Information Mining in Remote Sensing Image Archives—Part A: System Concepts

Mihai Datcu, Herbert Daschiel, Andrea Pelizzari, Marco Quartulli, Annalisa Galoppo, Andrea Colapicchioni, Marco Pastori, Klaus Seidel, Pier Giorgio Marchetti, and Sergio D'Elia

***Abstract*—In this paper, we demonstrate the concepts of a prototype of a knowledge-driven content-based information mining system produced to manage and explore large volumes of remote sensing image data. The system consists of a computationally intensive offline part and an online interface. The offline part aims at the extraction of primitive image features, their compression, and data reduction, the generation of a completely unsupervised image content-index, and the ingestion of the catalogue entry in the database management system. Then, the user's interests—semantic interpretations of the image content—are linked with Bayesian networks to the content-index. Since this calculation is only based on a few training samples, the link can be computed online, and the complete image archive can be searched for images that contain the defined cover type. Practical applications exemplified with different remote sensing datasets show the potential of the system.**

***Index Terms*—Content-based image retrieval (CBIR), image information mining, information extraction, statistical learning.**

I. Introduction

DURING THE LAST decades, the imaging satellite sensors have acquired huge quantities of data. Optical, synthetic aperture radar (SAR), and other sensors have delivered several millions of scenes that have been systematically collected, processed, and stored. The state-of-the-art systems for accessing remote sensing data and images, in particular, allow only queries by geographical coordinates, time of acquisition, and sensor type [3]. This information is often less relevant than the content of the scene, e.g., structures, patterns, objects, or scattering properties. Thus, only few of the acquired images can actually be used [6]. In the future, the access to image archives will even become more difficult due to the enormous data quantity acquired by a new generation of high-resolution satellite sensors. As a consequence, new technologies are needed to easily and selectively access the information content of image archives and finally to increase the actual exploitation of satellite observations [7].

Manuscript received December 3, 2002; revised May 25, 2003.

M. Datcu, H. Daschiel, and M. Quartulli are with the Remote Sensing Technology Institute, German Aerospace Center (DLR), D-82234 Oberpfaffenhofen, Germany.

A. Pelizzari is with Appunto Solucoes Informaticas, Lda., 1000-112 Lisbon, Portugal.

A. Galoppo, A. Colapicchioni, and M. Pastori are with the Earth Observation Division, Advanced Computer Systems S.p.A. (ACS), 00139 Rome, Italy.

K. Seidel is with the Computer Vision Laboratory, Swiss Federal Institute of Technology (ETH), CH-8092 Zürich, Switzerland.

P. G. Marchetti and S. D'Elia are with the Earth Observation Applications Department, European Space Agency Directorate of Earth Observation Programmes, ESRIN, 00044 Frascati, Italy.

Digital Object Identifier 10.1109/TGRS.2003.817197

For many years, it has been known that classical image file text annotation is prohibitive for large databases. The last decade is marked by important research efforts to develop content-based image retrieval (CBIR) concepts and systems: query by image and video content (QBIC) [9], VisualSEEK [16], Virage [1], etc. Images in an archive are searched by their visual similarity with respect to color, texture, or shape characteristics. While image size and information content are continuously growing, CBIR was not any more satisfactory, and thus, region-based information retrieval (RBIR) has been developed [17]. Each image is segmented, and individual objects are indexed by primitive attributes like color, texture, and shape. Thus, RBIR is a solution to deal with the variability of image content.

However, both CBIR and RBIR have been computer-centered approaches, i.e., the concepts hardly allowing for any adaptivity to user needs. Furthermore, the image retrieval systems have been equipped with relevance feedback functions [4], [13]. The systems are designed to search images similar to the user conjecture. The algorithms are based on analyses of the probabilities of an image to be the search target. A feedback that takes this part into account is introduced.

Another interesting approach was developed. It is based on a learning algorithm to select and combine feature grouping and to allow users to give positive and negative examples. The method refines the user interaction and enhances the quality of the queries [11].

Both previously mentioned concepts are first steps to include the user in the search loop; they are information mining concepts. They are also methods in the trend of designing human-centered systems.

In addition to the operational state-of-the-art archive and database systems, we have developed a knowledge-driven information mining (KIM) system. KIM is a next-generation architecture to support the man–machine interaction via the internet and to adaptively incorporate application-specific interests. In the system, the user-defined semantic image content interpretation is linked with Bayesian networks to a completely unsupervised content-index. Based on this stochastic link, the user can query the archive for relevant images and obtains a probabilistic classification of the entire image archive as an intuitive information representation.

The paper is organized as follows. After presenting the basic concept behind KIM of hierarchical information modeling in Section 2, we will show the applied methods for primitive image feature extraction in Section 3. From the extracted parameters, we will obtain a global content-index by an unsupervised

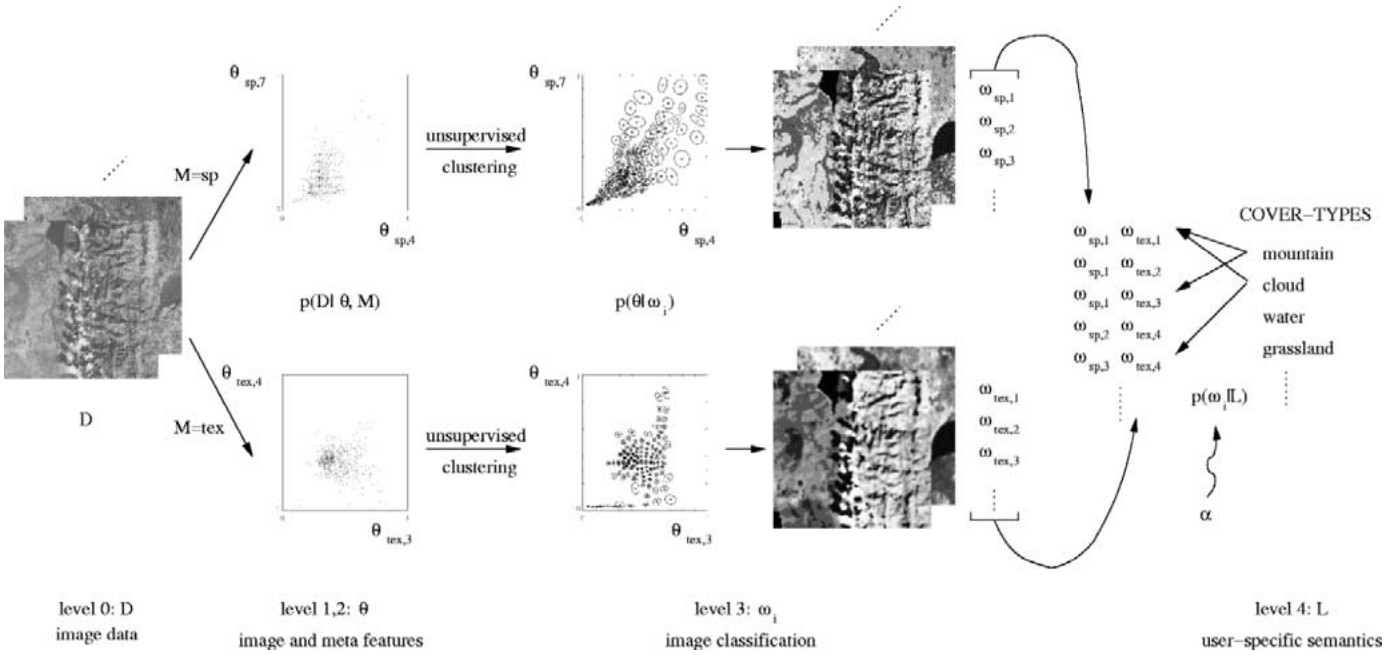


Fig. 1. Hierarchical modeling of image content and user semantic. First, primitive image features $\boldsymbol{\theta}$ and metafeatures are extracted from image data D, based on different parametric signal models M. Through an unsupervised clustering of the features, we obtain a vocabulary of signal classes ω_i. User-defined semantic cover types L are linked to the signal classes ω_i by probabilities $p(\omega_i \mid L)$ using Bayesian networks.

across-image clustering and subsequentely the catalogue entry as described in Section 4. How this content-index can be associated with user-specific interests by interactive learning will be pointed out in Sections 5 and 6. By making this association, all images in the archive that have the trained semantic information can be queried as shown in Section 7. The organization and connections of user, interactive training, and image query in the DBMS are demonstrated in Section 8 where we emphasize the information transmission between different parts of the system. After Section 9, a section of practical applications, we will conclude with a short summary.

