
CONAE

Work Plan: Advanced classification techniques for urban areas using optical and SAR images

Fabricio Herrera
Paolo Gamba
Anabella Ferral
Gabriel Platzeck

Maestría en Aplicaciones de Información Espacial
Instituto de Altos Estudios Espaciales “Mario Gulich” –
CONAE/UNC

Introduction

The increase of the global population affects the environment through the growth of population in urban zones and the corresponding increase of natural resource consumption. The precise monitoring of these changes is fundamental for the implementation of effective decision making processes.

Methodology

Since the goals of the work is procude update urban maps using VHR optical and SAR data, fusion methodologies aimed at classification will be considered and compared. To this aim, parametric and non-parametric approaches will be considered, with particular stress on deep learning, Hierarchical Binary Decision Trees, multi-kernel SVM, and other update machine learning algorithms. The target is to use the full potential of multitemporal SAR and optical data sets by using SPOT and COSMO-SkyMed images, thanks to agreements in place between CNR, ASI and CONAE. In addition to improve the classification algorithms currently used to analyze heterogeneous multitemporal data sets, this research will also be important to understand whether it will be possible and viable to use for urban mapping purpose the images provided by the prototype UAV (SARA) in development by INVAP.

Schedule

February: Land use classification of the Capital Department of the city of Córdoba based on data from SPOT 4 (March-2012) and SPOT 5 (March-2015). Classes: water, urban, crops, bare soil, vegetation.

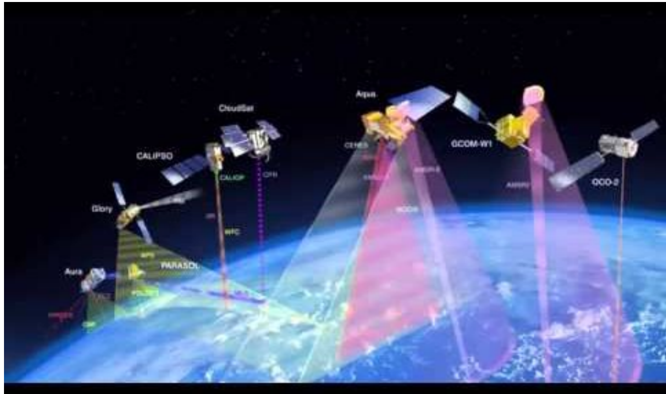
March: Description of results and table with the comparison of methods and elaboration of land use change map.

April: Optical Image Analysis (2012 and 2015). Add methods: Herarchical binary decision three, multi-kernel SVM.

May: SAR COSMO Add methods: Herarchical binary decision three, multi-kernel SVM.

June: Comparison of Optical and Radar methods.

July: Elaboration of final report.



Image/Sensor: Satellite data

<u>Study zone</u>: City of Cordoba, Argentina.

<u>Optical images:</u> Data from SPOT 4 ,5 and 6 will be used in order to classify built up areas in Córdoba city for three different dates. Table 1 shows three images from Córdoba city which were download from CONAE catalogue.

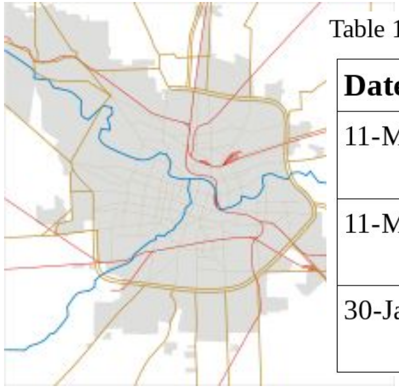
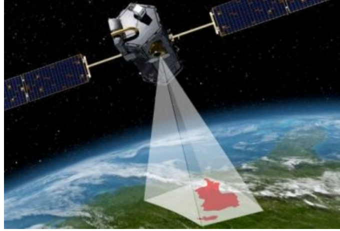


Table 1: Optical images to be used for Córdoba city classification

Date	Source	Product name	Resolution	Place
11-Mar-2012	SPOT4	SPOT4_HRVIR1_2012-03-11_13-58-44_M+I_685_413	10 m	Córdoba
11-Mar-2015	SPOT5	SPOT5_HRG1_2015-03-11_13-12-44_J_684_413_S0_L2A	10 m	Córdoba
30-Jan-2017	SPOT6	SPOT6_20170130_1353122_BUNDLE_W064S32_L2A_16JR_023x051_32720	10 m	Cordoba

Image/Sensor: Satellite data

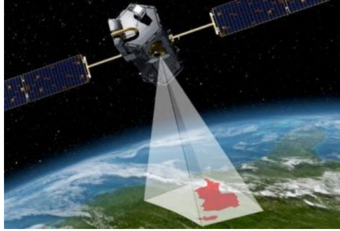


Cosmo-SkyMed

Request and reception of the images Cosmo SkyMed of the city of Cordoba. Between the years 2010 to 2017. The acquisition mode is STRIPMAP.

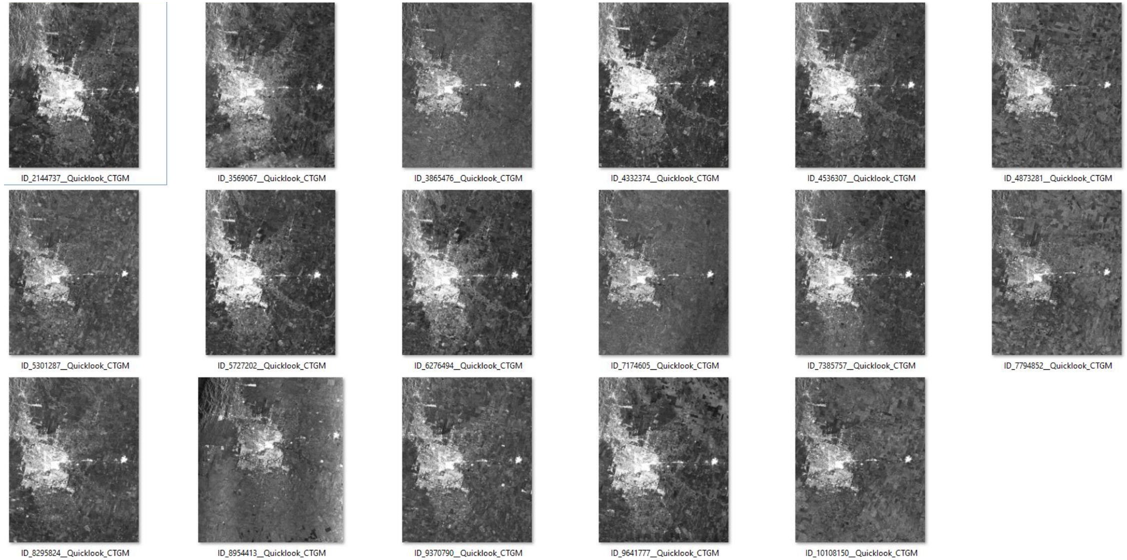
Number	Quick Look ID	Satellite ID	Record Number (Imágenes de archivo)	Fechas	Zona	Modo	Ang.	Polariz.	Dir. Orb.	Enfoque	Nivel
1	10108150	Cosmo-SkyMed 4	101599274	2017-01-31	Córdoba	STR_HIMAGE		HH	Ascending		SCS
2	9641777	Cosmo-SkyMed 1	101491794	2016-09-29	Córdoba	STR_HIMAGE		HH	Ascending		SCS
3	9370790	Cosmo-SkyMed 4	101442821	**2016-07-07**	Córdoba	STR_HIMAGE		HH	Ascending		SCS
4	8954413	Cosmo-SkyMed 2	101366604	2016-03-01	Córdoba	STR_HIMAGE		HH	Ascending		SCS
5	8295824	Cosmo-SkyMed 4	101246393	**2015-07-05**	Córdoba	STR_HIMAGE		HH	Ascending		SCS
6	7794852	Cosmo-SkyMed 4	101160788	2015-01-10	Córdoba	STR_HIMAGE		HH	Ascending		SCS
7	7385757	Cosmo-SkyMed 1	101123974	2014-09-08	Córdoba	STR_HIMAGE		HH	Ascending		SCS
8	7174605	Cosmo-SkyMed 2	101104456	**2014-07-14**	Córdoba	STR_HIMAGE		HH	Ascending		SCS
9	6276494	Cosmo-SkyMed 1	101019051	2013-11-24	Córdoba	STR_HIMAGE		HH	Ascending		SCS
10	5727202	Cosmo-SkyMed 1	100962592	**2013-07-19**	Córdoba	STR_HIMAGE		HH	Ascending		SCS
11	5301287	Cosmo-SkyMed 1	100916466	2013-03-29	Córdoba	STR_HIMAGE		HH	Ascending		SCS
12	4873281	Cosmo-SkyMed 4	100881967	2013-01-04	Córdoba	STR_HIMAGE		HH	Ascending		SCS
13	4536307	Cosmo-SkyMed 4	100845486	2012-09-30	Córdoba	STR_HIMAGE		HH	Ascending		SCS
14	4332374	Cosmo-SkyMed 4	100822086	**2012-07-28**	Córdoba	STR_HIMAGE		HH	Ascending		SCS
15	3865476	Cosmo-SkyMed 4	100770448	2012-03-06	Córdoba	STR_HIMAGE		HH	Ascending		SCS
16	3569067	Cosmo-SkyMed 2	100738424	2011-11-27	Córdoba	STR_HIMAGE		HH	Ascending		SCS
17	2144737	Cosmo-SkyMed 2	100626584	**2010-07-14**	Córdoba	STR_HIMAGE		HH	Ascending		SCS

