confidence index under 50%. The same difficulty was observed for the other tested algorithms.

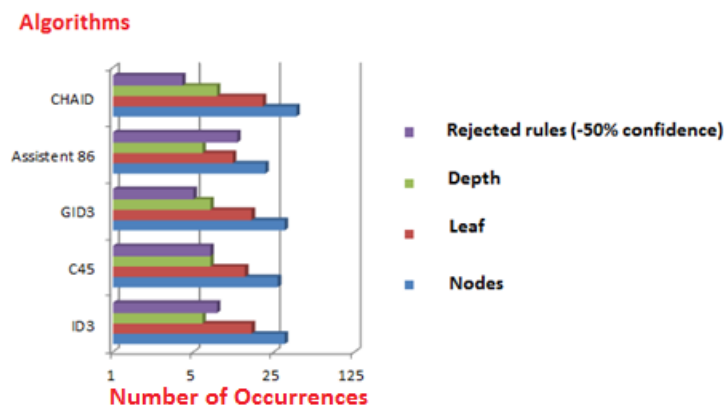
**Table 2:** Basic decision tree statistics with algorithm ID3.

Classes	Samples	Best rule	Confidence (Leaf)	Rejected rules (- 50% confidence)
Asbestos roof	30	IF div4_3 < 2.06 and ratio3 >=0.20 and div4_1 >=2.00 and brightness < 993.93	100%	NÃO
Asbestos roof_1	30	IF div4_3 >=2.06 and ratio2 >=0.26 and mean1 >=266.58 and ratio3 < 0.15 and minpixel1 >=226.00	100%	NÃO
Ceramic_clear	30	IF div4_3 >=2.06 and ratio2 < 0.26 and mean3 >=317.63 and ratio2 < 0.23	100%	SIM (40%)
Ceramic_dark	30	IF div4_3 >=2.06 and ratio2 >=0.26 and mean1 < 266.58	100%	NÃO
Metallic	30	IF div4_3 < 2.06 and ratio3 >=0.20 and div4_1 < 2.00	100%	NÃO
Metallic_1	30	IF div4_3 < 2.06 and ratio3 < 0.20 and ratio1 >=0.22 and compacity < 0.07	100%	SIM (17%)
Metallica_2	4	IF div4_3 < 2.06 and ratio3 < 0.20 and ratio1 < 0.22 and brightness >=916.20	100%	SIM (33%)
Metallica_3	30	IF div4_3 < 2.06 and ratio3 < 0.20 and ratio1 < 0.22 and brightness < 916.20 and ratio3 < 0.17	50%	SIM (17% e 20%)
Pool	2	IF div4_3 < 2.06 and ratio3 < 0.20 and ratio1 >=0.22 and compacity >=0.07	33%	SIM (33%)
Bare soil	30	IF div4_3 >=2.06 and ratio2 < 0.26 and mean3 >=317.63 and ratio2 >=0.23	60%	SIM (40%)
Concrete pavement	25	IF div4_3 >=2.06 and ratio2 >=0.26 and mean1 >=266.58 and ratio3 < 0.15 and minpixel1 < 226.00	100%	SIM (40%)

The Table 3 and Figure 9 show the performance of each of the analyzed SIPINA algorithms.

**Table 3:** Performance summary of each SIPINA decision tree algorithm.

Algorithms	Nodes	Leaf	Depth	Reject rules (<50% confidence)
ID3	31	16	6	8
C45	27	14	7	7
GID3	31	16	7	5
Assistent 86	21	11	6	12
CHAID	39	20	8	4

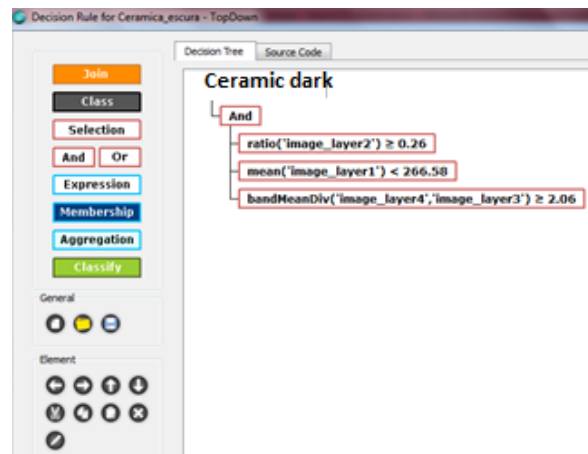
**Figure 9:** Performance summary of each SIPINA algorithm.