II. Application-Free Hierarchical Modeling of Image Information

In order to build a system that is free of the application specificity, to enable its open use in almost any scenario and to accommodate new scenarios that are required both by the development of sensor technology and the growing user expertise, we start from an application-free hierarchical modeling of the image content (Fig. 1).

The concept of information representation on hierarchical levels of different semantic abstraction is based on a five-level Bayesian learning model [15].

- First, primitive image features $\boldsymbol{\theta}$ (level 1) and metafeatures (level 2) are extracted from image data D (level 0) using different signal models M (Section 3).
- Next, by a completely unsupervised clustering of the preextracted image parameters $\boldsymbol{\theta}$, we obtain a vocabulary of signal classes ω_i separately for each model M (Section 4).
- Finally, user-specific interests, i.e., semantic cover type labels L (level 4), are linked to combinations of these vocabularies by simple Bayesian networks (Section 5).

Levels 1–3 are obtained in a completely unsupervised and application-free way during data ingestion in the system. The information at level 4 can be interactively defined by users with a learning paradigm (Section 6) that links (objective) signal classes ω_i and user (subjective) labels L.

III. Extraction of Primitive Image Features

Automatic interpretation of remote sensing images and the growing interest for image information mining and query by image content from large remote sensing image archives relies on the ability and robustness of information extraction from the observed data. We focus on the modern Bayesian way of thinking and introduce a pragmatic approach to extract structural information from remote sensing images by selecting those prior models that best explain the structures within an image. On the lowest level, the image data D, we apply stochastic models to capture spatial, spectral, and geometric structures in the image. These models are given as parametric data models $p(D \mid \boldsymbol{\theta}, M)$ and assign the probability to a given realization of the data D for a particular value of the parameter vector $\boldsymbol{\theta}$.

A. Optical Images

To apply parametric stochastic models [8] in order to extract primitive image features, the data are understood as a realization of a stochastic process. The Gibbs–Markov random field (GMRF) family of stochastic models assumes that the statistics of the gray level of a pixel in the image depends only on the gray levels of the pixels belonging to a neighborhood with a restricted dimension. The probability of the gray level of the pixel x_s is given by

$$p(x_s \mid \partial x_s, \boldsymbol{\theta}) = \frac{1}{Z_s} \exp(-H(x_s \mid \partial x_s, \boldsymbol{\theta})) \qquad (1)$$

where Z_s acts as a normalization factor given by the sum over all the possible states for the pixel x_s. Assumptions have to be made for the functional form of the energy function H. In this approach, we use an autobinomial model with its energy function

$$H(x_s \,|\, \partial x_s, \boldsymbol{\theta}) = -\ln \binom{G}{x_s} - x_s \eta \tag{2}$$

$$\eta = a + \sum_{ij} \theta_{ij} \frac{x_{ij} + x'_{ij}}{G} \tag{3}$$

as the joint influence of all neighbors weighted by the elements of the parameter vector $\boldsymbol{\theta}$. Each element θ_{ij} of the parameter vector describes the interaction between the pixel x_s and the pair x_{ij}, x'_{ij}, while the parameter a represents a sort of autointeraction. G indicates the maximum gray value, e.g., 255 for an eight-bit image.

A fitting of the model on the image is performed in order to obtain the best fitting parameters. For the estimation a conditional least squares (CLS) estimator [14] is obtained by

$$\hat{\boldsymbol{\theta}}_{\mathrm{CLS}} = \arg\min_{\boldsymbol{\theta}} \sum_s \left(x_s - \sum_{x_s=0}^{G} x_s p(x_s \,|\, \partial x_s, \boldsymbol{\theta}) \right)^2. \tag{4}$$

The evidence of the model, e.g., the probability of the model given the data, can be calculated by

$$p(M \,|\, D) = \frac{p(D \,|\, M) p(M)}{p(D)} \tag{5}$$

where the probability of the data D can be obtained via the integration

$$p(D \,|\, M) = \int p(D \,|\, \boldsymbol{\theta}, M) p(\boldsymbol{\theta} \,|\, M)\, d\boldsymbol{\theta}. \tag{6}$$

From the estimated parameters, we derive several features to describe the image content: the norm of the estimated parameters $|\hat{\boldsymbol{\theta}}|$ as the strength of the texture, the estimate of the variance $\hat{\sigma}_M^2$ as the difference between signal and model energy [12], the evidence of the model M, (5), and the local mean of the estimation kernel (Fig. 2).

To describe the image content by using the spectral properties, we do not have to explicitely estimate the parameter vector $\boldsymbol{\theta}$. Instead, we can directly assign the individual spectral channels (after a normalization) to elements of the vector $\boldsymbol{\theta}$, e.g., the six spectral channels in the visible spectrum of the Landsat Thematic Mapper (TM) result in a six-dimensional vector $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_6\}$.

B. SAR Images

In KIM, there exist coregistered optical and SAR images. To include radar information in the retrieval process for an entire exploitation of the image archive and to enable the mining of multisensor data for sensor qualification, we have to extract content-based image parameters from SAR data, too.

The information extraction is achieved as a model-based Bayesian approach [8], [18]. The system models and reconstructs an estimated backscatter image that is free of speckle noise, while still completely preserving its most important structural information.

Since the system takes both the statistics of the noisy and the noise-free data in a Bayesian framework into account, the choice of an appropriate model for the estimated backscatter image plays an important role and affects the obtained results directly. In order to filter out speckle, the Bayesian formula

$$p(x \,|\, y, \boldsymbol{\theta}) = \frac{p(y \,|\, x) p(x \,|\, \boldsymbol{\theta})}{p(y \,|\, \boldsymbol{\theta})} \tag{7}$$

is used, where we try to estimate the noise-free image that best explains the noisy observation assuming some prior information. By x we describe a noise-free pixel of the image; y indicates a pixel of the noisy observation, e.g., the European Remote Sensing 1 (ERS-1) image; and by $\boldsymbol{\theta}$ we characterize the parameters of the applied model.

The Bayes' equation (7) allows the formulation of the information extraction problem as a maximum *a posteriori* (MAP) estimation

$$\hat{x}_{\mathrm{MAP}} = \arg\max_x p(x \,|\, y, \boldsymbol{\theta}) \tag{8}$$

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \,|\, y). \tag{9}$$

Analytically computed MAP estimates of the cross section are generated from the filter. Subsequently, they are employed to produce parameters for $p(x \,|\, y, \boldsymbol{\theta})$ by iterative maximization of the evidence [18]. Expectation maximization is used to estimate the nonstationary texture parameters that provide the highest evidence value. The estimated model parameters express the characteristics of the texture and the strength of geometrical structures in the data.

The model used as a prior is the Gauss–Markov random field (GMRF) texture model [8], [18]

$$p(x_s \,|\, \partial x_s, \sigma, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_s - \sum_{ij} \theta_{ij}(x_{ij} + x'_{ij}))^2}{2\sigma^2} \right) \tag{10}$$

specified by σ and the parameter vector $\boldsymbol{\theta} = (\theta_{11}, \theta_{12}, \theta_{21}, \ldots)$. The latter is defined on a neighborhood of cliques centered on the generic pixel x_s so that the scalar parameters are symmetric around the central element. The main strength of the Gauss-Markov model lies in its ability to model structures in a wide set of images while still allowing analytical tractability. The likelihood used in the Bayes equation (7) is the gamma distribution

$$p(y \,|\, x) = 2 \left(\frac{y}{x}\right)^{2K-1} \frac{K^K}{x\Gamma(K)} \exp\left(-K \left(\frac{y}{x}\right)^2 \right) \tag{11}$$

with K the number of looks of the data. From the estimated parameters, we take the model-based filtered intensity image and the norm of the model parameter $|\hat{\boldsymbol{\theta}}|$ as exemplified in (Fig. 3).