Image/Sensor: Satellite data

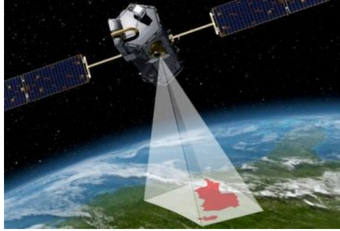


Cosmo-SkyMed

Request and reception of the images Cosmo SkyMed of the city of Cordoba. Between the years 2010 to 2017. The acquisition mode is STRIPMAP.



Image/Sensor: Satellite data



SPOT 4, 5 & 6

Años 2012 - 2015 - 2017.

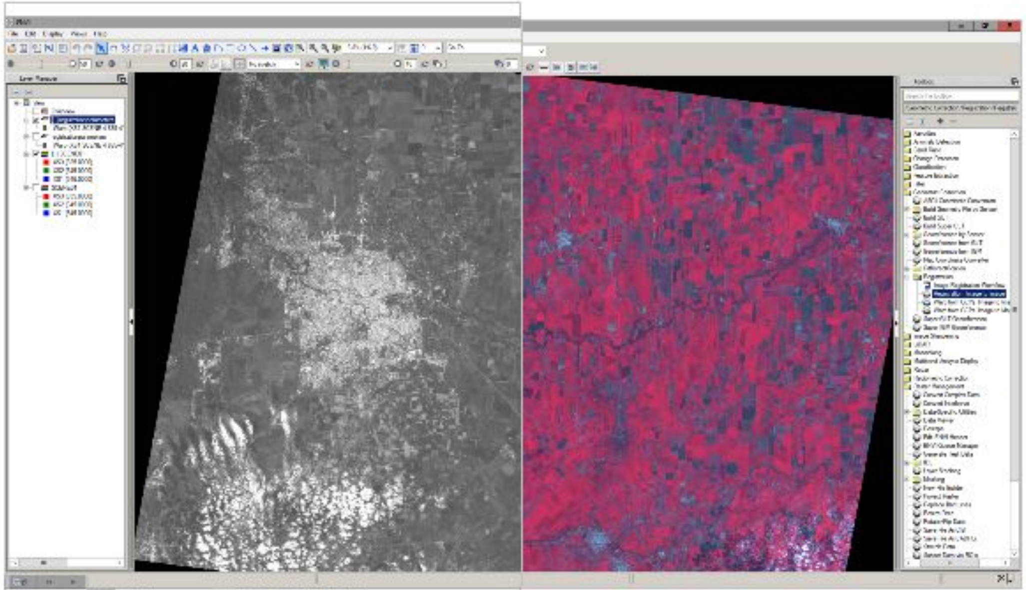
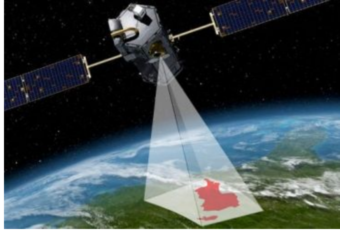
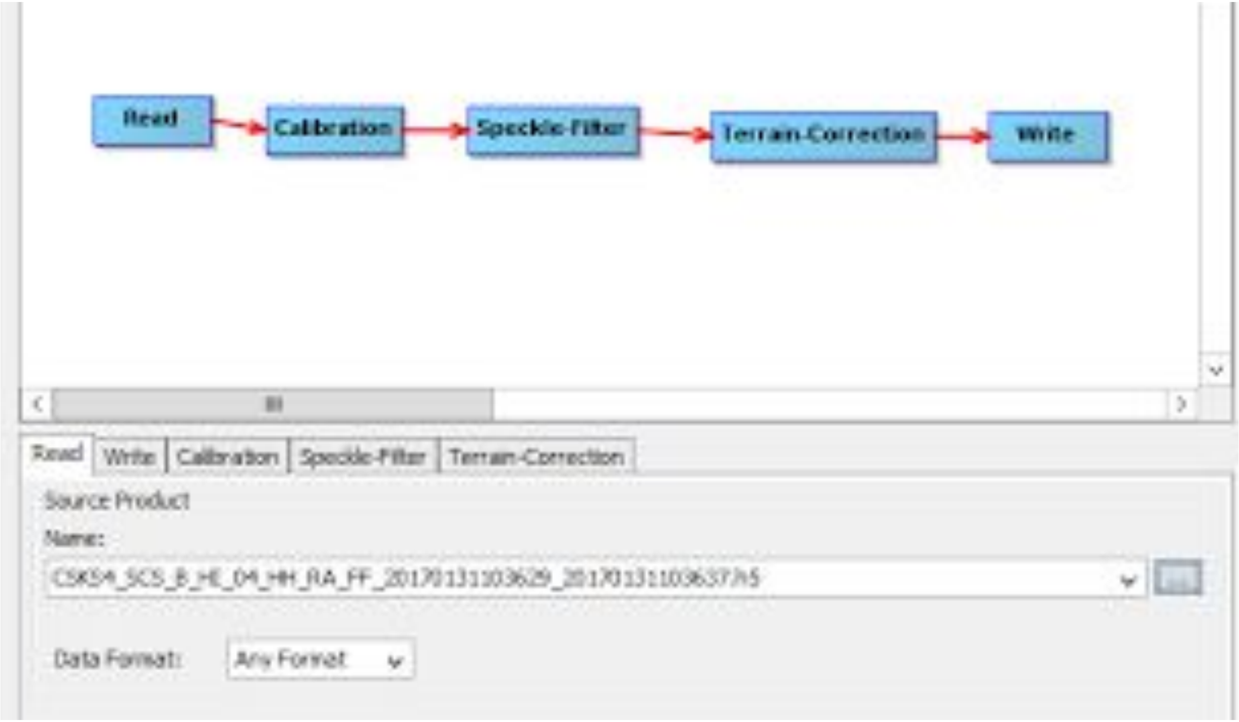


