25th June 2010

HAIS 2010 – San Sebastián, Spain

A SVM and k-NN restricted stacking to improve land use and land cover classification

AUTORES

Jorge García Gutiérrez

Department of Computing Languages and Systems – University of Seville

Daniel Mateos García

Department of Computing Languages and Systems – University of Seville

José C. Riquelme Santos

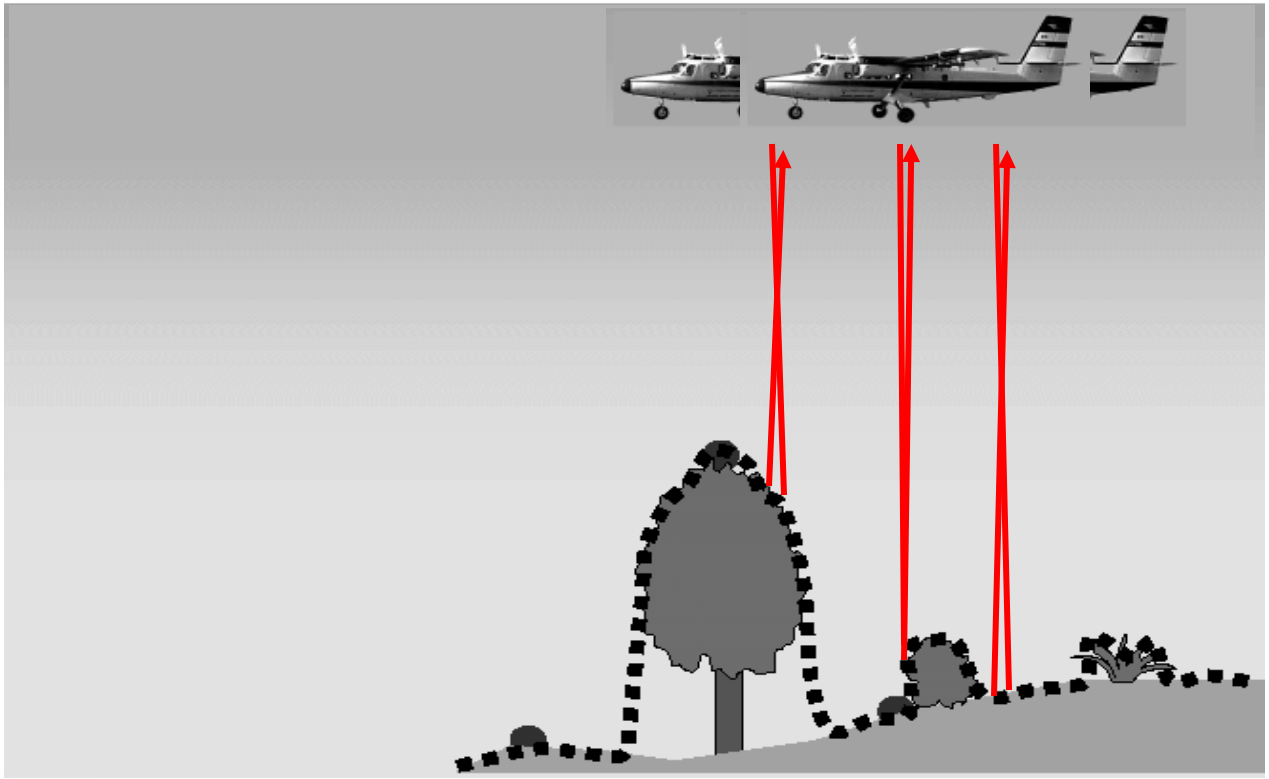
Department of Computing Languages and Systems – University of Seville

Background I

- **Land use and land cover maps (LULC):**
  - The most important product to monitor the Natural Environment.
  - Very useful to develop policies to manage specially interesting areas.
- **Data source:**
  - Aerial or satellite images (spectral, multispectral, hyper...)
  - Research on other sources.
- **Automation:**
  - Most cases low level.
  - Almost two months to develop a LULC manually for a new area.
- **Accuracy:**
  - Should be over 85%.