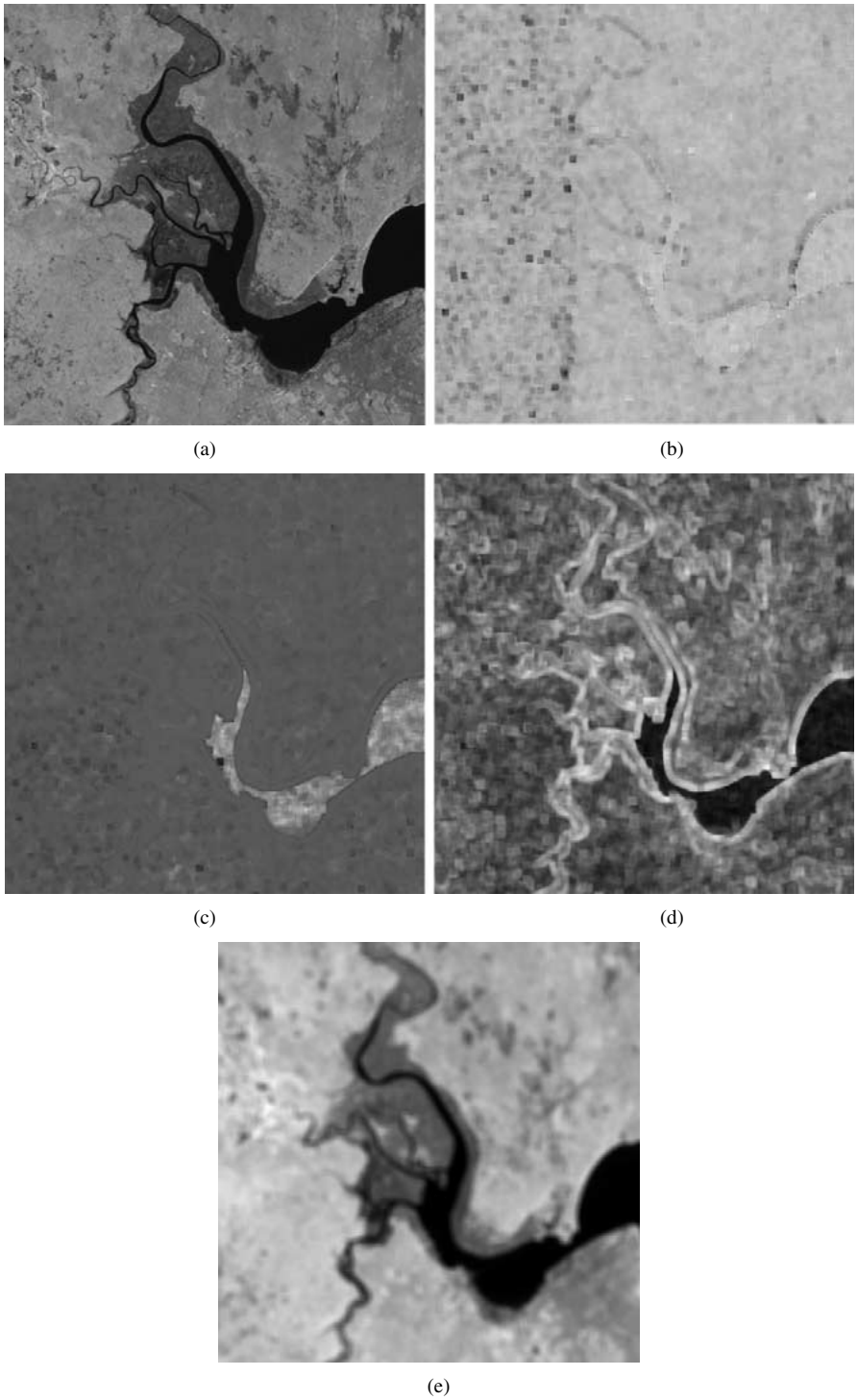
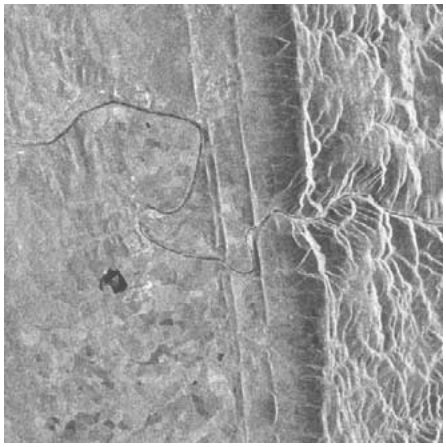


Fig. 2. Extracted textural features from a Landsat TM image. (a) Fourth band of Landsat as data D and (b) norm $|\hat{\boldsymbol{\theta}}|$ of the estimated texture parameters. (c) Variance $\hat{\sigma}^2_M$ and (d) evidence $p(M \mid D)$. (e) Local mean of the estimation kernel.

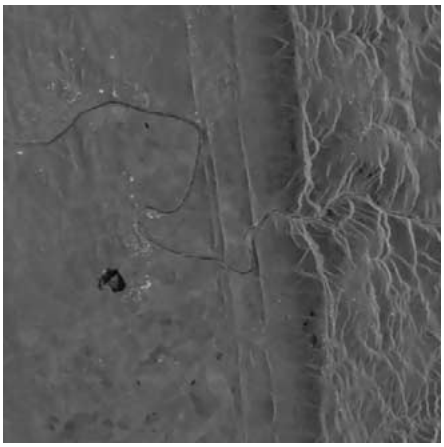
C. Information Extraction at Multiple Scales

We showed that parametric data models are suitable to characterize spatial information in images by its parameter vector $\boldsymbol{\theta}$. Capturing highly complex textures that have features at different scales, particularly large-scale structures such as mountains or rivers, requires high-order models. With an increasing neighborhood size, the number of parameters grows and leads to an averaging effect of different parameters. This results in a limited discrimination power of the extracted texture features and impairs the interpretation.

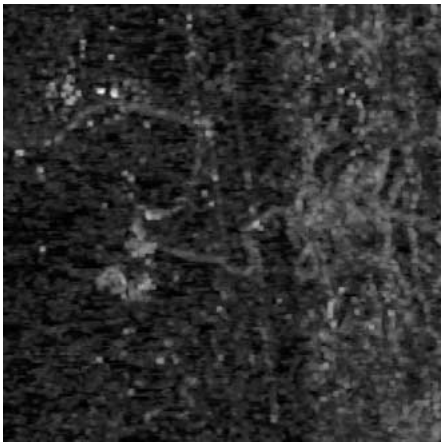
The approach we follow for a complete description of all texture structures is to generate a multiresolution image pyramid where the original image is located at the lowest layer and the reduced resolution representations of the image at higher layers [14]. If the same Gibbs random field texture model is applied with a limited neighborhood size at different layers, information for different structures is extracted. Thus, we can charac-



(a)



(b)



(c)

Fig. 3. Information extraction from radar images. (a) and (b) Original ERS-1 intensity image covering Mozambique with (a) mountains, rivers, and flat terrain and (b) the model-based despeckled image. (c) Norm $|\hat{\boldsymbol{\theta}}|$ of the estimated texture parameter vector.

terize large-extended spatial information by a restricted model order.

IV. Unsupervised Clustering of Primitive Image Features and Catalogue Entry Generation

In the previous section, we pointed out how the content of optical and radar images can be described by parametric data models. Since the feature extraction produces large volumes of data that cannot be managed in practice, estimated image parameters must be compressed and reduced. Clustering, which is similar to a quantization process, reduces the accuracy of the system, but justifies its practical use due to a large data reduction. In order to reject existing structures in the different feature spaces of the data and to avoid the time-consuming calculation of similarity functions, the unsupervised clustering is performed across all images as outlined in Fig. 1. We perform the global unsupervised clustering using a dyadic k means algorithm [5], which substitutes the "clouds" of primitive features by parametric data models $p(\boldsymbol{\theta} \mid \omega_i)$. Although our clustering method is slightly less accurate than k means, it significantly reduces the processing time. Especially for a large number of clusters, the algorithm has proved to be very efficient.

From the results of unsupervised feature classification, we derive a set of signal classes that describe characteristic groups of points in the parametric spaces of different models. This "vocabulary" of signal classes is valid across all images, ensured by the global across-image classification. The elements ω_i of this "vocabulary" are given by the cluster membership of all image points to one of the clusters. For each image, this results in as many classification maps as the number of models that are used. From these maps, we calculate the probabilities $p(\omega_i \mid I)$ of the ith class given a certain image I. These probabilities are separately computed for each signal model. We obtain the probabilities by calculating the histogram of the occurrence of signal classes ω_i in an image I. The elements of the histograms, the probabilities $p(\omega_i \mid I)$, are stored in a relational database system together with the classification maps. The latter are stored as binary large objects (BLOBs). Additionally, Quick-Looks (QLs) and their thumbnails as BLOBs in JPEG format, metainformation, such as sensor type, time of acquisition, geographical information, etc., are inserted.

V. User-Specific Semantic Labeling

The first three levels of our hierarchical modeling describe the image data D at level 0 in a completely unsupervised way (Section 2). Based on this objective representation, we can now link subjective user interests L (level 4) to the signal classes ω_i by probabilities $p(\omega_i \mid L)$. For a robust characterization of user-specific semantics L, several signal models (level 3) have to be applied.

