Human Centered Concepts for Exploration and Understanding of Satellite Images

Mihai Datcu[1] and Klaus Seidel[2]
[1]German Aerospace Center DLR
Oberpfaffenhofen, D-82234 Weßling, Germany
Phone: +49-8153-28 1388, Fax +49-8153-28 1444, Email: mihai.datcu@dlr.de
[2]Computer Vision Laboratory (CVL) - ETHZ
Gloriastr. 35, CH 8092 Zurich/Switzerland
Phone: +41-1-632 5284, Fax: +41-1-632 1199, Email: seidel@vision.ee.ethz.ch

***Abstract*-- The progress in information retrieval, computer vision and image analysis makes possible to establish very complete bases of algorithms and operators. A specialist in remote sensing or image processing has the tools now allowing him, at least in theory, to configure applications solving complex problems of image understanding. However, in reality, the Earth Observation data analysis is still performed in a very laborious way at the end of repeated cycles of trial and error. To this end we propose a novel advanced remote sensing information processing system, based on Human Centered Concepts, which implement new features and functions allowing improved feature extraction, search on a semantic level, the availability of collected knowledge, interactive knowledge discovery and new visual user interfaces.**

I. PRELIMINARIES

Madhok and Landgrebe introduced a novel process model for remote sensing data analysis [1]. The process is modular and analyst driven, it enables collaboration across disciplines and project elaboration and management. And, may be the most important characteristics of the process model, is the ability to concentrate on the user's skills and needs, more than on the deep mathematical background of the embedded algorithms, and to be in fact a "communication" media.

This article is a "letter" acknowledging the Madhok and Landgrebe work and sustain it with the presentation of a similar concept [2,4].

II. MOTIVATION

Modern imaging sensors, especially those aboard satellites, continuously deliver enormous amounts of data. The wide-spread of meter resolution images, is not only exploding the volumes of acquired data but also brings a new dimension in the image detail, thus fantastically growing the information content. These represent typical cases, where users need automated tools to explore and understand the contents of large and high complexity images. Without new theoretical concepts, methods and novel technological support, the existing volume of data prevents any systematic exploitation of Earth Observation (EO) data.

The progress in information retrieval, computer vision and image analysis makes possible to establish very complete bases of algorithms and operators. A specialist in remote sensing or image processing has the tools now allowing him, at least in theory, to configure applications solving complex problems of image understanding. However, in reality, the EO data analysis is still performed in a very laborious way at the end of repeated cycles of trial and error, without reaching the desired degree of adaptability and robustness. Also, the existing methodologies are prohibitive for the analysis of large volumes of data, and unable to discover relevant causalities or associations among the objects in the data sets, neither to reuse information or knowledge set up in other applications.

There is a strong need to have environments for development allowing a user to configure an application of image understanding, or to formulate information retrieval, without knowing the operators needed to represent the image information content at signal level, instead, the system can learn the user conjecture corresponding to interpretations which the user makes and to the concepts he is using in connection with this same image. This calls up synergetic integration of stochastic modelling of images, knowledge discovery and artificial intelligence, handling of formal representations, semantic representations of images and multi-agents systems implementing emerging functions.

III. PRINCIPLE AND METHODOLOGY

The decreasing cost of fast computers and large storage devices and the increasing speed of networks enable the management of large amounts of EO data, information and knowledge. However the capacity to understand data is further rather limited. Thus, an important and urgent priority is to develop new methodologies to metamorphose the computer from a machine used for solving difficult problems, into effective means for communicating with other people and machines. The concept to implement such a system, is to split the task in the most appropriate way to exploit the computer skills and to adapt to the human behavior.

The concept was applied by implementing the Knowledge Driven Image Information Mining (KIM) system and being developed in collaboration with Advanced Computer Systems (ACS) under ESA contract [2]. The basic principle is to split the processing in two parts, with the following characteristic:

Data driven processing

- off-line
- computational/time intensive task
- integration of libraries of novel algorithms for image parameter estimation
- the estimated parameters are generic, they represent characteristics of stochastic processes, and are application independent
- the processing is an objective coding of the image information content

User driven processing

- on-line
- interactive
- fast image analysis tool, for classification, data fusion, and object detection.
- real-time at user reaction time
- a machine learning process
- Human Machine Interaction as a dialog based on positive and negative examples
- communication via visual and semantic channels
- relevance feedback
- knowledge based and driven
- adaptation to the user conjecture

Thus the computer helps by determining the options, suggests solutions or results, selects and suggests actions and human may approve it, informs human about the course of actions, human may direct actions by abstract expression of interest.