As Table 3 and Figure 9 show, the algorithms presented a very little variation with respect to the corresponding tree structures. The CHAID algorithm produced the highest number of nodes (39), leaves (20) and greatest structural depth (8), yet the number of rejected rules was low (4).

The tree with the most compact structure was produced by the Assistant 86 algorithm. It had 21 nodes, 11 leaves and a maximum depth of 6, although with a high number of rejected rules (12).

SIPINA did not have any difficulties processing decision trees for the data considered in this study, but its limitation of 16,384 attributes and 500,000,000 registers requires attention. This limitation has to do with the fact that the system loads the whole data set to the memory before the learning process begins (Rakotomalala 2016).

As mentioned before, the decision rules with threshold values defined by the SIPINA algorithms were inserted into InterIMAGE using the TopDown Decision Rule. This process was performed manually. Figure 10 shows an example of the rule insertion for the dark ceramic roof class.

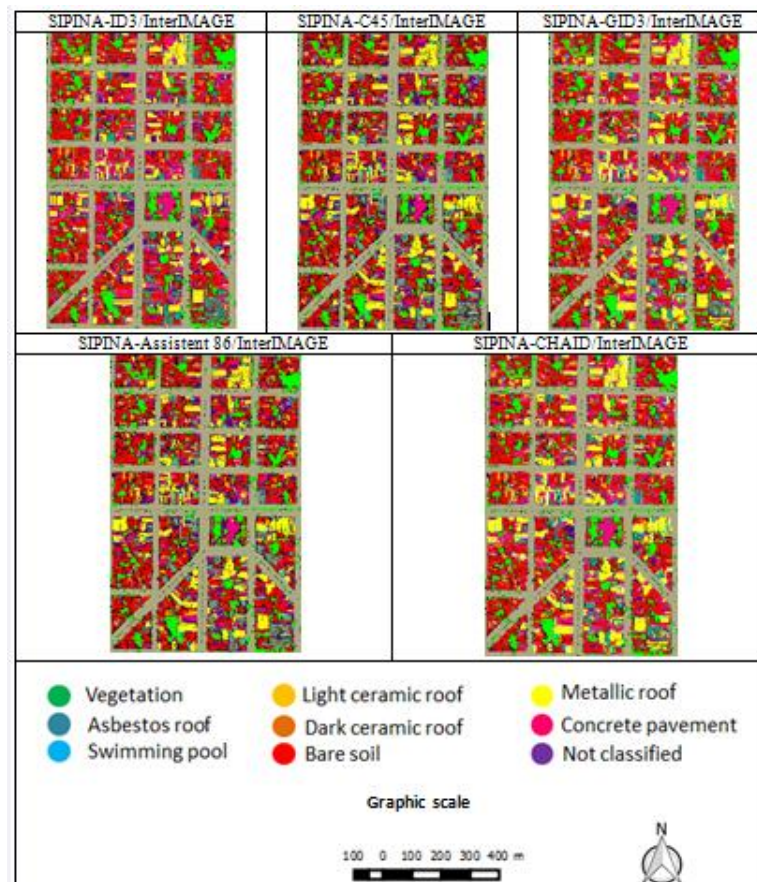


**Figure 10:** Example of a decision rule defined in SIPINA and inserted in InterIMAGE.

Image classification consists of separated segment sets that exhibit similar characteristics (e.g., spectral, morphological or textural). The classification result is a thematic map showing the geographical distribution of the classes (Tedesco et al. 2014).

Figure 11 visually shows each classification resulting from the SIPINA and InterIMAGE integration.

As expected, the confusion matrix derived coefficients, for each classification, showed greater confusion between ceramic roofs and bare soil classes, as they are composed of the same material (red clay): the spectral response also has a major influence in this respect. The vegetation and street classes were well classified, achieving a good separation, without much confusion.



**Figure 11:** Results from each classification. SIPINA-InterIMAGE integration.

The quality of each classification was calculated by means of the global accuracy, the TAU and Kappa agreement coefficients. Table 4 shows the corresponding accuracy values.