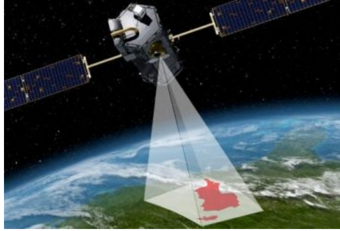
Figure 10: Images SPOT 4 & 5



Image/Sensor: Preprocessing

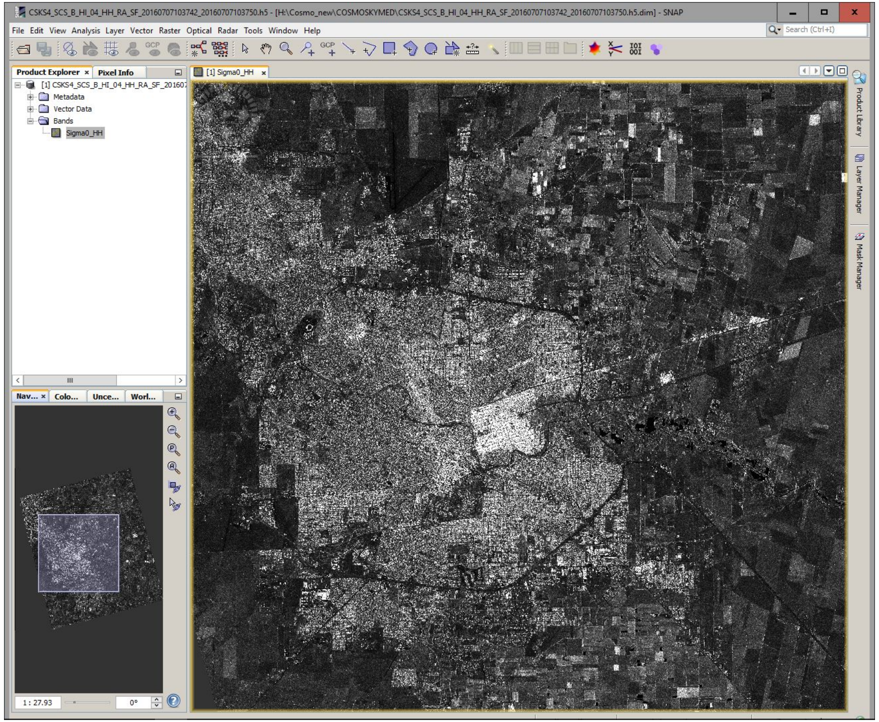
Process and steps to make the correction and calibration of SAR image in the SNAP software.



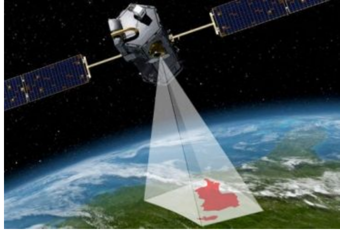


Image/Sensor: Preprocessing

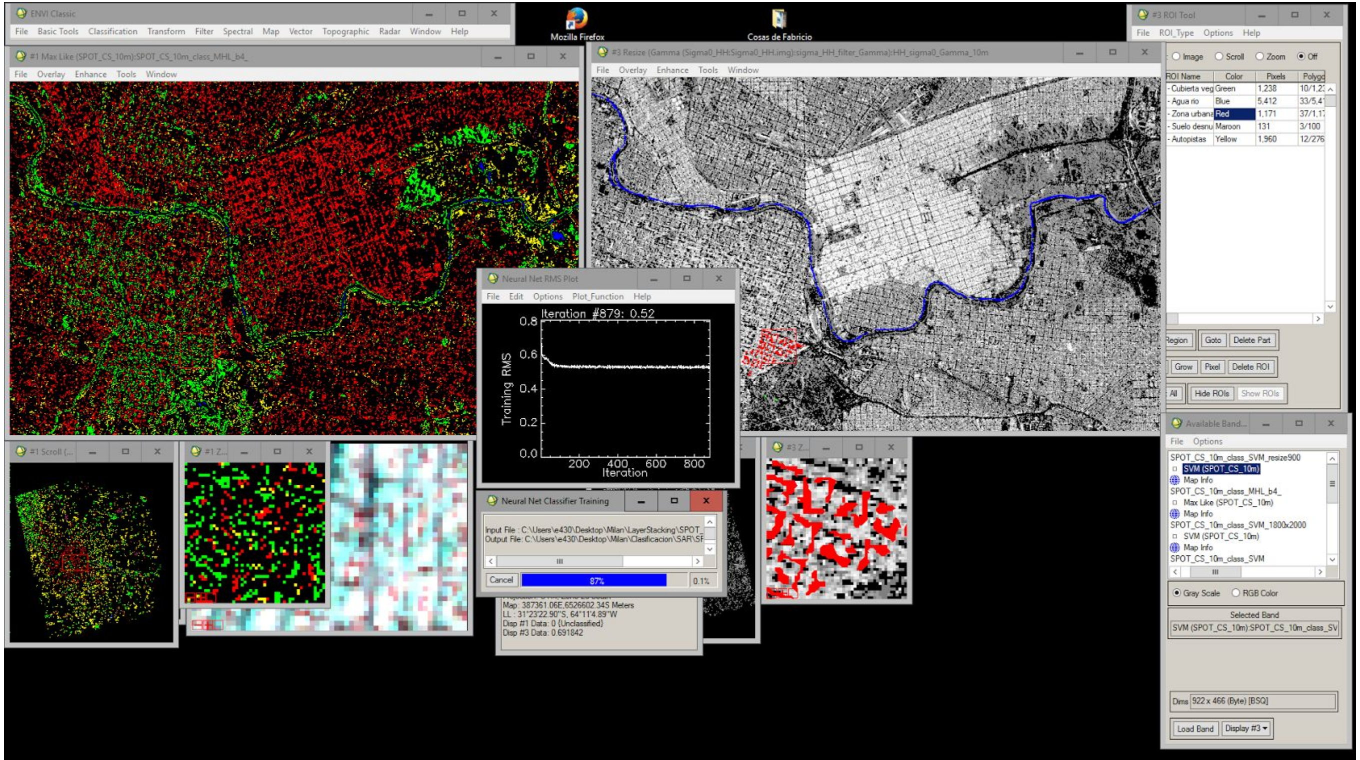
Reading the image SAR corrected and calibrated in the SNAP software.