Computer does intensive numeric tasks, but also learns form humans and later can help other users to take decisions or actions. However, the human is at the end in the centre of the task, by his continuous interaction, and by the knowledge he transfers to computers.

- Modelling the image formation process
- To arrive at a satisfactory solution adapted to the user's needs, it is necessary to describe and analyze the image formation and the data acquisition processes.

A simplified image formation process is presented in Fig. 1. Images are recorded in order to obtain useful information about a scene, or an environmental process. The image formation process assumes illumination of the scene with given radiation: light, laser, microwaves, etc. The reflected radiation contains information about the scene geometry, like shape, shadows, or textures, and also scattering information about the physical quantities, like gas concentrations, humidity, or speed of the wind. The reflected radiation, called cross-section, propagates to a sensor. The propagation can be affected by disturbances, e.g. refraction or absorption by the atmosphere. The sensor records the incident field and transforms it into an electrical signal, later visualized as an image.

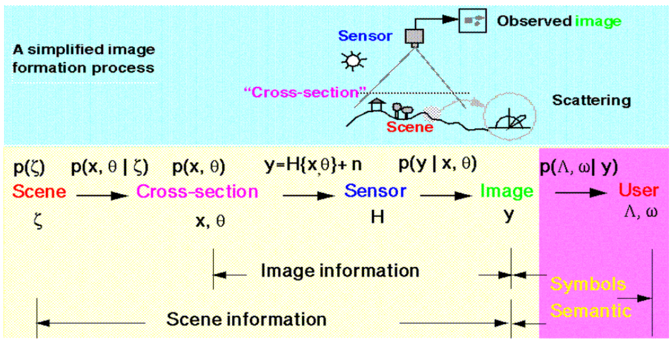

Figure 1. A simplified image formation process and the different levels of information models and the chain of causalities.

Thus, the model for the image formation process can be summarized as the chained causalities [3]. Where $p(\zeta)$ represents the scene model, ζ is the parameter vector characterizing the 3-dimensional geometry and radiometry of the scene; $p(x, \theta)$ is the cross-section model, x representing the intensity process of the *ideal* image and θ is the parameter vector describing its 2-dimensional structures. Both x and θ depend on ζ. This causality is modeled by the conditional process $p(x, \theta \mid \zeta)$. The propagation and the sensor are modeled by an operator H and the stochastic noise n, thus the observed image is $y=H\{x,\theta\}+n$.

Based on the above considerations, and assuming that the observed data is recorded by a certain sensor, we identify two types of *classical* problems:

- Image understanding, i.e. given an observed image, reconstruct and recognize the structures in the undisturbed image, the cross-section, e.g. image segmentation, edge or texture recognition, and
- Scene understanding, i.e. given an observed image, reconstruct the scene identity, e.g. shape from shading or interferometric methods for land surface reconstruction.

Both are information extraction problems of high complexity, and computational intensive. They are solved either as a parameter estimation or as an inverse problem. Their solutions are data driven, and presented many times in numeric form, thus requiring an additional level of interpretation adapted to the user conjecture.

Images contain quantitative, objective information, as acquired by an instrument. However, its perception and understanding is in form of symbols and semantics in a certain semiotic context.

Because *people maintain and operate discrete units of linguistic expression at several levels of abstraction, and people use the information from their self-monitoring and environmental-monitoring to dynamically modify and change their cognitively prepared sequence of linguistic expressions as they continue to execute sequentially ordered discrete expressions* [5], the image interpretation needs to adapt to users, to the symbols they are able to recognize and their domain specific semantics. The interpretation of EO data requires not only information retrieval for understanding of Earth cover structures, but, at a higher level, needs the aggregation with existing bodies of knowledge specific to the application fields. Thus, from this perspective, the interpretation of EO data is a knowledge driven task. The knowledge consists in the ensemble of existing information, known causalities and other type of associations between information and concepts. During the interpretation process new information and causalities are discovered, thus we have a dual process of knowledge acquisition. To formalize the knowledge both in knowledge driven interpretation and in the knowledge acquisition we need to define a hierarchy from the point of view of information categories, as proposed in the following diagram (Fig. 2).

