

IDES project: an advanced tool to investigate illegal dumping

G. Persechino¹, M. Lega², G. Romano¹, F. Gargiulo¹ & L. Cicala¹

¹*CIRA, Italian Aerospace Research Centre, Italy*

²*Department of Environmental Sciences, University of Naples, Italy*

Abstract

IDES – Intelligent Data Extraction System – is the name of a new project, founded by Campania Regional Government (Italy) and developed by the CIRA – Italian Aerospace Research Centre that aims at implementing an advanced tool to support Government Bodies in the discovery and localization of environmental criticalities. In detail, this project will develop a software and hardware platform for image, data and document analysis in order to support law enforcement investigations. The fight against illegal waste dumping is the first objective of the IDES project and the main goal is to develop algorithms for image analysis in order to automatically identify this environmental criticality. The first task is to identify the statistical link between specific features extracted from satellite images or acquired by airborne sensors, possibly in combination with punctual information (results of chemical analysis, physical property measurements, information pertaining to the population), and environmental criticalities (e.g. illegal dumping, illegal landfills etc). IDES will offer an integrated Geographic Information System (GIS) repository of information extracted and collected by various government bodies. Furthermore, an innovative aspect of IDES is the integration between information extracted from remotely sensed images and information extracted from textual documents through a semantic analysis. The results of this last analysis will be a taxonomy-driven document searching and browsing system in which all documents are categorized and all relevant entities in documents (e.g. people, organizations, places and events) are extracted and represented by means of concept maps. Finally, these results: a) provide support to government bodies “end-users”; b) can be included in a GIS; c) can constitute the basis for the application of Artificial Adaptive Systems used in the domain of semantic analysis of



unstructured information providing unknown relationships between entities (e.g. criminal associations, people, places, activities, buildings, towns and dates).

Keywords: illegal dumping, landfills monitoring, interoperability, unmanned aerial systems, synthetic aperture radar, multispectral images, infrared images, text semantic search, geographical information systems.

1 Introduction

A recent census of the Italian illegal dumping sites estimates the presence of 4866 illegal dumping. Only 21% of surveyed landfills have been reclaimed and more than 700 contain hazardous waste (Persechino *et al.* [1]).

In Italy, the Campania Region might be associated with a history of illegal dumping and disposal of hazardous waste.

The citizens' exposure to toxic waste is a major public health problem; therefore, the responsible public authorities have the need to act as soon as possible in the identification of such environmental issues. Identifying the problem, as soon as it occurs, would reduce reclamation costs. So it is necessary a more frequent and more targeted monitoring of large areas.

On the other hand, public authorities do not have large budgets, so the required solution must also be economically feasible. Many employers and gears are involved in territory control requiring high costs for this activity. But control of the territory is a necessary action for the prevention of such unlawfully. It is a difficult challenge to perform long-term environmental monitoring with today's manned aircraft because of vehicle, cost, and mission limitations.

Moreover, where there are small sources of pollution and contamination over a wide area, the illegal dumping is very difficult to detect. The use of satellite monitoring exceeds the current limits of traditional methods of detection. The satellite images are able to continuously monitor, in terms of space and time, large part of the territory. Early intervention on the waste accumulation is the only way to prevent its transformation in illegal landfill. The continuous satellite scanning allows the detection of anomalies connected to the dumping area employing a smaller number of operators. In this way the operators can arise only in aimed interventions having more time for other important police activities. Another plus of early warning in the contaminated area is a more effective remediation so that the territory can be back to his old self again. The treatment of small accumulations of waste requiring disposal and reclamation entails much smaller costs than those for the remediation of large volumes of waste and it's more safety for exposure to the risk of contamination from the people living around there.

Early warning also prevents the contamination of wider areas due to the dispersion phenomena mediated by trophic chain and atmospheric agents. government bodies must ensure to citizens the healthy avoiding their exposure to contaminants.