Then, we link the elements of the joint space of signal classes to the user's interests. The stochastic link can be achieved with different models for $p(\omega_i \mid L)$, but only if we suppose a full statistic independence written as

$$p(\omega_{jk\ldots} \mid L) = p(\omega_j \mid L) \cdot p(\omega_k \mid L) \cdot \ldots \tag{12}$$

a fast computation is possible. In the following, we will restrict ourselves to a statistic independence for $p(\omega_{jk\ldots} \mid L)$ with two models j and k.

With the results of unsupervised classification (level 3), we obtain the posterior probabilities $p(\omega_{jk} \mid D)$ for the signal classes ω_{jk} given the data D. With these results and the assumption that the signal characteristics of the semantic label

L are fully representated by ω_{jk}, we can calculate the posterior probability as

$$p(L\,|\,D) = \sum_{jk} p(L\,|\,\omega_{jk}) p(\omega_{jk}\,|\,D). \tag{13}$$

With Bayes' formula, (13) can further be expressed as

$$p(L\,|\,D) = p(L) \sum_{jk} \frac{p(\omega_{jk}\,|\,L)\, p(\omega_{jk}\,|\,D)}{p(\omega_{jk})} \tag{14}$$

where $p(L)$ indicates the prior probability of semantic labels L and $p(\omega_{jk}) = \sum_L p(\omega_{jk}\,|\,L) p(L)$ the prior of signal classes ω_{jk}. Since the posterior probability can be calculated for each image pixel, we can visualize $p(L\,|\,D)$. The spatial visualization of $p(L\,|\,D)$ is named in the following as "posterior map." This map gives the system operator feedback of how strong and accurate the cover type label has been already defined.

VI. Interactive Learning

In order to make the inference from the image data D (level 0) to the cover type label L (level 4), the system first has to learn the probabilistic link $p(\omega_i\,|\,L)$ based on user-supplied training samples. As mentioned in the last section [see (12)], we assume conditional independence for the signal classes ω_{jk} as a combination of two features. In the following, we denote the classes by ω_i. We perform the probabilistic learning with a simple Bayesian network [10]. Assume we have a set of user-supplied training data T expressed by $\{N_1, \ldots, N_r\}$ with N_i being the occurrence of ω_i in T. Then, the vector of N_i has a multinomial distribution, since ω_i is a variable with r states [2], if we consider the parameter vector $\boldsymbol{\phi} = \{\phi_1, \ldots, \phi_r\}$ as a model for the set of probabilities

$$p(\omega_i\,|\,L, \boldsymbol{\phi}) = \phi_i. \tag{15}$$

Now, we change our discussion from determining the probabilities of the signal classes ω_i to the parameter vector $\boldsymbol{\phi}$. For a newly defined label, we start with a constant initial prior distribution

$$p(\boldsymbol{\phi}) = \Gamma(r) = (r-1)! \tag{16}$$

where r indicates the number of signal classes ω_i and $\Gamma(\,\cdot\,)$ the gamma function. With our observed training set T and its instances N_i, we obtain the posterior probability

$$\begin{aligned} p(\boldsymbol{\phi}\,|\,T) &= \frac{\Gamma(r+N)}{\prod_i \Gamma(1+N_i)} \prod_i \phi_i^{N_i} \\ &= \mathrm{Dir}(\boldsymbol{\phi}\,|\,1+N_1, \ldots, 1+N_r) \\ &= \mathrm{Dir}(\boldsymbol{\phi}\,|\,\boldsymbol{\alpha}) \end{aligned} \tag{17}$$

with the total sum of training samples $N = \sum_i N_i$, the Dirichlet function $\mathrm{Dir}(\boldsymbol{\phi}\,|\,\boldsymbol{\alpha})$, and the hyperparameters

$$\alpha_i = 1 + N_i. \tag{18}$$

If we observe another training set T' that is considered to be independent on T, we obtain by

$$\begin{aligned} p(\boldsymbol{\phi}\,|\,T', T) &= \frac{p(T'\,|\,\boldsymbol{\phi}, T)\, p(\boldsymbol{\phi}\,|\,T)}{p(T', T)} \\ &= \mathrm{Dir}(\boldsymbol{\phi}\,|\,\alpha_1 + N_1', \ldots, \alpha_r + N_r') \end{aligned} \tag{19}$$

an additional update of the hyperparameters by adding the number of times ω_i' occurs in the training dataset T'

$$\alpha_i' = \alpha_i + N_i'. \tag{20}$$

The initial state of the hyperparameters is given by

$$\boldsymbol{\alpha}_0 = \{1, \ldots, 1\} \tag{21}$$

and a new set of training samples updates the hyperparameters as given in (20).

Having a definition of the hyperparameters $\boldsymbol{\alpha}$ by some training sets T, we can finally calculate the probabilities as expectation over all possible values of $\boldsymbol{\phi}$ as

$$p(\omega_i\,|\,L, T) = \frac{\alpha_i}{\sum_i \alpha_i}. \tag{22}$$

The fast computation of the probabilistic link $p(\omega_i\,|\,L)$ [see (22)] and the updating after observing new training data [see (20)] make the hyperparameters $\boldsymbol{\alpha}$ a very advantageous tool to describe the stochastic link between objective signal classes and subjective user semantics (Fig. 1).

In order to allow high-precision training specified on full-resolution images, an online training interface has been developed (Fig. 4).

A human trainer can define an arbitrary number of (pairwise disjunct) cover types (e.g., "lake" and "not lake") on a set of images in full resolution. After selecting a combination of signal classes of feature models, the trainer can ask for the posterior map of a particular cover type label or an assessment of the selected features classes. Since the image content has already been extracted up to level 3, only the probabilistic link has to be recalculated, and the response is pretty fast. This allows an iterative refinement of the training regions and "simultaneous" observation of the consequences for the posterior probabilities.

VII. Probabilistic Search

We can calculate the posterior probability of L given the image I as

$$p(L\,|\,I) = \sum_i p(L\,|\,\omega_i)\, p(\omega_i\,|\,I) \tag{23}$$

in a similar way as we calculated the posterior probability of L given a particular data D [see (14)]. The posterior probability is a measure of how probable an image "is of" a particular cover type.

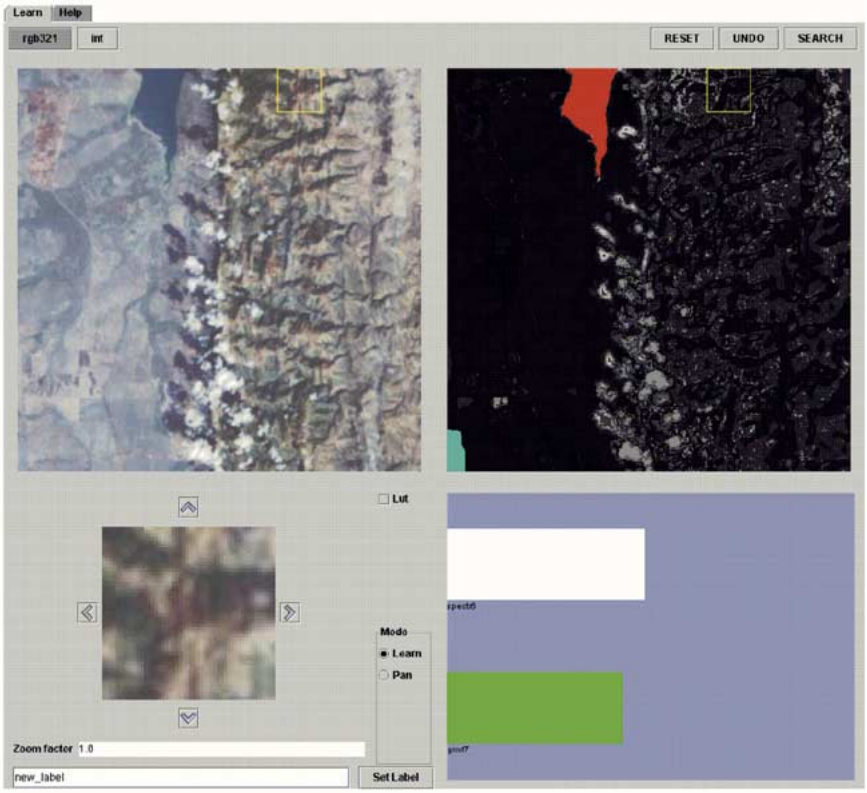


Fig. 4. Graphical user interface in KIM for interactively learning the contents (water) of remote sensing images. The system operator can specify his interests by giving positive and negative training samples, either directly (top left) into the original image, (bottom left) the zoom window, or (top right) the posterior map. After each mouse click, the hyperparameters $\boldsymbol{\alpha}$ and the probabilities $p(\omega_i \mid L)$ of the stochastic link are updated, and the posterior map on the right side is recomputed and redisplayed. As a further quality measurement, we display (bottom right) the divergence between positive and negative training samples. If the users are satisfied with the posterior map, they can search the entire archive for relevant images by clicking on the "SEARCH" button.