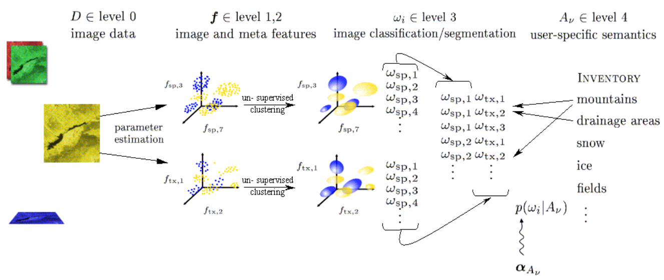

Figure 2. The hierarchy of information used to enable information extraction and knowledge representation adapted to the user needs.

The levels of information enable a simple and systematic representation of the image content.

The quasi-complete image content description using features extracted as the basic image primitives, i.e. spectral signatures, texture parameters, geometrical characterization as geometrical features, shape descriptors and elements of topology.

The completeness of the description is augmented with metainformation, i.e. the type of model used for the extraction of the primitive features, and the scale at which the analysis was done. The multiscale approach is one of the prerequisites of the quasi-complete image description.

The clustering is applied for each feature space apart for all images in an archive, thus is simultaneously a data reduction and generalization by similarity over the image space. The cluster model is key information to be included in a data base catalogue.

At the semantic level the meaning of image objects or structures is obtained by an interactive learning process fusing the relevant information extracted from one sensor image data set or multisensor data and information.

IV. KNOWLEDGE DRIVEN INFORMATION RETRIEVAL

To this end we propose a novel advanced EO information processing system which implements new features and functions allowing improved feature extraction, search on a semantic level, the availability of collected knowledge, interactive knowledge discovery and new visual user interfaces [2].

The Human Centered Concepts, knowledge driven methods and the associated data management are changing the paradigms of user / data interaction by providing simpler and wider access to EO data. Emerging technologies could now provide a breakthrough, permitting automatic or semi-automatic information extraction supported by intelligent learning systems.

Images and other non-visual EO data are realizations of stochastic processes, thus information extraction relies on the accuracy of the models used for the information content. An important requirement is the robustness of the derived models for very large volumes of data. Models must describe a broad data variability, and variety of artifacts, e.g. noise, sensor distortion, processing artifacts. A quasi-complete representation of the image content is used in order to build systems that are independent of any application specificity, and to enable its open use for almost any scenario.

Based on the extracted features derived from each model and from all images in an archive, a new abstract hierarchic level is defined: the set of signal classes describing characteristic groups of points in the feature space of the different models. This is a code, i.e. a vocabulary of characteristic signal classes, which is valid across all images. It represents the key condensed and generalized information to be used further for the mining process.

Based on this objective representation, a link can be created from the subjective user interest to the signal classes. This link is modeled probabilistically, and is a two level hierarchical model driven by the user. The model learns the user semantics, in a certain conjecture, and creates a meaningful structure by fusing clusters obtained in different feature spaces. A data base index stores both the identity of the image pixels (spatial index) and of the image (image index). In this way the learning process establishes associations between all the hierarchic levels of information representation.

The hierarchy of information in conjunction with learning process results in knowledge structures, as the example in Fig. 3.

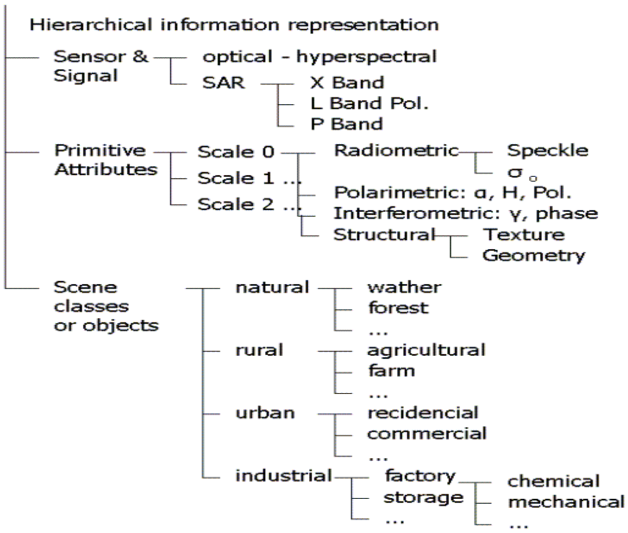

Figure 3. Knowledge acquired based on the hierarchical information representation.