2 IDES – Intelligent data extraction system

IDES – Intelligent Data Extraction System – project aims at implementing an SW platform for data analysis within the domain of environmental criticalities.

IDES will offer an integrated repository of information extracted from heterogeneous, physically distributed, unstructured sources (satellite and airborne data, web pages, etc.) by means of a capture, extraction and analysis process. Based upon the integrated information stored in a Geographic Information System and elaborated with the aid of advanced data analysis techniques and tools, IDES will be able to extract the *hidden* information, that is information non immediately identifiable through a mere reading or a deeper analysis, even if performed by a domain expert; at this end, IDES will be able to uncover patterns, multi-disciplinary correlations not known a priori and it will be able to extract relevant information useful for government bodies.

There are available, with regard to these issues, a large amount of data that describe anthropic activities with heavy impact on environmental health; so to this end, the project team focussed on the following dataset:

- dumping and landfills data related to specific industrial installations;
- urban solid waste temporary deposits/facilities;
- statistical data about urban people.

All above data, coming from different sources available in the project, are different for semantics (chemical, emissions, county people, installations address, etc.) and structure (tabular, vector, raster, structured, unstructured). IDES is an environmental Geographic Information System (GIS) with the strategic mission to be a centralized and unified informative system containing all data coming from different sources (satellite, airborne, terrain data) and tools enriched with geographical information enabling detailed spatial analysis tasks.

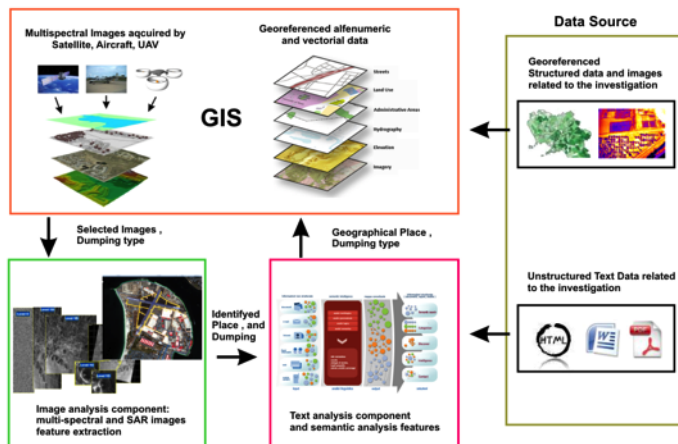


Figure 1: IDES components workflow.

In order to deal with different types of data and to enable user analysis, the IDES SW platform is based on different SW components each of which with a well defined responsibility:

- text analysis component with semantic analysis features; its objective is to analyse text document coming from intranet and web sources in order to automatically extract entities and relationship among them with which to build conceptual map useful for a primary illegal crime detecting;
- image analysis component for analysing multi-spectral and SAR images through advanced algorithmic features to mine pattern among data; this represents a powerful key to identify and geo-locate sites with potential illegal activity;
- geographical information system and geostatistical analysis component for spatial data exploration and analysis with the capabilities to create statistically valid prediction information from a limited number of data measurements.

2.1 IDES Geographical Information System

The IDES project involves several government agencies including: Autorità di Bacino, Arpa Campania, Genio Civile, Provincia di Caserta, etc. Each agency also participates in the project by providing its own data. Such data are always geo-referenced or they could be geo-referenced with a little effort. Then, the first step is to collect all this information into a GIS.

A geographic information system is a software system used to describe and characterize the earth and other geographies for the purpose of visualizing and analyzing geographically referenced information.

GIS is one of the most powerful of all information technologies because it focuses on integrating knowledge from multiple sources (for example, as layers within a map) and creates a crosscutting environment for collaboration. In addition, GIS is attractive because it is both intuitive and cognitive. It combines a powerful visualization environment, using maps to communicate and visualize geographic data through a strong analytic, statistical and modeling framework.

