The new generation of high resolution imaging satellites acquires huge amounts of data which are stored in large archives. The state-of-the-art systems for data access allow only queries by geographical location, time of acquisition or type of sensor. This information is often less important than the content of the scene, i.e. structures, objects or scattering properties. Meanwhile, many new applications of remote sensing data are closer to computer vision and require the knowledge of complicated spatial and structural relationships among image objects.

We are creating an intelligent satellite information mining system, a next generation architecture to help users to gather rapidly information during courses of actions, a tool to add value and to manage the huge amount of historical and newly acquired satellite data-sets by giving to experts access to relevant information in an understandable and directly usable form and to provide friendly interfaces for information query and browsing.

MOTIVATION

The most recent example which motivates the development of information mining technology is the Shuttle Radar Topography Mission - SRTM (http://www.dfd.dlr.de/srtm/index.html). SRTM is an important milestone in the history of remote sensing. In a few days it collected about 18 terabytes of radar measurements which allow scientists to virtually reconstruct a 3 dimensional model of 80% of the continental surface. The virtual Earth is reconstructed as a mesh of 30 m spacing, and is accompanied for each point by a measure of the reflected energy of the radar signal, the Synthetic Aperture Radar image. The data becomes an important reference for comparisons and correlations with older and future satellite recordings or other Earth observation data. SRTM is a status 2000 for many applications ranging from geology, tectonics, hydrology, cartography, to navigation and communication. The data acquired by the SRTM mission is a huge thesaurus which requires a careful management and innovative exploitation.

LOOKING FOR A NEEDLE IN A BUNDLE OF HAY

In recent years our ability to access and store large quantities of data has greatly surpassed our ability to meaningfully extract the information from the data. This has led to concerted efforts to develop new concepts and methods to deal with large data sets: query by image content, data mining, knowledge discovery, information visualization. A broad range of techniques was developed to deal either with particular data types, like text, numerical records, or voice signatures, and also with heterogeneous data types, e.g. combining video and sound. One of the most complex tasks still remaining is the access of image information. Image data information systems require both database and visual capabilities, but a gap exists between these systems. The theory of databases, until recently, did not deal with multi-dimensional pictorial structures, and vision systems do not provide database query capabilities. Most existing image databases have been created using some extensions of the relational data model. Meanwhile, with the explosion of multimedia systems, scientific applications, and especially the growing interest in spatial data (GIS, remote sensing images, digital cartography), a new dimension came to the problems of accessing the information content in a database.

In addition to the operational state of the art archive and database systems we develop image information mining systems. The objective of information mining is to extract essential information that is implicitly stored in large data archives. The objective of information mining is to extract essential information that is implicitly stored in large data archives. We are creating an intelligent satellite information mining system: a next generation architecture to help the user to gather relevant information rapidly and a tool that can manage and add value to the huge amounts of historical and newly acquired satellite data-sets [1].

CONCEPT AND SYSTEM

The concept we elaborated for information mining and retrieval from remote sensing image archives is based on a hierarchical Bayesian learning model and is demonstrating a system with two levels:
1) interactive training of the desired image content in terms of image features, followed by,
2) query by image content using as content the image features defined in step 1.

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 [4]. 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 [2]. 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. 1 presents the logical diagram of the system.

The system integrates several original solutions, e.g. feature extraction from SAR images, texture features estimation in presence of noise [3], hierarchy of information representation for image content characterization [1], supervised classification and interactive training using Bayes networks [2]. Other solutions implemented in the system following the most advanced results obtained until now are feature extraction from optical multispectral data, clustering, and part of the user adaptation techniques.

This concept was implemented and successfully demonstrated with an experimental system, see http://isis.dfd.dlr.de/mining/ and http://www.vision.ee.ethz.ch/~rsia.

FUNCTIONS

The classical task in the interpretation of remote sensing data generally assumes that the source of information is just one image. 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 pass, results in an increasing number of large multimission remote sensing archives. Thus, the problem of image content extraction should be reformulated taking into consideration the new source of information: the image archive.

The methods for searching 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 images 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 archives.

APPLICATIONS

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. For exemplification we present two scenarios.

Scenario 1: The user has at his disposal a collection of 66 high resolution optical images (areal 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 hole image collection. In Fig. 2 an example of the result of such a search from the above mentioned demonstrator database is presented.

Scenario 2: The study of dynamic of inhabited areas requires
in a preliminary step the detection of build-up areas. The example in Fig. 3 shows the result of a query combined with a classification of SAR (X-SAR) images from an archive of 110 scenes of 2048x2048 pixels. The result is obtained by interactive learning the behaviour 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 online demonstrator.

CONCLUSIONS

The field of data mining reaches the maturity for integration in commercial products, however mining 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.

REFERENCES