The concept is based on statistical models and machine learning methods which "explore and explain" the image information. The user can build an image semantics in a free-form text manner as an visual language for image processing. The resulted associated semantic content is difficult to derive using standard computer vision methods. Basically, information comparable to the results gained during the interactive user training will be used as existing knowledge for the control of the retrieval process.

The concept enables automatic categorization and ontology learning, symbolic representations for abstract bodies of knowledge, and cognitive visualization techniques. The methods allow the elaboration of sophisticated communication and computational concepts designed to amplify human cognitive and perceptual capabilities in relation with human-machine interaction. The system enables sharing knowledge.

V. SYSTEM ARCHITECTURE

The concept we elaborated for information retrieval and understanding of remote sensing images is based on a hierarchical Bayesian learning model and is implemented in a system with two levels:

1. interactive training of the desired image content in terms of image features, and user semantics, followed by,
2. generalization and knowledge formalization.

Both levels make use of pre-extracted image parameters. For computational complexity reasons, the image parameters are extracted off-line at the time of data ingestion in the archive. The parameters are extracted for different image scales. In the next processing step the image parameters are clustered, and further a signal content index is created using the cluster description, the scale information, and the type of stochastic model assumed for the image parameters. A Bayesian hierarchical decision algorithm (naive Bayes) allows a user to visualize and to encapsulate interactively his prior knowledge of certain image structures and to generate a supervised classification in the joint space of clusters, scales, and model types. The user is enabled to attach his meaning to similar structures occurring in different images, thus adding a label in the archive inventory. This label is further used to specify queries. Fig. 4 presents the logical diagram of the system.

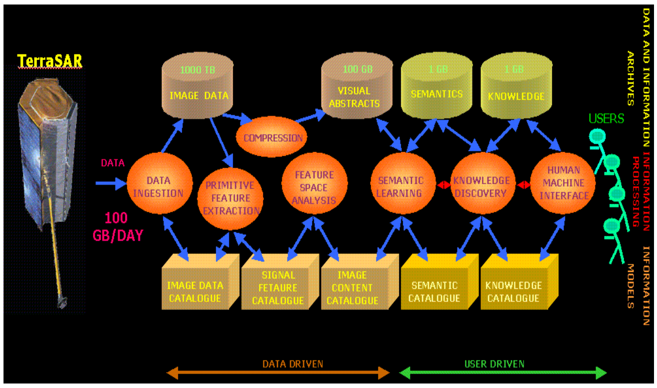

Figure 4. The system consists of two main modules: 1) the first, data driven, is responsible for data acquisition, and preprocessing for feature extraction, 2) the second, user driven, solves the information fusion and interactive interpretation operations and it supports knowledge based functions.

The information processing follows the model of our advanced communication system. Thus, it is splitted up in two parts: the first one is the data-driven processing, corresponding to the objective modelling of the communication channel, and the second one is user-driven, corresponding to the subjective part of the channel model.

The core of the system is the database that will host the different catalogues, from image data, features, semantics, to knowledge catalogues, following a progressive distillation of information. The database has specific design and customizing in order to fully exploit advanced concepts, such as semantics and knowledge management that are not included in standard spatio-temporal databases.

The overall system is based on a client-server structure and is intrinsically designed to support web-based remote access which enables communication at different levels.

VI. CASE STUDIES

The classical task in the interpretation of remote sensing data generally assumes that the source of information is a collection of images. The methods applied for information extraction are image enhancement, image segmentation, feature extraction, fitting physical models to the data, etc. The explosion in sensor technology, both high resolution and frequently repeat paths, results in an increasing number of large multimission remote sensing images available for an application.