**Table 4:** Global accuracy and TAU and Kappa agreement coefficients for the obtained classification results.

Algorithm	Global accuracy (%)	TAU	Kappa
ID3	70	0.66	0.65
C45	72	0.68	0.67
GID3	73	0.70	0.69
Assistent 86	73	0.70	0.69
CHAID	73	0.70	0.69

As Table 4 shows, classification using ID3 algorithm had the worst agreement coefficients (TAU 0.66 and Kappa 0.65). The GID3, Assistant 86 and CHAID algorithms all reached the same agreement indices (TAU 0.70 and Kappa 0.69).

## 4. Conclusion

The results obtained in this study allowed us to evaluate the integrated use of the SIPINA data mining package and the InterIMAGE system in an object-based image analysis application. The investigation led to the following conclusions:

a) SIPINA proved to be an easy-to-use software, working directly in Excel spreadsheets (without the need of installing other applets) and providing ways for visual analysis of decision trees and corresponding rules confidence values associated to each tree node, thus allowing the analyst to inspect the credibility of rules for each class of interest.

b) SIPINA offers a number of algorithms for decision tree induction: ID3, C4.5, GID3, Assistant86 and CHAID. With the exception of CHAID, from Morgan and Sonquist (1963), all the other algorithms are based on the Quinlan seminal algorithm.

c) The TAU and Kappa coefficients obtained with the GID3, Assistant86 and CHAID algorithms, 0.70 and 0.69, respectively, represent satisfactory classifications, not excellent ones, however. The confusion between bare soil and ceramic roof was, for the most part, responsible for reducing the values of such coefficients. Antunes et al. (2015) used the same input data, but employed the J4.8 algorithm implemented in the WEKA package, and obtained a TAU of 0.78 for the same classification. The J4.8 algorithm is a Java implementation of the Ross Quinlan's C4.5 algorithm. In this project, using the same input data and the C4.5 algorithm, the TAU index of 0.72 was attained. The presented results are very close, showing a small difference (0.6). Several factors may influence this difference though, such as internal characteristics of the data mining software (WEKA and SIPINA), different random samplings and others.

d) Among the algorithms tested in this work, Assistant 86 was more suitable for integration with InterIMAGE (at least for this particular classification problem). It reached the same agreement coefficient as the CHAID and GID3 algorithms, but the tree structure was more compact, having fewer nodes, leaves and depth. This means there were fewer rules inserted in InterIMAGE for classification, implying that the integration does not require much effort on the part of the analyst. Additionally, there is a smaller chance of error in its operation and greater flexibility to interpret the rules translated into InterIMAGE.

## ACKNOWLEDGEMENT

The authors acknowledge the support provided by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

## REFERENCES

- Antunes, R. R. et al. 2014. *Desenvolvimento de técnica para monitoramento do cadastro urbano baseado na classificação orientada a objetos. Estudo de caso: Município de Goianésia, Goiás*. Revista Brasileira de Cartografia, 67/2, pp.357-372.
- Antunes, R. R. et al. 2016. *Integration of open-source tools for object-based monitoring of urban targets*. GEOBIA 2016: Solutions and Synergies. Enschede, The Netherlands. University of Twente Faculty of Geo-Information and Earth Observation (ITC).
- Baatz, M and Schäpe, A. 2000. *Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation*. In: XII Angewandte Geographische Informationsverarbeitung, AGIT Symposium. Proceedings. Karlsruhe, Germany: Herbert Wichmann Verlag, Salzburg, Áustria, pp. 12-23.
- Bias, E.S. et al. 2014. *Application of Imagery Analysis Based on Objects as a Tool for Monitoring the Urban Cadastre in Small Municipalities*. International Geographic Object-Based Image Analysis Conference, Thessaloniki, Greece, pp. 15-20.
- Blaschke, T. 2010. *Object based image analysis for remote sensing*. Journal of Photogrammetry and Remote Sensing, Falls Church, v. 65, n. 1, pp. 2–16.
- Blaschke, T. and Tomljenovic, I. 2012. *LidarScapes and OBIA*. In Proceedings of the ASPRS. Annual Conference, Sacramento, CA, USA, pp. 19–23.
- Cerqueira, J. A. C. and Alves, A. 2010. *Classificação de imagens de alta resolução espacial para o mapeamento do tipo de pavimento urbano*. III Simpósio Brasileiro de Ciências Geodésica e Tecnologias da Geoinformação. Recife, PE, Brasil.
- Cestnik, B. Kononenko I. and Bratko I. 1987. "ASSISTANT 86: A Knowledge Elicitation Tool for Sophistical Users". Proc. of the 2nd European Working Session on Learning, pp.31-45.
- Chen, Q. and Chen, Y. 2014. *Object-based Change Detection of WorldView-2 data for Urban Dynamic Monitoring*. South-Eastern European Journal of Earth Observation and Geomatics. Aristotle University of Thessaloniki, Greece, pp. 41-46.