Image/Sensor: Satellite data



Process classification

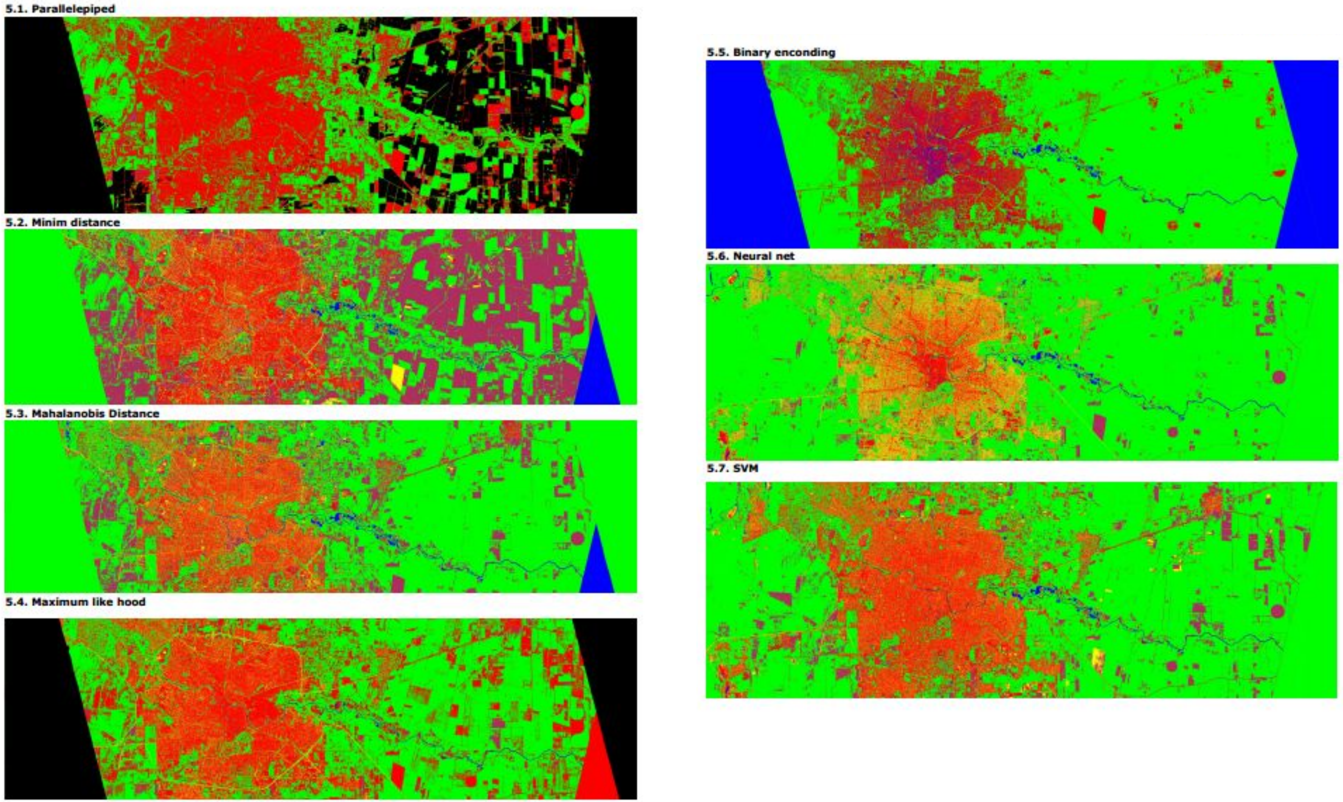


First Results

We expect to obtain a series of classification maps useful to analyze the growth of new built-up areas. and develop new techniques to fully exploiy multitemporal series of optical and SAR adat, understanding how many images are useful to achieve useful results, and how the data should be combined to improve the mapping results.

First Results

Creation of classified maps year 2012.



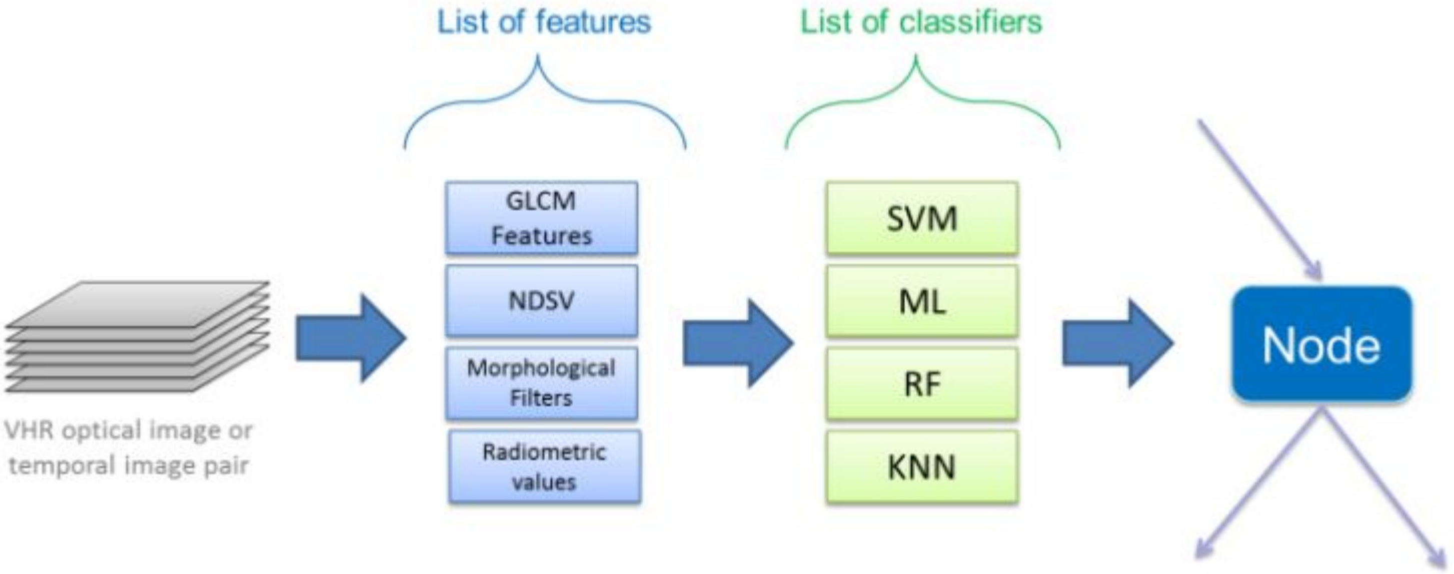
First Results

Classifier Results

Confusion Matrix - Layer Stacking SPOT5_CS		
Classifiers / Rates	**Overall accuracy**	**Kappa Coefficient**
1. Parallelepiped	25.0303%	0.1498
2. Minim distance	83.4847%	0.7505
3. Mahalanobis Distance	88.5291%	0.8219
4. Maximum like hood	79.6610%	0.7127
5. Binary enconding	61.0270%	0.4157
6. Neural net	82.4144%	0.7292
7. SVM	85.4478%	0.7825

Method Hierarchical binary decision tree (HBDT)

.Overview of the method. The best pair, composed by a feature set and a classifier is assigned to each node of the HBDT.



Method Hierarchical binary decision tree (HBDT)

.Papers

IMPROVING THE HBDT FRAMEWORK FUSING HR AND VHR SAR AND OPTICAL DATA FOR IMAGE CLASSIFICATION

G.C. Iannelli and P. Gamba

Department of Industrial and Information Eng., University of Pavia, Via Ferrata 5, Pavia, Italy

ABSTRACT

Land cover mapping is usually characterized by a multi-scale classification, generally because the image-objects are better identified at multiple scales. Consequently, the features are usually extracted at different resolutions, and then classified by means of the same classifier (or ensemble of it). Unfortunately, this approach does not adapt for very different scenes, because some classes are better defined by specified features, and other classes by different ones. In these situations, the only multi-scale analysis may not be sufficient. The Hierarchical Binary Decision Tree (HBDT) approach, instead, combines different processing chains (composed by feature selection and classification steps) and automatically adapts to the spatial and spectral properties of the classes to be recognized in a scene. The HDBT algorithm has already be proven to be efficient for single sources of data. In this paper we propose two different data fusion techniques exploiting HBDT. These two techniques refer one to the fusion of results at the decision level, the other to fusion at the feature level. This work compares the performances of these two techniques using a data set of two city areas, i.e. Pavia (Italy) and Beijing (P.R. China), using HR and VHR optical and SAR data.

with nodes representing spatial resolution values. In [2,3], the multi-scale analysis is exploited by developing segmentation-based decision tree techniques. In doing so, features are extracted from segments delineated at different scales. A similar region-based multi-scale approach is also available in [4], but its results are computationally expensive. In [5,6] the classification is performed by extracting the same features at multiple scales. In [7], the multi-scale data fusion is applied trough wavelet-based methods and hierarchical Markov random field. A similar HBDT methodology has been proposed in [8] for hyperspectral data, but without any multi-scale analysis.