The methods for understanding the image content we developed are intended to overcome the informational bottle-neck of classical approaches and also to stimulate the user in finding new scenarios for data interpretation, e.g. *find all sites containing cities surrounded by forest*. The novel functions presently provided by the system are:

- Search by Scale - find all images with relevant structures at specified scales,
- Image Content Search - find all images containing a specified structure or object, e.g. lakes, cities, types of forest, etc.,
- Cover-Types by Application Area - the same as the previous, but the catalogue inputs are clustered by application interests, e.g. Meteorology, Hydrology, Geology,
- Cover-Type Training - interactive generation of new catalogue inputs in terms of image content. This is an information mining function, allowing the exploration of unknown image content in large image collections.

These functions are demonstrated by several projects, using a broad variety of image data, and addressing very different users.

CASE A. Estimation and classification of mined areas from high resolution airborne images

The project demonstrates tools for the use of airborne images to help to reduce the suspected mined areas. Millions of mines are infesting over seventy countries on all continents. They still maim or kill decades after they were laid. They have enormous and long-term extremely negative effects on a country.

The main activities carried in the project have been:

1. the data collection, planning and performing fights over several suspect areas with high-resolution SAR, and hyperspectral sensors, together with a ground-truth mission the data processing.
2. Data processing for geometric and radiometric correction, co-registration, and extraction of a set of quasi-complete set of image primitive features, e.g. radiometric, spectral, or polarimetric signatures, texture parameters at different scales.
3. Ingestion of the image data and extracted features in an integrated environment to support and make the work of analysts easier

User interactive derivation of the suspected minefields

The human centered procedure proposed was:

1. Selection of learning data sets used to develop and tune the procedures, the methods and the algorithms.
2. Selection of validation sites to be used to validate the procedures, the methods and the algorithms.
3. The system is operated by the domain expert knowledge and the ground truth specialists.

The following table presents a summary of the data processed and ingested in the KIM system [2].

Sensor name	Spectral resolution	Spatial resolution	Swath width	Radiometric resolution
Daedalus 8 bits line-scanner	11 channels	<1 m	450 m	0.380 & 2.7 m
SAR L-band 6 bits	HH-VV-HV-VH	2.5 m	3.5 km < 2 dB	16 bits
SAR X-band	VV	2 m	3.5 km < 2 dB	16 bits

The following generic features have been extracted:

- SAR despeckled images
- SAR image textures at resolutions 2 m and 4 m
- SAR polarimetric signatures

The delimitation using airborne imagery of areas suspected to be mined is not an easy task. The ontology of the sensor data and signal classes correspond to the KIM strategy.

In this assumption the Domain Knowledge can be structured as:

- Sensor and phenomenology of image formation
- Mitigation: de-mining
- Military: mining

Sensor and phenomenology of image formation

The detection of mined fields, compared with other scene classification tasks, has very high complexity. The difficulties are in the heterogeneous nature of the observed scene, thus requiring both sensor and information fusion. The sensor fusion aims at the observation of physical and geometrical scene characteristics, trying to capture complementaries. The information fusion has as goals first a fission of the observed signals in a quasi-complete set of primitive attributes, and their aggregation (fusion) such to be able to detect all, or at least, all possible, scene structures. The fission is based on utilization of a library of models for the data, capturing the main attributes: radiometric or polarimetric, structural, and multiscale.

The large number of models, and the huge dimensionality of the feature space, mainly for hyperspectral data, makes the selection, by classical methods, of the appropriated data sets difficult. The task can be solved by utilization of the categories of information stored in KIM (the horizontal ontology). The expert user - sensor domain - is helped to collect the information under generic categories, at semantic level, thus communicable to users in other domains.

Mitigation domain knowledge: de-mining

The expert in the de-mining action is using a special set of objects, structures and interrelations among them, to express a degree of believe in the fact that a field is mined or not. In order to automatize such a process in the case of utilization of image data, it is needed to detect and translate the symbols he is using in signal attributes, and more to communicate the results to a non expert user. The key step is the identification of the domain knowledge ontology. In this case the following concepts are relevant:

- bare soil
- agriculture fields used and not-used
- roads fields used and not-used
- farms fields used and not-used

The domain knowledge in the field of sensor data shall be transferred to the appropriate level to enable the communication.

Military domain knowledge: mining

To increase the degree of goodness of classification of mine fields, mining domain knowledge should be used. As example, the important categories can be:

- road
- bridge
- power line
- access to narrow river site

The translation of these concepts in image symbols, i.e. structures and associated characteristic primitive attributes is a modality of knowledge communication.