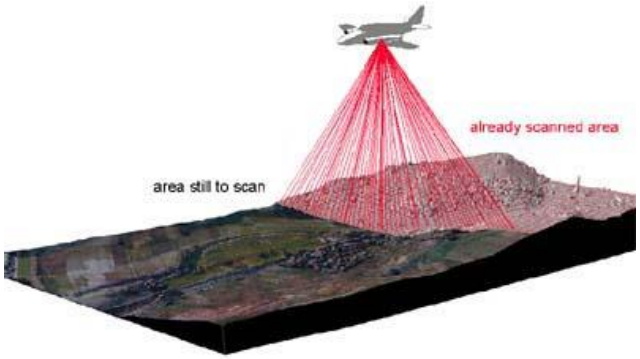
Background II

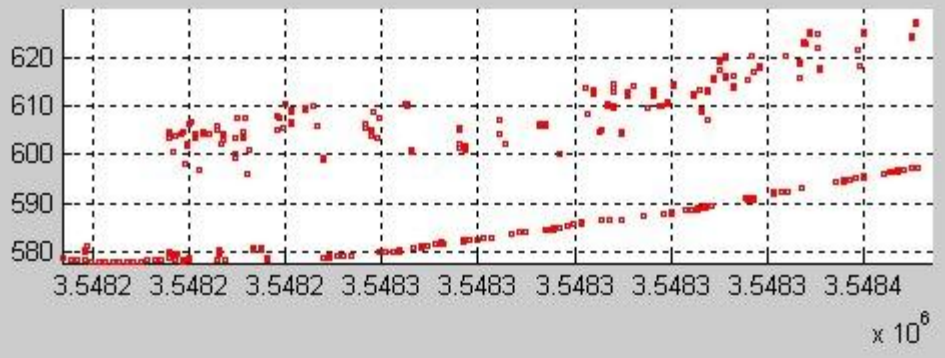
- **Light Detection And Ranging (LIDAR)**



Background II

- LIDAR point cloud:

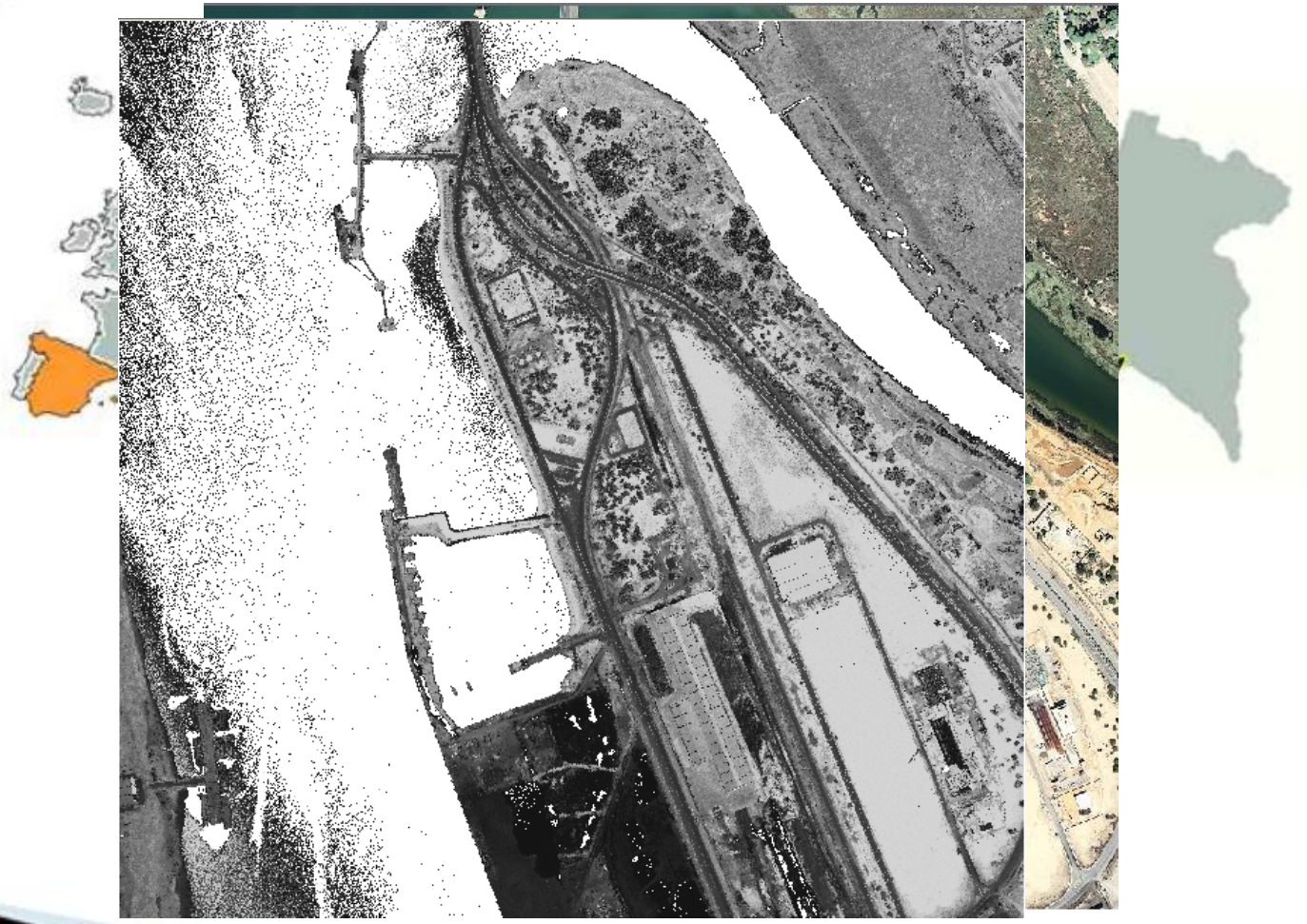




Introduction

- **OBJECTIVE:** Use data fusion (aerial images and LIDAR) and intelligent techniques to develop LULC maps automatically.
- **WHAT FOR:** Improve the results and decrease development time and costs on classical LULC classification.
- **WHY:** More quality and efficiency for environmental products is needed specially when investment risks being reduced.
- **HOW:** Using well-known intelligent techniques from data mining.