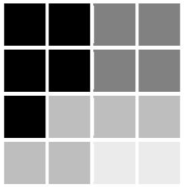
It must be noted that all these techniques do not consider the variability of the features useful to identify a specific class. If only a single predefined set of features, even at multiple scales, is computed, it may not be the most suitable one for all of the classes. Similarly, it may not be the most efficient option for other scenes, whose interpretation will presumably request to identify different classes. To cope with this variability, which is becoming more and more important due to the huge amount of data collected and the need to analyze them quickly, the method proposed in this paper combines different processing chains and automatically selects the spatial/spectral features most suited

Gray-Level Co-Occurrence Matrix (GLCM)

Characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix.

- **Contrast:** Measures the local variations in the gray-level co-occurrence matrix.
- **Correlation:** Measures the joint probability occurrence of the specified pixel pairs.
- **Energy:** Provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment.
- **Homogeneity**: Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

The image as it appears:



The GL (digital numbers) associated with each pixel:

0	**0**	**1**	**1**
0	**0**	**1**	**1**
0	**2**	**2**	**2**
2	**2**	**3**	**3**

Method Hierarchical binary decision tree (HBDT) processing results

1- InstallPython 2.7. According to Operating System (recommended in 64bit).

2- Install IDE as Anaconda.

3- Install required libraries: scikit_learn, numpy, scipy, GDAL-2.1.3 , scikit_image, matplotlib y opencv_python.

4- Instalar QGIS version OSGEO with includes GRASS.

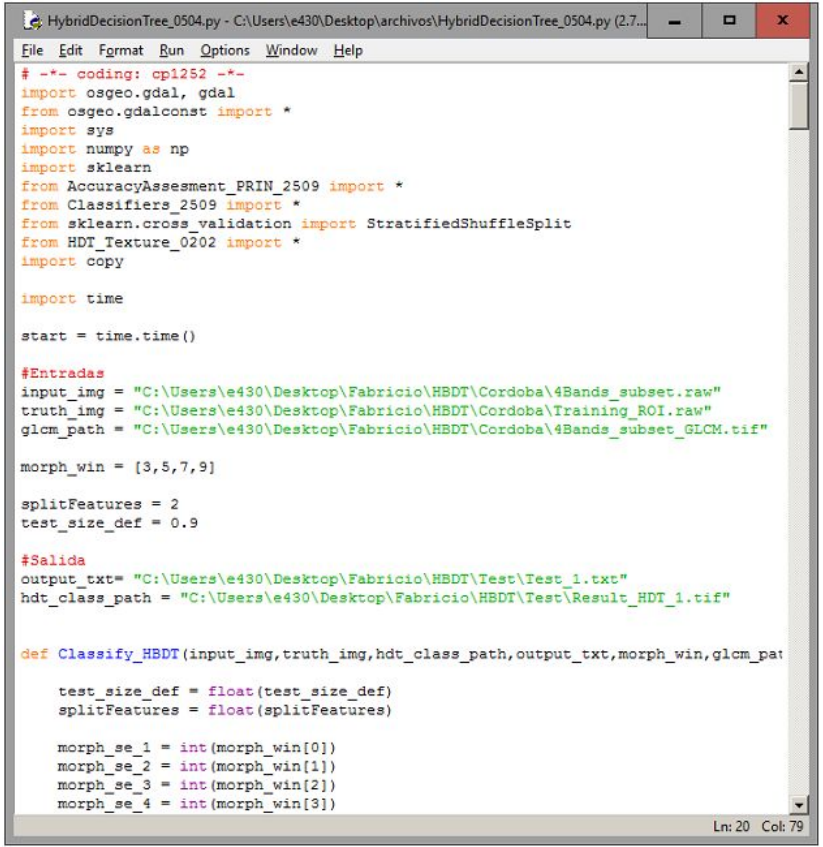
5- Advanced configuration with packages and libraries

6- Add plugin to folder *user>.qgis2/python/plugin/*

7- Enable plugin in the Qgis add-ons menu

6- To carry out the first test, the entries Data.raw and training.raw

7- Add name to the outputs name.tif and name.txt

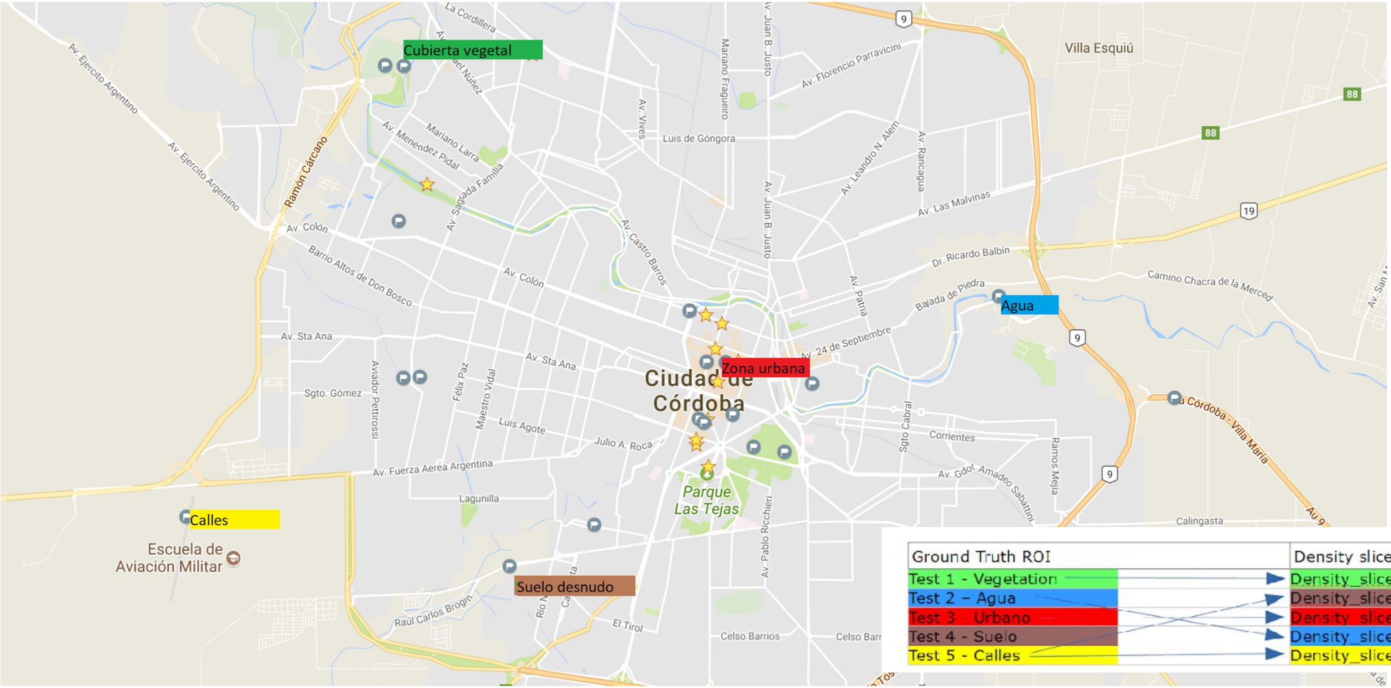



```
# -*- coding: cp1252 -*-
import osgeo.gdal, gdal
from osgeo.gdalconst import *
import sys
import numpy as np
import sklearn
from AccuracyAssesment_PRIN_2509 import *
from Classifiers_2509 import *
from sklearn.cross_validation import StratifiedShuffleSplit
from HDT_Texture_0202 import *
import copy

import time

start = time.time()

#Entradas
input_img = "C:\Users\e430\Desktop\Fabricio\HBDT\Cordoba\4Bands_subset.raw"
truth_img = "C:\Users\e430\Desktop\Fabricio\HBDT\Cordoba\Training_ROI.raw"
glcm_path = "C:\Users\e430\Desktop\Fabricio\HBDT\Cordoba\4Bands_subset_GLCM.tif"

morph_win = [3,5,7,9]

splitFeatures = 2
test_size_def = 0.9

#Salida
output_txt= "C:\Users\e430\Desktop\Fabricio\HBDT\Test\Test_1.txt"
hdt_class_path = "C:\Users\e430\Desktop\Fabricio\HBDT\Test\Result_HDT_1.tif"


def Classify_HBDT(input_img,truth_img,hdt_class_path,output_txt,morph_win,glcm_pat

    test_size_def = float(test_size_def)
    splitFeatures = float(splitFeatures)

    morph_se_1 = int(morph_win[0])
    morph_se_2 = int(morph_win[1])
    morph_se_3 = int(morph_win[2])
    morph_se_4 = int(morph_win[3])
```