A GIS utilizes a layer-based geographic information model for characterizing and describing our world and models geographic information as a logical set of layers or themes. For example, a GIS can contain data layers for the following:

- streets represented as centerlines;
- land-use areas that represent vegetation, residential areas, business zones, and so forth;
- administrative areas;
- water bodies and rivers;
- parcel polygons representing landownership;
- a surface used to represent elevation and terrain;
- an aerial photo or satellite image for an area of interest.

A GIS uses maps to visualize and work with geographic information.



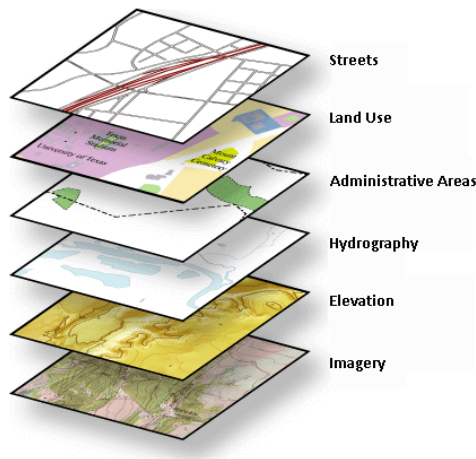


Figure 2: GIS layers.

Each GIS includes a set of interactive maps and other views (such as 3D globes) that show features and relative relationships on the earth's surface. Various map views of the underlying geographic information can be constructed to support query, analysis, and editing of geographic information. Maps can also be used to access geographic modeling tools that are used to derive new information.

A GIS has a comprehensive set of analytic and data transformation tools to perform spatial analysis, data processing and geostatistical analysis. Geo-processing is the methodical execution of a sequence of operations on geographic data to create new information. GIS includes a large set of geo-processing functions to take information from existing datasets, apply analytic functions, and write results into new result datasets. Each geo-processing tool takes existing information as input and derives a new result, which can be used in subsequent operations. This ability to string together a logical sequence of operations so that you can perform spatial analysis and automate data processing, all by assembling a model, is one of the key elements of GIS.

The emphasis of spatial analysis is to measure properties and relationships, taking into account the spatial localization of the phenomenon under study in a direct way. That is, the central idea is to incorporate space into the analysis to be made (Camara *et al.* [7]). Spatial analysis is useful for evaluating suitability and capability, estimating and predicting, and interpreting and understanding.

Geostatistics is a branch of statistics focusing on spatial or spatiotemporal datasets. It is related to interpolation methods, but extends far beyond simple interpolation problems. Geostatistical techniques rely on statistical method to model the uncertainty associated with spatial estimation and simulation. Geostatistical algorithms are incorporated in GIS.

For example, starting with industrial dumping registered in the province of Caserta, we use the geostatistical analysis to identify areas where the distribution

of these dumping is “abnormal” with respect to the density of population and productive activities. This may provide a guideline for the presence of potential illegal dumping.

2.2 Text analysis component and semantic analysis features

An innovative aspect of IDES is the integration between information extracted by remotely sensed images and information extracted from textual documents. There is a huge amount of information in unstructured documents (for example, it has been estimated that about 80% of corporate information is unstructured and contained in spreadsheets, presentations, e-mails, web pages and .pdf files).

Thanks to search methodologies based on deep linguistic analysis (semantic engines), it is possible to grasp the “meaning” of the content and go beyond traditional search tools

Such a kind of semantic engines relies on a semantic network, (e.g. ontologies), that contains all information about terms, grammar rules, meaning relations (i.e. synonyms, hyponyms, hypernyms, and so on) of the language (i.e., Italian, English). The semantic network is the core of the semantic engine, it contains about half a million concepts and four million relations.

In addition to basic linguistic analysis (grammar and morphology), the semantic engine used in the IDES project performs: logical and sentence analysis, disambiguation of polysemic terms and entity extraction (i.e. people, place and organization).