Data

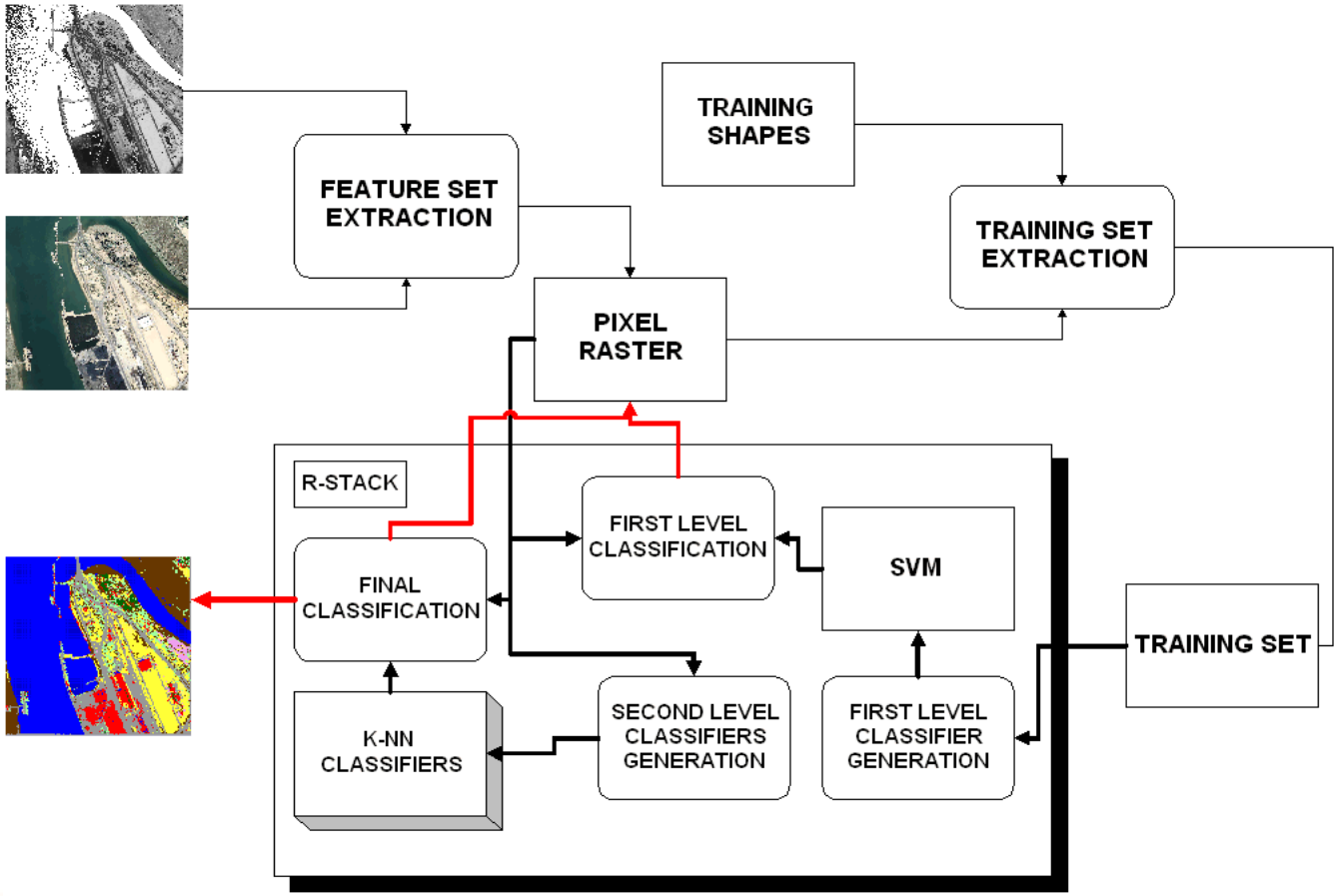


Method I

- Feature set

Variable	Description	Variable	Description
SNDVIMIN	SNDVI minimum	ICV	Intensity coefficient of variation
SNDVIMAX	SNDVI maximum	HCV	Height coefficient of variation
SNDVISTD	SNDVI Standard deviation	SLP	Slope
SNDVIAVG	SNDVI average	CRR	Canopy relief ratio
MIN(*)	Minimum	PEC	Penetration coefficient
MAX(*)	Maximum	TOTALR	Total of returns
STD(*)	Standard deviation	PCTN1	Unique return percentage
AVG(*)	Average	PCTN2	Double return percentage
VAR(*)	Variance	PCTN3	Three or more returns percentage
SKEW(*)	Skewness	PCTR1	First return percentage
KURT(*)	Kurtosis	PCTR2	Second return percentage
RANGE(*)	Range	PCTR3	Third or later return percentage
NOTFIRST	Second or later return	PCTR31	PCTR3 over PCTR1
EMP	Empty neighbours	PCTR21	PCTR2 over PCTR1
		PCTR32	PCTR3 over PCTR2

Method II


TRAINING SHAPES
FEATURE SET EXTRACTION
TRAINING SET EXTRACTION
PIXEL RASTER
R-STACK
FIRST LEVEL CLASSIFICATION
SVM
FINAL CLASSIFICATION
TRAINING SET
K-NN CLASSIFIERS
SECOND LEVEL CLASSIFIERS GENERATION
FIRST LEVEL CLASSIFIER GENERATION