To provide a more practical measure for image retrieval, we compute the "coverage"

$$C = \sum_i p(\omega_i \mid I) \cdot \text{Heaviside}\,(p(L \mid \omega_i) - p_{\text{th}}) \tag{24}$$

which specifies the approximate percentage of the image that definitely contains the desired cover type. The degree of "definitely" is determined via the threshold p_{th}.

Since the distribution of $p(\omega_i \mid L)$—resulting from limited training data—is known in detail, we can specify both the probability of a label in a particular image and the expected degree of variation. We do this by calculating the expected variance of the posterior

$$\delta^2 p(L \mid I) = \sum_i \delta^2 p(L \mid \omega_i) p(\omega_i \mid I) \tag{25}$$

with the δ^2 symbol denoting the variance. As a measure of how well L is separated from $\neg L$ in a particular image I, we use the separability

$$S(L \mid I) = \frac{\delta^2 p(L \mid I)}{p(L \mid I)(1 - p(L \mid I))} \tag{26}$$

which is the variance in units of the maximal possible variance. The smaller $S(\cdot)$, the "better" we call the separability. The separability measurement is very useful for further learning, since retrieved images with low separability are related to performed positive training samples and images with high separability are connected to negative training. The user can either decide to enforce the positive training because of bad query results for low separability or enforce negative training due to bad search results for high separability.

VIII. System Description and Configuration

In this section, we describe the KIM system from the technical point of view with its main compontents (Fig. 5).

To access KIM, a user has to register first by choosing a user id and a password. After successful login, a personal welcome page is displayed. The user can decide to perform some administration to start the interactive learning process. The latter requires the selection of a combination of up to four signal models. In KIM, the information extracted from one single sensor as well as the information from multiple sensors can be used for interactive training and probabilistic search. Having selected a certain model combination, the user has to pick out a starting image from a gallery of randomly chosen images. If the gallery does not contain an image of the user's favor, they can choose another set of random images.

Once clicked on an image, the interactive learning process begins as depicted in Fig. 6.

In a first step, the following objects are downloaded by the interactive learning applet: the QL image in JPEG format and the corresponding classification maps (image content catalogue)

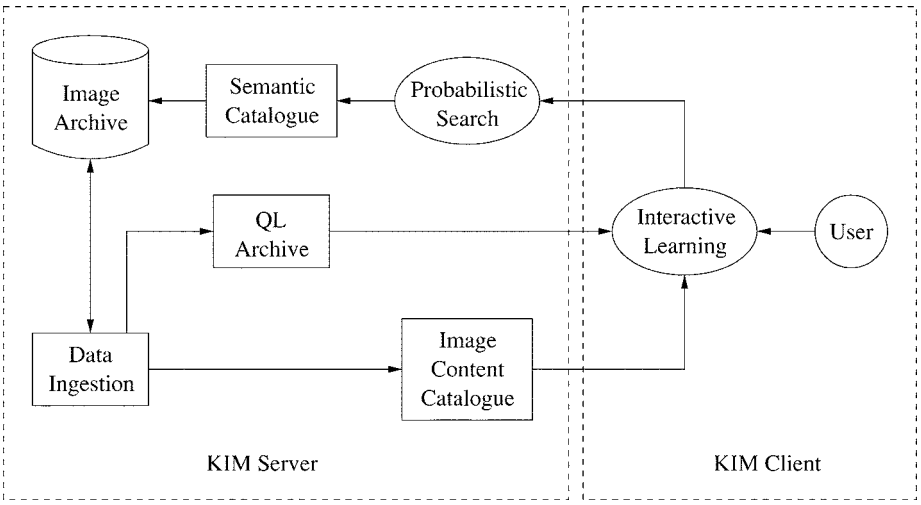


Fig. 5. KIM client–server system architecture.

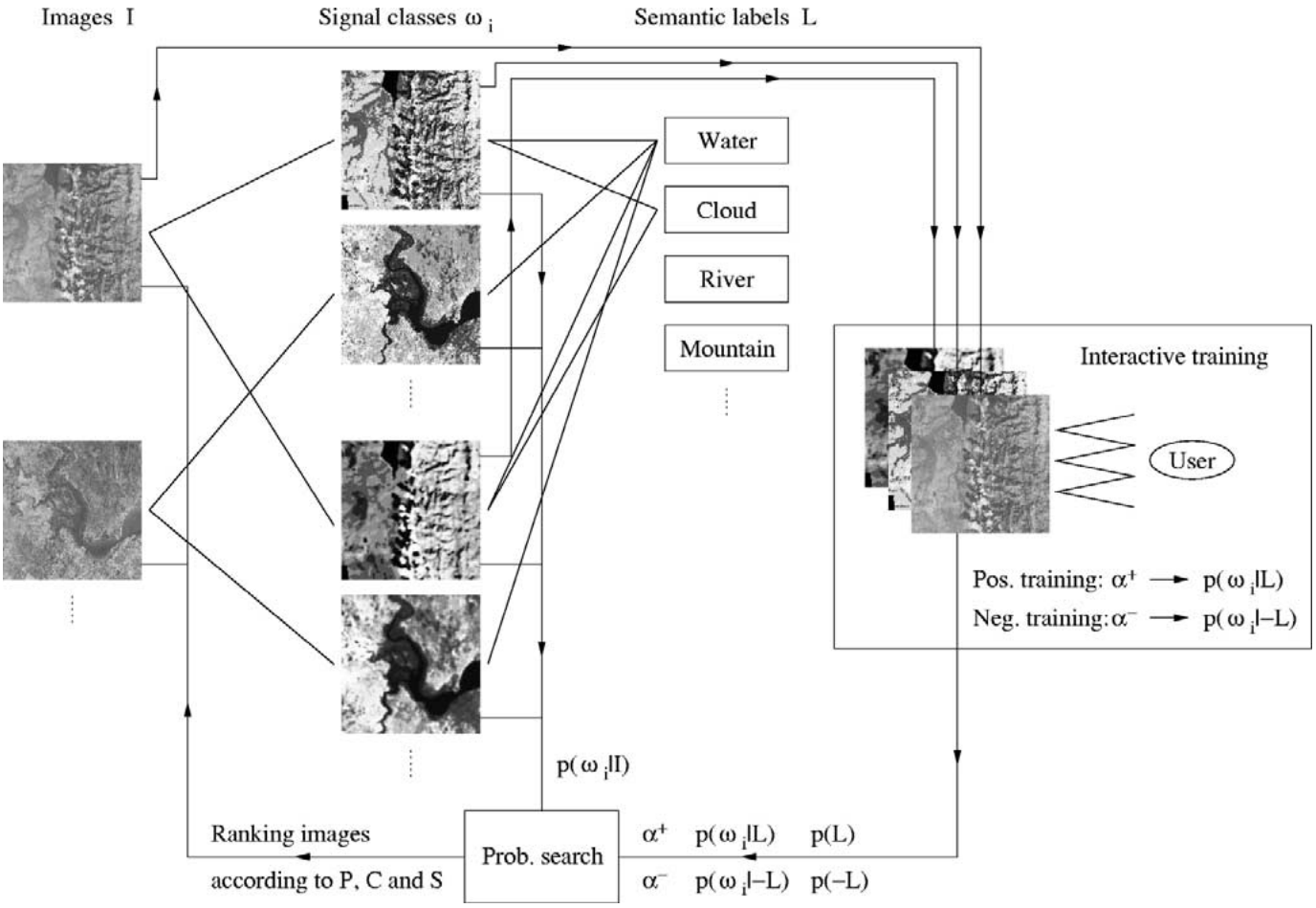


Fig. 6. Dataflow during interactive learning and probabilistic search. After identifying the user and choosing a combination of signal models, the learning applet is downloaded from the server to the client browser. The system operator can continously train a specific label of interest by giving positive and negative training samples. After each training (mouse click), the hyperparameters $\boldsymbol{\alpha}$ are updated, and the redisplayed posterior map indicates the current state of the label. If the users are satisfied with the trained label, they can query the entire image archive for the defined cover type. The system delivers the top ranked images according to (C) coverage, (P) posterior probability, and (S) separability. For further label definition, the user can select another image from the search results.