Method HBDT processing results

Training classes:

Vegetation (833 pixels), water (811 pixels) , urban (797 pixels), soil (833 pixels), streets(726 pixels).



Ground Truth ROI	Density slice
Test 1 - Vegetation	Density_slice 1_vegetacion
Test 2 – Agua	Density_slice 2_suelo
Test 3 – Urbano	Density_slice 3_urbano
Test 4 - Suelo	Density_slice 4_agua
Test 5 - Calles	Density_slice 5_calles

Method HBDT processing results

To make the confusion matrix, we proceeded by code using the Python language, obtaining the following results from each of the simulations performed.

Simulación 0					
Overall Accuracy = (8109/9643) 84.0921%					
Kappa Coefficient = 0.7674					
Duration = 32 min					
	1	2	3	4	5
Vegetación 1	1238	1345	0	0	43
Agua 2	0	3975	0	0	0
Urbano 3	0	40	1020	0	62
Suelo 4	0	0	0	73	3
Calles 5	0	41	0	0	1803
Totals	1238	5401	1020	73	1911

Method HBDT processing results

Result of the first simulations:

·

6.4. Results of the simulations

On average you can see that round in values at 60% accuracy, with an average duration of 39 minutes of execution each. The simulation 0 was the one that obtained better value of precision with approximately 80%. The execution time of each result is conditioned by the resources of the computer, realized in processor Intel Core i3 with 4Gb of RAM.

Results of the simulations			
N#Simulations	Overall Accuracy (%)	Kappa Coefficient (0-1)	Duration (min)
Simulación 0	84,0921	0,7674	32
Simulación 1	55,2214	0,4476	44
Simulación 2	67,7590	0,5698	34
Simulación 3	59,8569	0,4914	43
Simulación 4	51,0318	0,4043	43
Simulación 5	58,9547	0,4792	34
Simulación 10	62,8020	0,5205	39
Simulación 40	57,3991	0,4676	47
Simulación 50	46,7800	0,3694	36
Simulación 75	55,8540	0,4522	37
Simulación 99	54,0807	0,4316	47
Simulación 100	71,9382	0,6190	31
Average	**60,4808**	**0,50167**	**38,92**

Method HBDT processing results

New results obtained: With parameter settings. He was given to take all the training data. Longer times to obtain results.

Simulación_.0					
Overall Accuracy = (6781/9643) 70.3204%					
Kappa Coefficient = 0.5946					
Duration = 89 min					
	1	2	3	4	5
Vegetación 1	1230	400	0	0	2
Agua 2	0	2582	0	0	0
Urbano 3	0	406	1020	0	32
Suelo 4	0	0	0	73	0
Calles 5	8	2013	0	0	1877
Totals	1238	5401	1020	73	1911

Method HBDT processing results

Summary of new simulations:
The first 4 simulations were done with the data of Training, but with taking all the set of them, visualizing a greater precision at the cost of other times.

N#Simulations	Overall Accuracy (%)	Kappa Coefficient (0-1)	Duration (min)
Simulación_.0	70,3204	0,5946	89
Simulación_.2	69,1486	0,5865	103
Simulación_.5	70,2271	0,6001	87
Simulación_.100	70,1856	0,5996	114
Simulación_.2T	72,4357	0,6582	62
Simulación_.5T	81,5475	0,7692	137
Simulación_.10T	89,7684	0,8719	75
Average	**74,8048**	**0,66859**	**68,41**

Method HBDT processing results

The method was adjusted and optimized, obtaining only the values of the matrix of confusion and of the measurers.
Test set 50% - Training set 30%

	Test Set 50% - Training Set 30 %			
Types of Data	***Overall accuracy***	***Average accuracy***	***Kappa***	***Processing time (sec)***
HBDT_SAR	88,43	84,92	0,8491	1309
HBDT_Multiespectral	93,13	91,57	0,9107	1347
HBDT_SAR+Multiespectral	99,3	99,31	0,9941	1269

Method HBDT processing results

The method was adjusted and optimized, obtaining only the values of the matrix of confusion and of the measurers.
Test set 50% - Training set 20%

	Test Set 50% - Training Set 20 %			
Types of Data	*Overall accuracy*	*Average accuracy*	*Kappa*	*Processing time (sec)*
HBDT_SAR	88,32	84,8	0,8477	1394
HBDT_Multiespectral	92,99	91,31	0,9089	1404
HBDT_SAR+Multiespectral	99,26	99,28	0,9904	1285

Method HBDT processing results

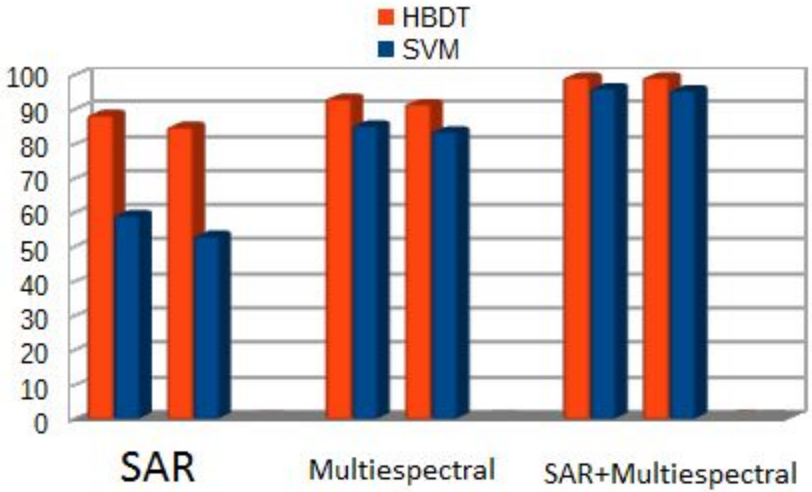
The method was adjusted and optimized, obtaining only the values of the matrix of confusion and of the measurers.
Test set 50% - Training set 10%