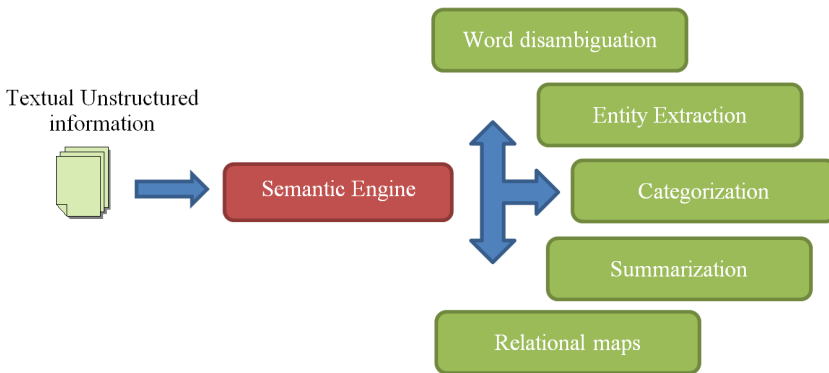


Figure 3: The relationships between the extracted entities represented by the “relational map”.

A customization, also called “training”, of the semantic engine must be performed in order to obtain: a better automatic comprehension of the content, a correct disambiguation of terms and a high level of quality of semantic-based categorization in specific domains, such as environmental criticalities domain.

There are two types of customization: Extensions of the semantic network and Definition of domain specific taxonomies. The former is a deeper

personalization because a relevant number of new concepts (e.g. pollutant, criminal associations) and their relations are hard-coded in the semantic network. Through the semantic network, the semantic engine reconstructs the meaning of each word and each sentence of text. The definition of specific domain taxonomies instead allows us to classify each document in one or more categories with a certain percentage of membership. The documents are classified by the application of specific rules. A training set allows analysts to derive the classification rules. For example, the taxonomy may include a category for each environmental criticality and, during the step of classification, the semantic engine inserts each document in the right categories. In this way the user can access the documentation in a selective manner. Furthermore, advanced automatic semantic summarization allows the user to examine a summary of a set of documents. Then, the user can decide very quickly whether or not these documents are useful.

In the context of the IDES project, the semantic analysis of large amounts of unstructured information is guided by information obtained from the analysis of images.

By means of the semantic engine specifically trained documents both in the public domain (e.g. news and technical reports from public agencies or industry, blogs) and reserved domain (e.g. documentation of real investigations) will be analysed.

The results of this analysis will be a taxonomy-driven document searching and browsing system in which all document are categorized and all relevant entities in documents (e.g. persons, organizations, places and events) are extracted and represented by means of concept maps. These results: a) provide support to end-users; b) can be included in the GIS layer if they could be geo-referenced; c) can constitute the basis for the application of Artificial Adaptive Systems used in the domain of semantic analysis of unstructured information that allow to visually represent obvious connections and probable links between entities (e.g. criminal associations, persons, places, activities, buildings, towns and dates).

2.3 Image analysis component: multi-spectral and SAR images feature extraction

Remote sensed images from satellite platforms have been used to derive interesting feature maps suitable for environmental applications.

The aim is to localize high environmental risk regions. The localization process is performed by terrain classification, change detection and data fusion with not imagery derived data. A GIS is used to combine the imagery derived data with the others.

Two kinds of very high resolution (VHR) imagery sensors have been used:

- multispectral (MS) Images, GeoEye-1 sensor, operating in blue, green, red and near infrared spectral bands;
- synthetic aperture radar (SAR) images, Cosmo SkyMed sensor, operating in microwave X band.



Different acquisitions have been committed for the same region to analyse.

In the MS case, we planned to acquire new data in the same month and with the same sensor pose with respect to available archive data, so that a multi-temporal analysis can be performed without introducing errors due to the different acquisition geometry.