for the selected signal models in raw binary format. When the downloading is finished (after a few seconds), the user can start the definition of a semantic cover type of interest by giving positive and negative samples using the left/right mouse button. After each click, the hyperparameters $\boldsymbol{\alpha}$, the likelihoods $p(\omega_i \mid L)$, and the posterior map are updated. The latter permanently gives the user an intuitive feedback about the quality of the learning process by marking regions corresponding to the cover type with red color. If the current label definition is satisfactory, the system operator can query the entire archive

TABLE I
INGESTED DATASETS IN KIM AND THE APPLIED SIGNAL MODELS

Sensor	Landsat TM	ERS1	Ikonos	Landsat TM	Landsat TM
Coverage	Mozambique	Mozambique	Mozambique	Nepal	Switzerland
No. of scenes	14	32	9	1	4
No. of images	438	438	207	144	184
Channels, size	$6, 2000^2$	$1, 2000^2$	$1, 2000^2$	$8, 500^2$	$7, 1024^2$
Resolution	25m, geo./co.	25m, geo./co.	1m, geo.	25m, geo.	25m, geo.
Signal models	spectral, GMRF at 3 scales	intensity, EMBD at 2 scales	intensity, GMRF at 3 scales	spectral, GMRF at 2 scales	spectral, GMRF 5 scales

Sensor	Ikonos	Ikonos	Daedalus ATM	E-SAR
Coverage	Germany	Germany	Kosovo	Kosovo
No. of scenes	1	1	1	1
No. of images	43	43	12	12
Channels, size	$1, 500^2$	$4, 500^2$	$12, 1000^2$	$4, 1000^2$
Resolution	1m, geo./co.	4m, geo./co.	1m, geo./co.	1m, geo./co.
Signal models	spectral, GMRF at 3 scales	spectral	spectral, GMRF at 3 scales	pol. intensity EMBD at 2 scales

for images containing similar structures or objects. For the computation of the probabilistic search measurements on the KIM server's site, only the hyperparameters with the derived posterior probabilities $p(L \mid \omega_i)$ and the probabilities $p(\omega_i \mid I)$ of the generated and inserted catalogue entries are necessary. At this time, the label is persistently stored in the database. The definition of the label is given by its name, the used image from training, the selected signal models, the hyperparameters, and the resulting (queried) images.

The user can continue the learning process until they are satisfied with the query result. In order to improve the definition of the semantic label, the system operator clicks on another image in the resulting image set and continues to feed in positive and negative examples. Every time the user selects an image from the query gallery, the QL and the assigned signal model classification maps are transmitted via the World Wide Web. We want to point out that the cover type learning using several images is important to obtain a well-defined semantic label. We call this "iterative incremental learning." Each time the user queries the image archive, the semantic label definition in the database is updated.

A tool worth to be mentioned in KIM is the tracking module that stores each human–machine interaction in the database. Based on the stored information, the system computes a number of statistical and information theoretical measures that indicate the goodness of the learning process. These measures give the user a further feedback about the learning progress (lower left part in Fig. 4).

IX. PRACTICAL APPLICATIONS

The applied concept of unsupervised indexing of image content and the user-specific semantic labeling of cover types have been extensively tested in KIM based on various remote sensing datasets (Table I).

In the performed experiments, the image data range from monochromatic high-resolution (Ikonos) to hyperspectral (Daedalus ATM) data and from medium-resolution SAR (ERS-1) to high-resolution polarimetric (E-SAR) image data. The fusion of different signal models from one sensor, as well as the fusion of multisensor image data, is applied for interactive learning and probabilistic retrieval. With it, we want to demonstrate the power of KIM for data-independent image mining applications.

In the following, we show examples of labeling user-defined semantics and query results from the image archive. We start with the analysis of a cover type "mountain" that was trained with different combinations of signal models as shown in Fig. 7.

The selected combination of signal models influences both the level of compactness and detail of the semantic label. The retrieved images for the defined cover type "mountain" are given in Fig. 8. By default, only the highest six top ranked images are delived for probability, coverage, and separability, but the user can ask for more results.

User-specific interactive learning with information from multiple sensors can be used for sensor qualification and further exploration of the image dataset. For this, the interactive training with high-resolution image data is exemplified in Fig. 9.

In a final application, we show the classification and retrieval of the label "water" from coregistered high-resolution hyperspectral and polarimetric radar data (Fig. 10).

The applied signal models are spectral from the hyperspectral data and from E-SAR the despeckled SAR backscatter (L-band, scale 2 m), the despeckled SAR backscatter (L-band, scale 2 m), and the norm of the SAR texture vector (L-band, scale 4 m). With an increasing number of signal models, the number of structural details grows.

X. CONCLUSION

In this paper, we have presented KIM, which is a prototype of a next-generation knowledge-driven image information mining system developed for the exploration of large image archives. We started our presentation with the offline part of KIM that

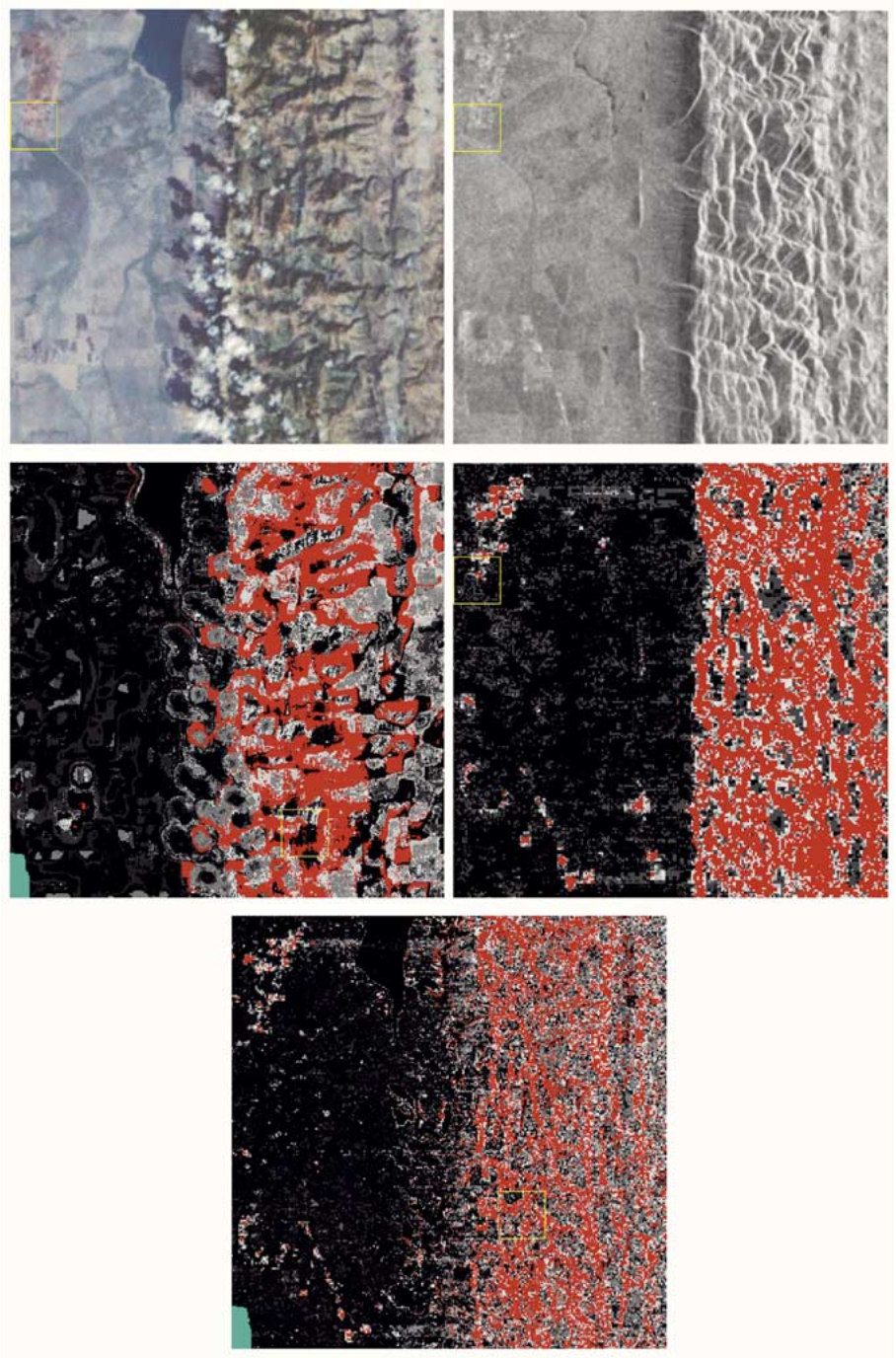


Fig. 7. Interactive training of mountainous areas with different combinations of signal models. Regions belonging to the semantic label are marked with red color. QL from (top left) Landsat TM and (top right) ERS-1 image data. (Middle left) Trained semantic label with spectral and texture information from Landsat TM and (middle right) the obtained results with only texture features from ERS-1 at different scales. (Bottom) Cover type with across-sensor model combination of spectral (Landsat) and texture (ERS-1).

consists of the extraction of primitive features from image data and their compression by an unsupervised classification. After generating the catalogue entries from the clustering results and ingesting them in the database, we described the online part of KIM: user-specific semantic image content labeling. The semantic cover type is defined based on an intelligent graphical user interface. If the users are satisfied with the trained semantic label, they can query the whole archive for images that contain the defined content. We demonstrated the system operation of KIM and its potential for practical applications on various datasets. Thereby, we included extracted information from multiresolution datasets as well as the information from multisensor data.