- Congalton, R. G. and Green, K. 1999. *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton-USA: Lewis Publisher.
- Costa, G.A.O.P. et al. 2010. *Knowledge-based Interpretation of Remote Sensing Data With the InterIMAGE System: Major Characteristics and Recent Developments. Proceedings of the 3rd GEOBIA*. Gent, Belgium.
- Fayyad, U. Piatetsky-Shapiro, G. and Smyth, P. 1996. *From Data Mining to Knowledge Discovery in Databases*. American Association for Artificial Intelligence. AI magazine: AI magazine, v. 17, n. 3, pp. 37.
- Fayyad, U. M. 1994. *Branching on attribute values in decision tree generation*. California Institute of Technology. In: AAAI (www.aaai.org), pp. 601-606.
- Ferreira, R.S. Costa, G.A.O.P. and Feitosa, R.Q. 2013. *Avaliação de critérios de heterogeneidade baseados em atributos morfológicos para segmentação de imagens por crescimento de regiões*. Boletim de Ciências Geodésicas, v. 19, n 3, pp.452-471.
- Francisco, C.N., Almeida, C.M. *Avaliação de desempenho de atributos estatísticos e texturais em uma classificação de cobertura da terra baseada em objeto*. Boletim. Ciências. Geodésicas. vol.18 no.2 Curitiba, Paraná, Brazil, 2012.
- Goldschmidt, R. and Passos, E. 2005. *Data Mining: Um Guia Prático*. Elsevier. Rio de Janeiro, Brasil.
- Hssina, B. et al. 2014. *A comparative study of decision tree ID3 and C4.5*. International Journal of Advanced Computer Science and Applications, v. 4, n. 2.
- Kaur, A. and Singh, S. 2013. *Classification and Selection of Best Saving Service for Potential Investors using Decision Tree – Data Mining Algorithms*. International Journal of Engineering and Advanced Technology (IJEAT), pp. 80-82, 2013.
- Kass G. 1980. *An exploratory technique for investigating large quantities of categorical data*. Applied Statistics, 29(2), pp. 119-127.
- Morgan, J. N. and Sonquist, J. A. 1963. *Problems in the analysis of survey data, and a proposal*. Journal of the American Statistical Association 58, 415-434.
- Orlando, P. and La Rosa, E. 2014. *Object oriented methodology for change detection technique: the case of Scopello-Silicy*. South-Eastern European Journal of Earth Observation and Geomatics. Aristotle University of Thessaloniki, Greece, pp. 65-68.
- QGIS Brasil, 2015. Comunidade de usuários QGIS Brasil. [online]. Available at: <<http://qgisbrasil.org/>>. [Accessed on: 08/11/2015].
- Quilan, J. 1986. *"Induction of Decision Trees"*, in Machine Learning, pp.81-106.
- Rakotomalala, R. 2008. *Introduction of a Decision Tree using SIPINA*. Tutorial. Departamento de Informática e Estatística. University Lyon, France.
- Rakotomalala, R. 2016. *SIPINA Overview*. Departamento de Informática e Estatística. University Lyon, France. Available at: <<http://eric.univ-lyon2.fr/~ricco/sipina.html>>. [Accessed on: 23/02/2016].
- Wilges, B. et al. 2010. Bastos, R. *Avaliação da aprendizagem por meio de lógica de fuzzy validado por uma Árvore de Decisão ID3*. Novas Tecnologias na Educação. Centro interdisciplinar de novas tecnologias na educação – CINTED. Universidade Federal do Rio Grande do Sul – UFRGS. v. 8, n 3.

Tedesco, A., Antunes, A. F. B. and Oliani, L. O. 2014. *Detecção de formação erosiva (voçoroca) por meio de classificação hierárquica e por árvore de decisão*. Boletim Ciências Geodésicas, v. 20, n. 4, pp.1005-1026.