In the SAR case, we planned to use a long time series (8 scene) of the Map Italy project (ScanSAR acquisition mode of the Cosmo SkyMed sensor), obtained with the same acquisition geometry and interleaved approximately by one month, and a pair of interferometric couple, acquired in a committed mission.

Table 1: Used sensors.

Sensor	Sensor type	Sensor mode	On ground resolution	Temporal acquisition	Time series length
GeoEye-1	Multispectral	MS (4 ch) + Panchromatic	4 m /pixel + 1 m/pixel	1 year	2
Cosmo Sky Med	SAR	ScanSAR	3 m /pixel	1 month	8
		Spotlight interferometric couple	1 m /pixel	1 month	2

The imagery raw data has been processed, obtaining, as final products, some indexed maps, each of them related to a specific phenomenon to be observed. The maps, originally rasters, can be easily vectorialized for an immediate use in a GIS.

The processing workflow can be summarized in the following table.

Table 2: Used processing workflow.

General case	MS	SAR
Pre-processing	Pan Sharpening	Spatial Multilooking
Temporal Alignment	Coregistration	Coregistration + Temporal Despeaking
Feature Extraction (spectral, spatial, temporal)	Spatial Feature Extraction, Spectral Feature Extraction	Interferometry
Segmentation	Terrain Classification Map	Coherence Maps
Geometry Correction (using DTM)	Ortorectification	Geocoding
Fusion		
Change Detection		

In general the processing workflow consists in five phases. The first step is the pre-processing of the images. The resolution of the MS images can be improved, using the panchromatic channel, by pan-sharpening. The SAR images, that suffer from the speckling phenomenon, can be enhanced using a spatial multilook strategy. The second step consists in the multitemporal coregistration of the time series. It can be performed, using control points. In the SAR case, the coregistration can be simultaneously performed with a temporal filtering, in order to further reduce the speckle. Then features of interest can be extracted.

The feature can be of spectral type (in the MS case), of temporal type (like in the SAR case, where interferometry is performed), of spatial type (in both cases, in order to exploit spatial patterns like that well characterize manmade objects like building or roads). Extracted the interest features, the source image is partitioned in homogeneous regions, in order to finally create polygons that can be easily handled by the GIS geostatistical analyser. At the end of the processing workflow, the geometry of the maps has to be corrected. For this purpose we planned to use the Digital Terrain Model of Italy (DTM), 20 m on ground resolution, produced by the Military Geographic Institute.

Two separate workflows have been designed for the Multispectral and the SAR cases, because the different geometry of the two sensors. The final product of the MS workflow is a terrain classification map (with associated probability levels), the final product of the SAR workflow is a coherence map. The geo-corrected segmentation maps are the final output of the two processing workflows, while the environmental analysis, consisting in a data fusion and a change detection, is performed with the GIS tools.

The used features are summarized in the following table.

Table 3: Used spectral/spatial/temporal features.

Environment feature	Image feature	Map index
Multispectral		
Vegetation / Culture	Spectral Angle	Vegetation Presence Probability Level
Water	Spectral Angle	Water Presence Probability Level
Soil	Spectral Angle, Corner Detectors (low response)	Soil Probability Level
Rocks	Spectral Angle, Corner Detectors (low response)	Rock Probability Level
Buildings	Spectral Angle, Corner Detectors (high response)	Urban Area Probability Level
Roads	Spectral Angle, Line Detectors (high response)	Urban Area Probability Level
Legal Garbage	Spectral Angle, Corner Detectors (high response)	Legal Garbage Probability Level
Illegal Garbage	Spectral Angle, Corner Detectors	Illegal Garbage Probability Level
SAR		
Incoherence in non vegetation areas	Interferometric Coherence	Change Probability Level

The analysis is focused mainly on temporal changes. The change detection is obtained statistically combining two approaches:

- comparing two MS segmentation map;
- analysing the SAR interferometric map, evaluating the coherence information only for the some terrain types (excluding, for example, roads, water or vegetation).