In the future, we will further develop KIM and use the knowledge and semantic information that is stored in the database system. During interactive learning and probabilistic search, the database management system records the user semantics, the combination of models able to explain the user's target, the classification of the target structure in each individual image, and a set of statistical and information theoretical measures of

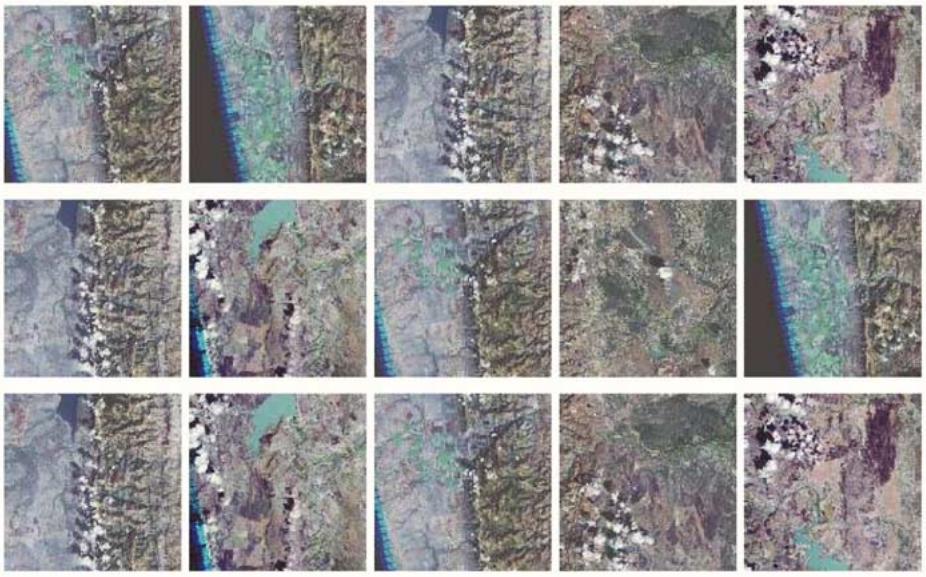


Fig. 8. Results of probabilistic search for the trained semantic label "mountain" in upper right panel of Fig. 7. The queried images are ranked according to (first row) coverage [see (24)], (second row) posterior probability [see (23)], and (third row)separability [see (26)]. The user can continue training the cover type by selecting one of the retrieved images.

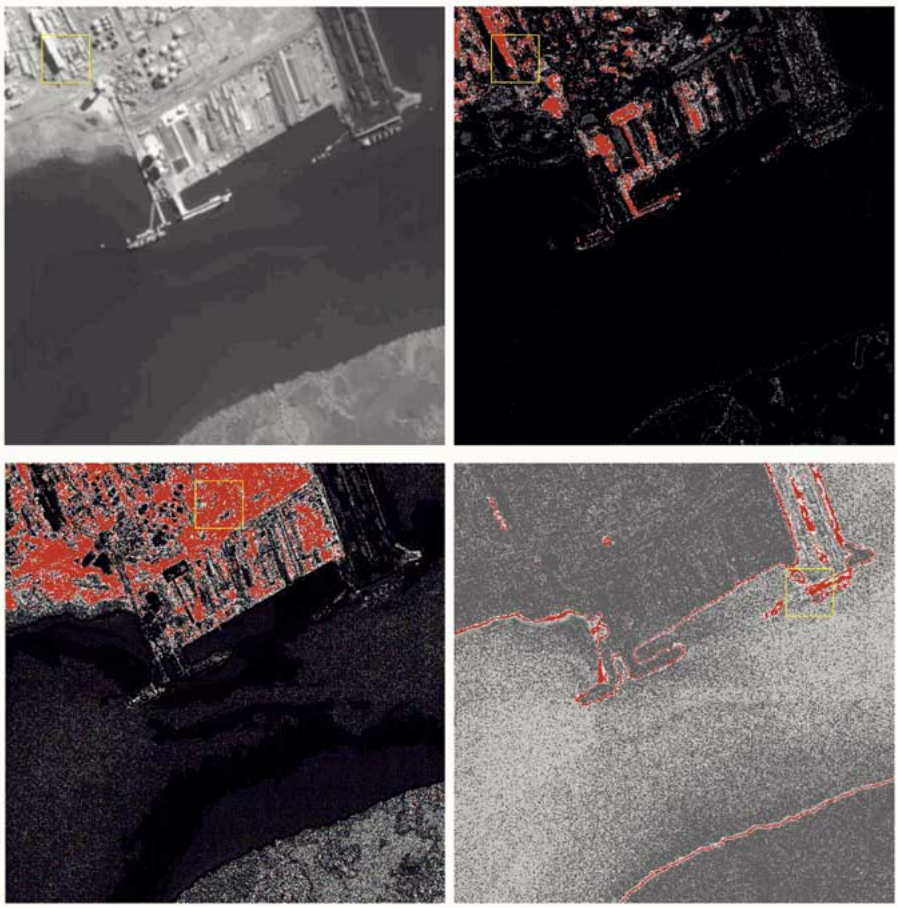


Fig. 9. Interactive training of semantic labels in an Ikonos image with spectral and texture information. (Upper left) QL of a panchromatic Ikonos image with a resolution of 1 m and (upper right) trained semantic label "industrial area." (Bottom left) Trained cover type labels "grassland" and (bottom right) "coastline." Consider that the defined semantic labels were obtained with just a few training samples.

the quality of the learning process. This information and associations represent a body of knowledge, either discovered or learned from the various system users. It will be further used for other mining tasks. The acquired knowledge will be the object of mining, e.g., grouping of semantic levels, relevance feedback, joint grouping between the semantic space, and the