The following images presents a selection of instances extracted from an interactive session aiming at detection of fields suspected to be mined using high resolution hyperspectral and Polarimetric L Band SAR images (Fig. 5, Fig. 6, Fig. 7).

The user was supported by the system who recorded the sequence of actions, selected automatically the weight of the information extracted from optical and SAR data, and the type of information, either radiometric or structural.

It is known that the distinction between the perception of information as signals and symbols is generally not dependent on the form in which the information is presented but rather on the conjecture in which it is perceived, i.e. upon the hypothesis and expectations of the user. Thus, the new technology requires a different attitude of the user of remote sensing data for searching or interpreting the image content.

CASE B. Landing fields for small airplanes

The user has at his disposal a collection of 66 high resolution optical images (aerial photographs) and searches for areas where landing with a small airplane would be possible. The prior knowledge the user implicitly is using is a generic description of a landing field: a flat, smooth, solid and reasonable large area. This description, by an interactive learning process, is translated in image (signal) texture and reflectance features, which are generalized over the whole image collection. In Fig. 8 an example of the result of such a search from the above mentioned demonstrator database is presented.

Figure 5. Site suspected to contain mined areas.
SAR L band polarimetric image.

Figure 6. Site suspected to contain mined areas.
Hyperspectral image.

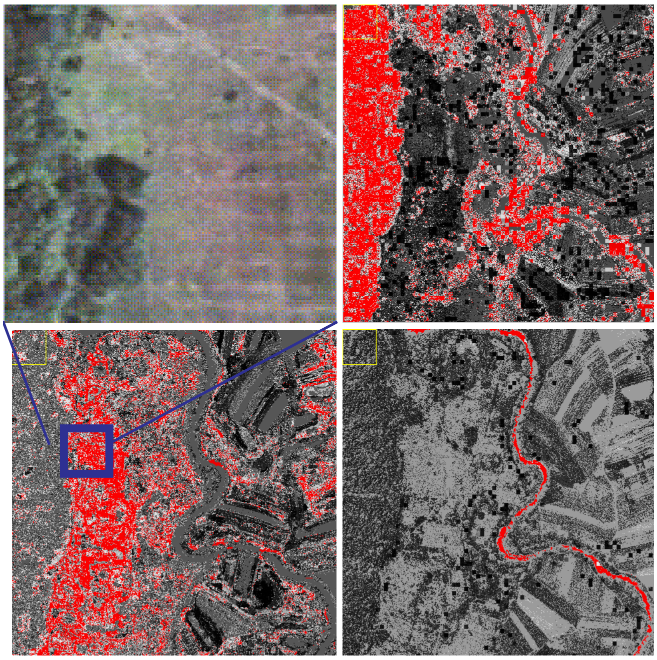

Figure 7. The images present the result of user interactive and knowledge driven task solving for detection of areas of risk to be mined. Left: in red area of high risk and image detail for confirmation. Right: top-detection of trees, bottom-detection of river. All classifications are obtained by fusion of information extracted from hyperspectral and PolSAR images.

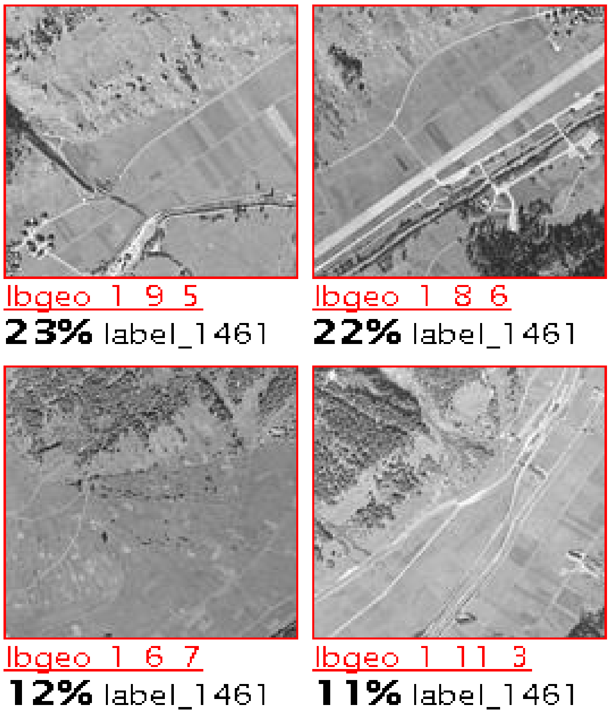

Figure 8. The images present the result of exploration of areas appropriate for landing of a small aircraft. The system was able to select three images presumable correct, however the probabilistic nature of the search resulted also in an answer unlikely to be correct (he bottom-left image).