The main changes of interest are reported in the following table.

Table 4: Temporal changes of interest (I=interesting, DC=don't care, -= not frequent, the not considered, E=excluded from the analysis.

Time 2 (right) Time 1 (down)	Vegetation	Water	Soil	Rock	Building	Road	Legal garbage	Illegal garbage	Coherence information
Vegetation	DC	DC	I	I	DC	I	I	I	E
Water	DC	DC	DC	DC	DC	DC	I	I	E
Soil	DC	DC	DC	-	DC	I	I	I	I
Rocks	I	-	-	DC	DC	DC	I	I	I
Buildings	DC	-	DC	-	DC	DC	-	-	DC
Roads	DC	-	I	DC	DC	DC	-	-	E
Legal Garbage	DC	-	-	I	-	-	I	I	I
Illegal Garbage	I	-	I	I	-	I	I	I	I

3 Preliminary results

The project IDES is now in the test phase of the software components. Some preliminary results of individual software components are outlined here.

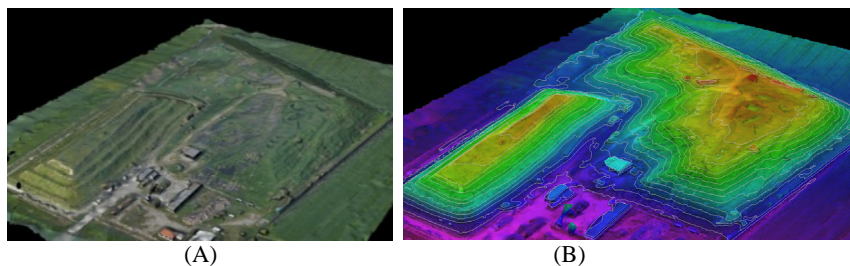


Figure 4: (A) Example of a landfill 3D reconstruction by set of aerial images. (B) Example of a landfill characterization and geometrical analysis.

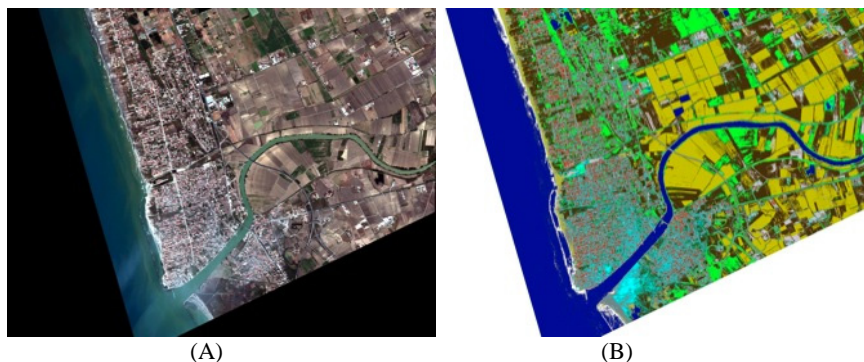


Figure 5: (A) Example of multispectral remote sensed images from satellite: Volturno (2011). (B) Example of a multispectral remote sensed images feature extraction: vegetation, soil, com, stone, sand, houses, water, etc.

Thanks to the reconstruction of the investigated site by a 3D model, it's possible to create a multi-layer structure that permits to represent in one shot and in overlap all the data acquired during the mission. This approach simplifies the interpretation of data and allows us to filter the anomalies related to the specific morphology of the land (Lega *et al.* [2]).

Starting with industrial dumping registered in the province of Caserta, we use the geostatistical analysis to identify areas where the distribution of this dumping is "abnormal" with respect to the density of population and productive activities.

This may provide a guideline for the presence of potential illegal dumping.