Results I

- **SVM: hold-out**

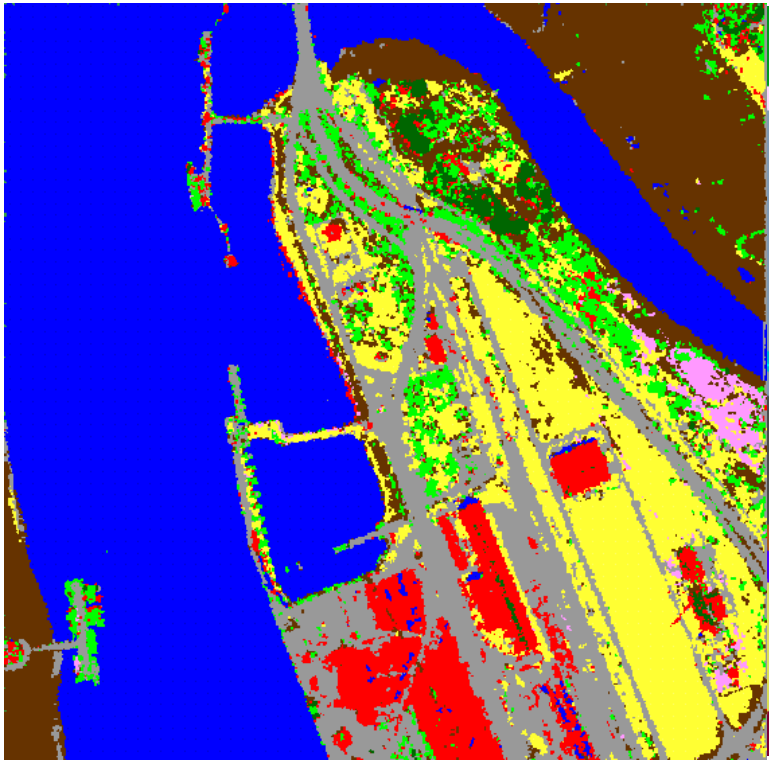
User class \sample	Water	Marsh	Roads or railways	Low Veg.	Middle Veg.	High Veg.	Buildings and other urban areas	Dumps
Water	2148	0	0	0	0	0	3	0
Marshland	2	1171	45	8	38	0	2	0
Roads or railways	6	25	962	39	24	0	26	1
Low Veg.	0	114	0	572	0	0	0	0
Middle Veg.	0	112	22	12	310	0	8	0
High Veg.	0	8	2	0	49	234	36	0
Buildings and other urban areas	42	11	50	20	4	115	1072	0
Dumps	0	58	7	2	0	0	0	142
KIA	0.855							
Correctly classified	**0.881**							

Results II

- **R-STACK: hold-out**

User class \sample	Water	Marsh	Roads or railways	Low Veg.	Middle Veg.	High Veg.	Buildings and other urban areas	Dumps
Water	2151	0	0	0	0	0	0	0
Marshland	2	1199	32	9	24	0	0	0
Roads or railways	7	16	1000	38	8	0	13	1
Low Veg.	0	111	0	575	0	0	0	0
Middle Veg.	0	82	23	13	346	0	0	0
High Veg.	0	2	2	0	42	261	22	0
Buildings and other urban areas	30	6	34	16	2	104	1122	0
Dumps	0	33	6	0	0	0	0	170
KIA	0.89							
Correctly classified	**0.91**							

Results III



Conclusions

- Remote data mining techniques can lower time to develop environmental products.
- Data fusion and ensembles are an excellent tool to improve the general accuracy of thematic maps.
- Future work:
  - Apply evolutive computation to control feature weighting.
  - Test more complex ensembles.
  - Prior phase to detect critical objects: docks, buildings,...

THANK YOU VERY MUCH

jgarcia@lsi.us.es