CASE C. Study of dynamics of inhabited areas

The study of dynamics of inhabited areas requires in a preliminary step the detection of build-up areas. The example in Fig. 9 shows the result of a query combined with a classification of SAR (X-SAR) images from an collection of 110 scenes of 2048x2048 pixels. The result is obtained by interactive learning the behavior of a build-up area using the estimated SAR backscatter and density of targets.

Due to the flexibility of the system the number of possible scenarios is very large. The reader is encouraged to experiment the above mentioned on-line demonstrator http://www.acsys.it:8080/kim, or http://isis.dlr.de/mining.

VII. CONCLUSIONS

The field of human centered computing reaches the maturity for integration in EO applications, helping the understanding and exploration of image data is a highly complex task. We developed a new concept for image information mining and demonstrated it for a variety of remote sensing applications. Image information mining opens new perspectives and a huge potential for information extraction from remote sensing images and the correlation of this information with the goals of applications. The user has fast, interactive and friendly access directly to the information content of the images, can interactively add value and evaluate the appropriateness of a sensor acquisition and the feasibility of data for a certain application.

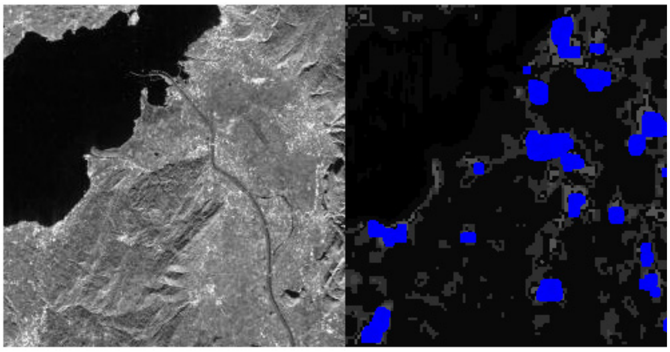

Figure 9. Example of detection of buildup regions using SAR (X-SAR) observations. The result of the query the buildup regions are marked, thus the user can very fast and easily pre-evaluate the selected images for further detailed interpretation.

ACKNOWLEDGMENTS

This work was supported by ESA contracts KIM: Knowledge driven Image Information Mining in Remote Sensing archives, and KES Knowledge Enabled Services. The KIM system is developed by ACS Advanced Computer Systems S.p.A, Rome. Part of data and experiments are in the frame of the EC project SMART.

REFERENCES

[1] V. Madhok, D. A. Landgrebe, 2002, A Process Model for Remote Sensing Data Analysis, IEEE Trans. on Geoscience and Remote Sensing, ISSN 0196 2892, Vol. 40, pp. 680-686

[2] M. Datcu, K.Seidel, S. DElia, P. G. Marchetti, 2002, Knowledge-driven Information-Mining in remote sensing image archives, ESA Bulletin 110, pp. 26-35. Available on http://esapub.esrin.esa.it/bulletin/bullet110.htm/

[3] M. Datcu, G. Schwarz, 2000, The Application of Forward Modelling as a Data Analysis Tool, in Information Processing for Remote Sensing, Ed. C.H. Chen, World Scientific Publishing Company Pte Ltd, pp. 51-65.

[4] M. Schröder, H. Rehrauer, K. Seidel, M. Datcu, 2000, Interactive learning and probabilistic retrieval in remote sensing image archives, IEEE Trans. on Geoscience and Remote Sensing, ISSN 0196 2892, Vol. 38, pp. 2288-2298.

[5] D. C. McFarlane, 1998, Interruption of People in Human-Computer Interaction, Ph.D. Thesis, The George Washington University. Available on http://www.aic.nrl.navy.mil/hail/McFarlane.1998.pdf