	Test Set 50% - Training Set 10 %			
Types of Data	*Overall accuracy*	*Average accuracy*	*Kappa*	*Processing time (sec)*
HBDT_SAR	87,64	84,11	0,839	1495
HBDT_Multiespectral	92,49	90,61	0,9023	1517
HBDT_SAR+Multiespectral	98,88	98,82	0,9854	1333

Method HBDT processing results

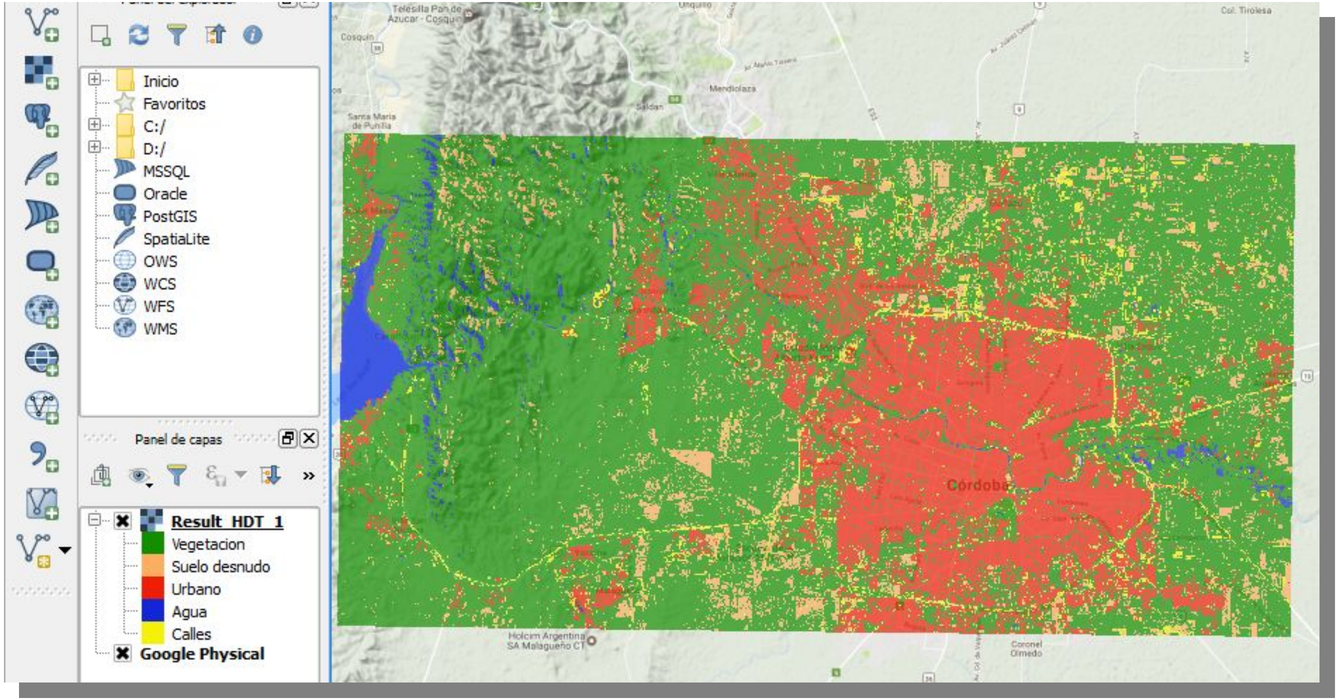
The HBDT is compared with the SVM method, where the following values are observed.

Comparison between classifiers			
Type of Data	**Measurers**	**HBDT**	**SVM**
SAR	Overall accuracy	88,43	59,31
	Average accuracy	84,92	53,24
	Kappa	0,8491	0,4548
Multiespectral	Overall accuracy	93,13	85,38
	Average accuracy	91,57	83,69
	Kappa	0,9107	0,8095
SAR + Multiespectral	Overall accuracy	99,3	96,18
	Average accuracy	99,31	95,59
	Kappa	0,9941	0,9504



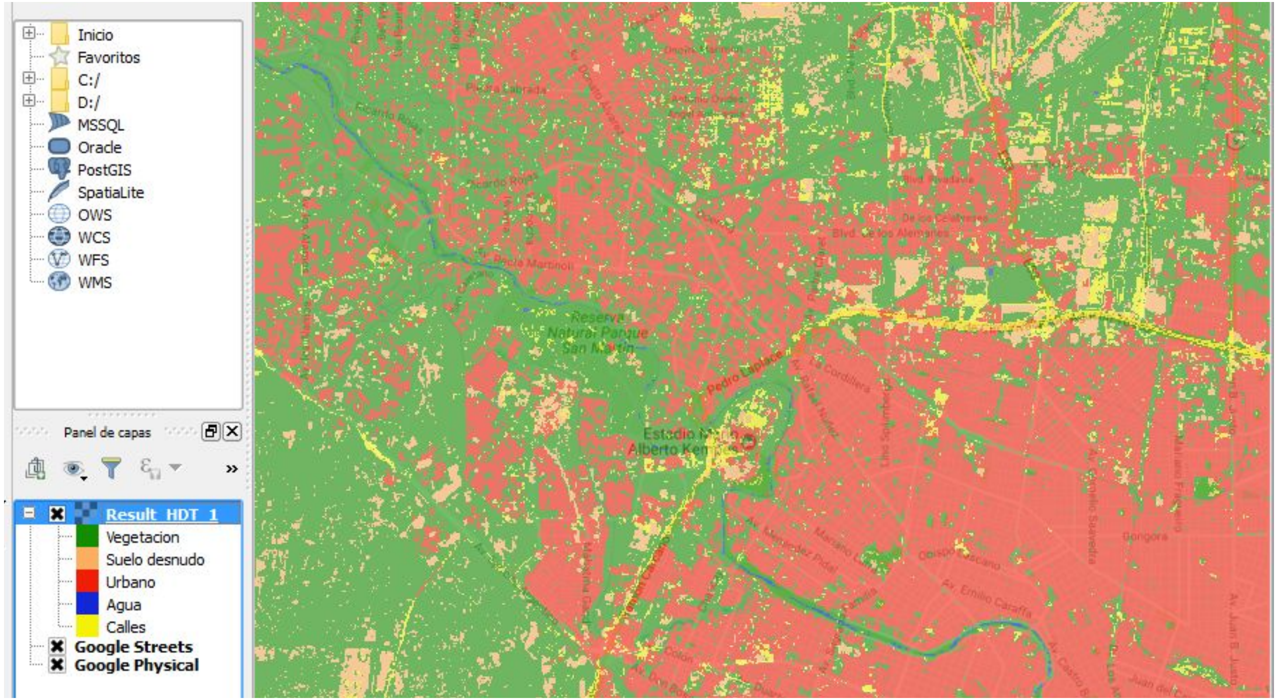
Method HBDT processing results

Map obtained from the HBDT classification



Method HBDT processing results

Map obtained from the HBDT classification. Visualization in nearer areas, with the Google streets layer.