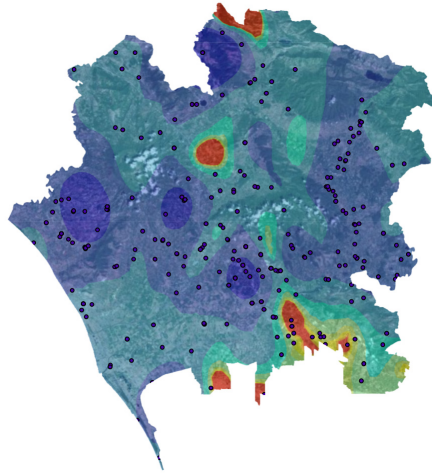


Figure 6: Example of a geostatistical map.

Geostatistical techniques rely on statistical model to exceed the uncertainty associated with spatial estimation and simulation. Geostatistical algorithms are incorporated in GIS.

4 Conclusions

If today a conventional investigation begins with the evidence of the crime that defines the scenario to be investigated, in the environmental field the investigation often begins with the evidence of the damage and not of the illegal polluting act. Often the consequences of pollution appear indeed in a different place or a long time after the polluting act has been committed and the correlation between the source and the damage depends from the morphology of the scenario and the physical phenomena that permit the transport of the pollutants. For these reasons the environmental police investigations must be realized alternating a bottom-up and top-down approaches (Lega *et al.* [2]).

In this paper we demonstrate the added value of the use of an integrated approach that combines several IT tools in the environmental police investigations.

IDES offer an integrated Geographic Information System (GIS) repository of information extracted and collected by various Government Bodies. The main novelty of this research lies in the integration between information extracted from remotely sensed images and information extracted from textual documents through a semantic analysis.

Thanks to a GIS multi-layer structure it's possible an advanced analysis of the investigated macro-area that permits to represent in one shot and in overlap all the data acquired and/or correlated.

This approach simplifies the interpretation of data and allows filtering the results.

The results of this last analysis will be a taxonomy-driven document searching and browsing system in which all documents are categorized and all relevant entities in documents (e.g. people, organizations, places and events) are extracted and represented by means of concept maps. Finally, these results: a) provide support to government bodies "end-users"; b) can be included in a GIS; c) can constitute the basis for the application of Artificial Adaptive Systems used in the domain of semantic analysis of unstructured information providing not obvious relationships between entities (e.g. criminal associations, people, places, activities, buildings, towns and dates).

Moreover, this IDES project provides an example of where law enforcement and research teams can collaborate on developing enhanced environmental protection methods.

References

- [1] Persechino G., Schiano P., Lega M., Napoli R.M.A., Ferrara C., Kosmatka J. "Aerospace-based support systems and interoperability: the solution to fight illegal dumping", Waste Management 2010, Tallin, Estonia, 2010.
- [2] M. Lega, C. Ferrara, J. Kosmatka, G. Persechino and R.M.A. Napoli "Thermal Pattern and Thermal Tracking: fingerprints of an environmental illicit", Quantitative Infrared Thermography (QIRT), Naples, Italy, 2012.
- [3] M. Lega, R M.A. Napoli. Aerial infrared thermography in the surface waters contamination monitoring.
- [4] G. Persechino, P. Schiano, M. Lega, R.M.A. Napoli, "Environment Monitoring performed by Advanced Hybrid Airship at low altitude", TIES 2009 – the 20th Annual Conference of The International Environmetrics Society – Conference 2009, Bologna, Italy, July 5–9, 2009.
- [5] M. Lega, R.M.A. Napoli, G. Persechino, P. Schiano, "EMPA project: the conquering of the third dimension in the ambient air Monitoring", Airnow 2009: Air Quality Conferences: Air Quality Forecasting and Mapping, Dallas, USA, 2–5 March, 2009.
- [6] M. Lega, R.M.A. Napoli – "A new approach to solid waste landfills aerial monitoring", Waste Management 2008, Granada (Spain) 2–4 June, 2008.
- [7] Câmara, G., Monteiro, A. M., Fucks, S. D. and Carvalho, "Spatial Analysis and GIS: A Primer", 2004.