Method HBDT processing results

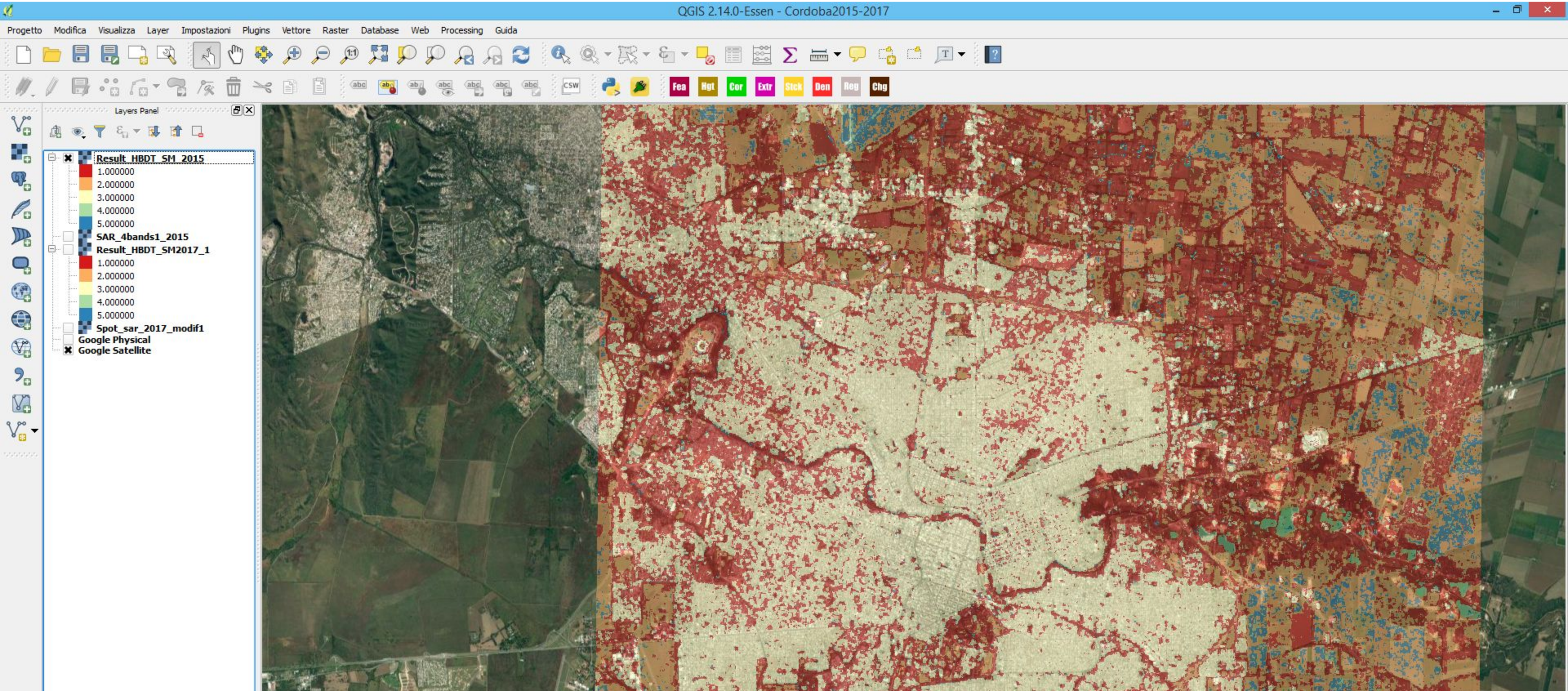
Comparison between different types of data over the years using HBDT.

.

		2012		2015		2017	
Data	Percentage types	HBDT (%)	Time processing (min)	HBDT (%)	Time processing (min)	HBDT (%)	Time processing (min)
SAR	SAR 50_10%	87,104467	43	84,1198377	39	86,039613	39
	SAR 50_20%	86,32645	40	85,8978356	41	84,654135	38
	SAR 50_30%	86,330654	39	87,8436286	43	85,917475	40
	Average	**86,58719**	**40,7**	**85,953767**	**41,0**	**85,5370741**	**39,0**
Multiespectral	Multiespectral 50_10%	89,496722	49	90,6187542	47	92,433076	44
	Multiespectral 50_20%	91,384823	47	89,3487635	46	92,473077	45
	Multiespectral 50_30%	90,959664	46	91,0392876	48	91,909877	43
	Average	**90,613736**	**47,3**	**90,335602**	**47**	**92,272010**	**44**
SAR+Multiespectral	SAR+Multiespectral 50_10%	89,370347	51	98,88203364	49	97,032104	45
	SAR+Multiespectral 50_20%	89,202479	48	96,7889997	50	97,084603	45
	SAR+Multiespectral 50_30%	89,035557	46	94,98777656	48	96,876576	44
	Average	**89,202794**	**48,3**	**96,88627**	**49**	**96,997761**	**44,7**

Method HBDT processing results

Differences visible in the city of Córdoba.

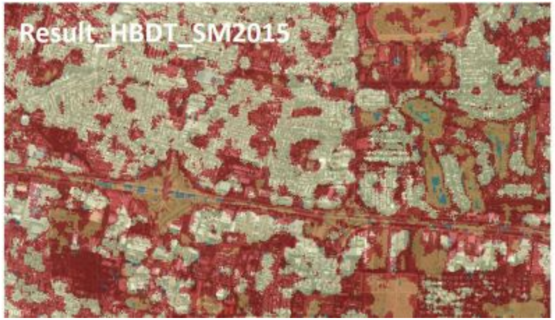


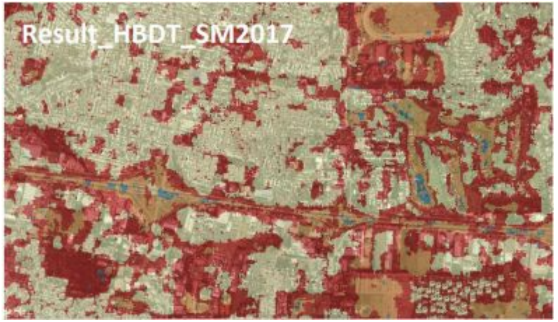
Method HBDT processing results

Classification Maps years 2012 - 2017
Nueva Cordoba - CSM & SPOT

Method HBDT processing results

Classification Maps years 2012 - 2017
Cordoba south zone - CSM & SPOT





Method Hierarchical binary decision tree (HBDT) processing results

Conclusions

- The use of free software next as Python, SNAP, allowing to adapt according to the needs and or requirements of the user.
- Some of the problems presented were not validated ground truth data, generating problems in the correct generation of classification.
- The HBDT method obtained excellent results applied to the study area, obtaining greater results in the combination of data SAR+Multiespectral.
- The times of computations allow to make multiple classifications in a single step.
- Ability to apply to other data sources.

References

- G.C. Iannelli, G. Lisini, F. Dell'Acqua, R. Queiroz Feitosa, G.A.O.P. Costa, P. Gamba, “Urban area extent extraction in spaceborne HR and VHR data using multi-resolution features”, Sensor, doi:10.3390/s141018337, vol. 14, pp. 18337-18352, 2014.
- G.C. Iannelli, P. Gamba, “Hierarchical hybrid decision tree multiscale fusion for urban image classification”, Proc. of IGARSS’16, Beijing (P.R. China), July 2016, pp. 1800-1803, doi: 10.1109/IGARSS.2016
- Thomas Lillesand, Ralph W. Kiefer, Jonathan Chipman: “Remote Sensing and Image Interpretation”. Wiley. ISBN- 13: 978-0470052457 - Capitulos 1.
- COSMO-SkyMed SAR Products Handbook. Link: http://www.e-geos.it/products/pdf/csk-product%20handbook.pdf
- Quick Guide for the SPOT Product User
- https://catalogos4.conae.gov.ar/spot6/Docs/Guia-para-el-Usuario-de-Productos-SPOT-6y7.pdf
- Telecommunications & Remote Sensing Laboratory, Department of Electrical, Computer and Biomedical Engineering - University of Pavia. Link: http://tlclab.unipv.it/sito_tlc/downloads.do

Acknowledgments