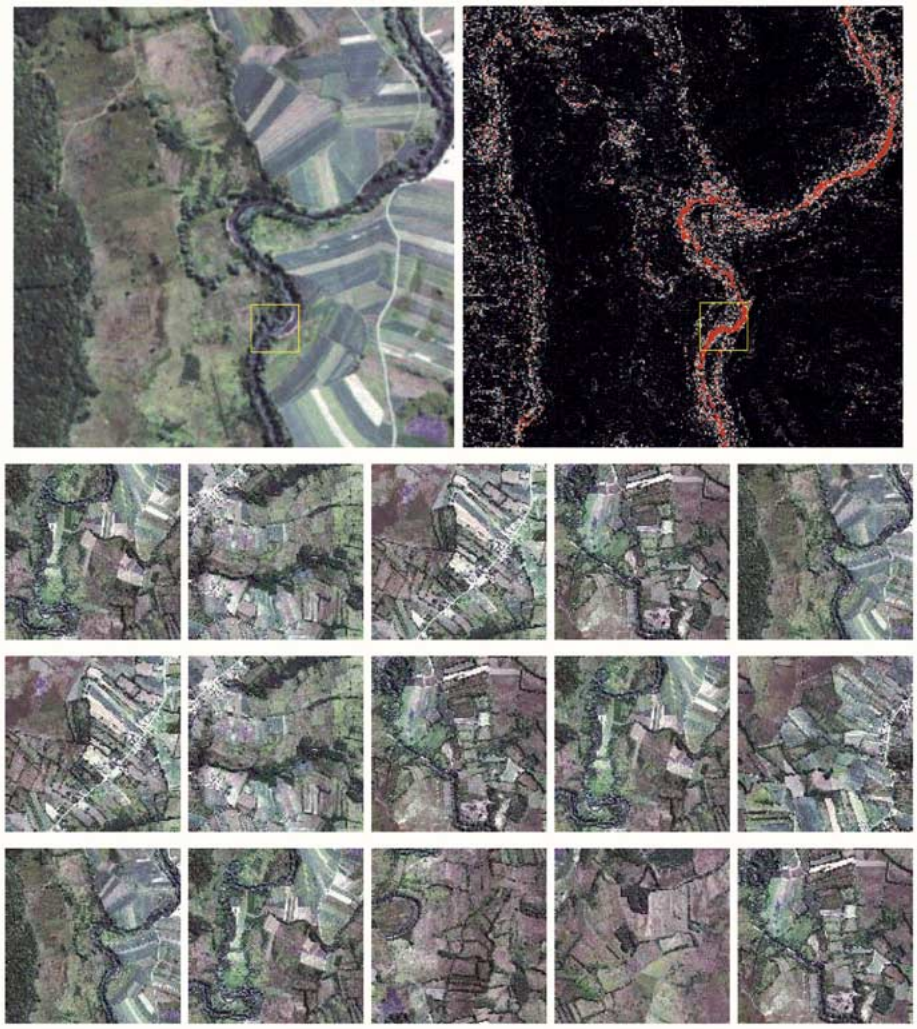


Fig. 10. Interactive learning and probabilistic search of semantic label "water" from hyperspectral and polarimetric data using four different signal models. (Upper left) QL and (upper right) posterior map. (Second row) Top ranked retrieved images according to coverage. (Third row) Posterior probability. (Bottom row) Separability.

statistical or information theoretical measures of the quality of the learning process. The KIM system is available online at http://www.acsys.it:8080/kim.

ACKNOWLEDGMENT

The authors want to thank the European Union Satellite Center EU SC (Madrid, Spain) and the Nansen Environmental and Remote Sensing Center NERSC (Bergen, Norway) for providing image datasets, additional relevant information, and participating with the evaluation of the KIM system.

REFERENCES

[1] J. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, and R. Jain, "The virage image search engine: An open framework for image management," in *Proc. SPIE Conf. on Storage and Retrieval for Image and Video Databases*, 1996, pp. 76–87.

[2] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York: Wiley, 2001.

[3] C. Chang, B. Moon, A. Acharya, C. Shock, A. Sussman, and J. Saltz, "Titan: A high-performance remote sensing database," in *Proc. 13th Int. Conf. Data Engineering*, 1997, pp. 375–384.

[4] I. J. Cox, M. L. Miller, S. M. Omohundro, and P. N. Yianilos, "Pichunter: Bayesian relevance feedback for image retrieval," in *Proc. Int. Conf. Pattern Recognition*, vol. 3, 1996, pp. 361–369.

[5] H. Daschiel and M. Datcu, "Cluster structure evaluation of a dyadic k-means algorithm for mining large image archives," in *Proc. SPIE Remote Sensing Symp.*, 2002.

[6] M. Datcu, A. Pelizzari, H. Daschiel, M. Quartulli, and K. Seidel, "Advanced value adding to metric resolution SAR data: Information mining," in *Proc. 4th Eur. Conf. Synthetic Aperture Radar (EUSAR 2002)*, 2002.

[7] M. Datcu and K. Seidel, "New concepts for remote sensing information dissemination: Query by image content and information mining," in *Proc. IGARSS*, vol. 3, 1999, pp. 1335–1337.

[8] M. Datcu, K. Seidel, and M. Walessa, "Spatial information retrieval from remote sensing images. Part I: Information theoretical perspective," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1431–1445, Sept. 1998.

[9] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkini, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: The QBIC system," *IEEE Comput.*, vol. 28, pp. 23–32, Sept. 1995.

[10] D. Heckermann, D. Geiger, and D. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data," Microsoft, Redmond, WA, Tech. Rep. MSR-TR-94-09, 1994.

[11] T. P. Minka and R. W. Picard, "Interactive learning using a "society of models"," *Pattern Recognit.*, vol. 30, no. 4, pp. 565–581, 1997.

[12] J. J. Ó. Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*. New York: Springer, 1996.

[13] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, pp. 644–655, Sept. 1998.
[14] M. Schröder, H. Rehrauer, K. Seidel, and M. Datcu, "Spatial information retrieval from remote sensing images. Part II: Gibbs–Markov random fields," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 1446–1455, Sept. 1998.
[15] ——, "Interactive learning and probabilistic retrieval in remote sensing image archives," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 2288–2298, Sept. 2000.
[16] J. R. Smith and S.-F. Chang, "VisualSEEk: A fully automated content-based image query system," *ACM Multimedia*, pp. 87–98, 1996.
[17] C. R. Veltkamp, H. Burkhardt, and H.-P. Kriegel, *State-of-the-Art in Content-Based Image and Video Retrieval*. Norwell, MA: Kluwer, 2001.
[18] M. Walessa and M. Datcu, "Model-based despeckling and information extraction from SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 2258–2269, Sept. 2000.

Mihai Datcu received the M.S. and Ph.D. degrees in electronics and telecommunications from the University "Politehnica" of Bucharest UPB, Bucharest, Romania, in 1978 and 1986, and the title "Habilitation a diriger des recherches" from Université Louis Pasteur, Strasbourg, France.

He holds a Professorship in electronics and telecommunications with UPB since 1981. Since 1993, he has been a Scientist with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. He is currently developing algorithms for model-based information retrieval from high-complexity signals, methods for scene understanding from SAR and interferometric SAR data, and he is engaged in research in information theoretical aspects and semantic representations in advanced communication systems. He has held Visiting Professor appointments from 1991 to 1992 with the Department of Mathematics, University of Oviedo, Oviedo, Spain, from 2000 to 2002 with the Université Louis Pasteur, and the International Space University, Strasbourg, France. In 1994, he was a Guest Scientist with the Swiss Center for Scientific Computing (CSCS). Manno, Switzerland, and in 2003, he was a Visiting Professor with the University of Siegen, Siegen, Germany. From 1992 to 2002, he had longer Visiting Professor assignments with the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland. He is involved in advanced research programs for information extraction, data mining and knowledge discovery, and data understanding with the European Space Agency, Centre National d'Etudes Spatiales, the National Aeronautics and Space Administration, and in a variety of European projects. He is currently a Senior Scientist and Image Analysis research group leader with the Remote Sensing Technology Institute (IMF), DLR. His research interests are in Bayesian inference, information and complexity theory, stochastic processes, model-based scene understanding, image information mining, for applications in information retrieval and understanding of high-resolution SAR and optical observations.

Herbert Daschiel received the diploma degree in geodetic engineering from the Technical University Munich, Munich, Germany, in 2001. He is currently pursuing the Ph.D. degree.

Since spring 2001, he has been with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany. His thesis is on image information mining system upgrade and evaluation. His research interests are in image processing, statistical data analysis, and information theory.

Andrea Pelizzari received the laurea degree in electronics and telecommunications from the Politecnico di Milano, Milano, Italy, in 1985.

After working in the software industry, he worked for the German Aerospace Center (DLR), Oberpfaffenhofen, Germany, from 1999 to 2002, where he participated in the SRTM Shuttle mission as a Signal Processing Expert and was a member of the X-Data capture team at the Jet Propulsion Laboratory, Pasadena, CA. He has recently been involved in R&D projects in the field of data mining in large image databases and web applications for image archive browsing. He is currently a Consultant in the software and web development industry.

Marco Quartulli received the laurea degree in physics from the University of Bari, Bari, Italy, in 1997 with a thesis on differential SAR interferometry. He is currently purusing the Ph.D. degree.

From 1997 to November 2000, he worked on SAR and InSAR processing systems at Advanced Computer Systems, Rome, Italy. In 2000, he joined the Image Analysis Group at the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germnay. His research interests include Bayesian modeling and scene understanding for high-resolution SAR and InSAR.

Annalisa Galoppo received the laurea degree and the Ph.D. degree in mathematics from the University of Naples "Federico II," Naples, Italy, in 1995 and 1999.

From 1996 to 1997, she was a Visiting Scholar at University of Kentucky, Lexington. In 2000, she attended a course in multimedia organized by the University of Florence, Florence, Italy, and RAI (Radio Televisione Ialiana). From 2000 to 2002, she was a Virtual Reality Software Engineer with Infobyte, Rome, Italy. She is currently a Software Engineer with the Earth Observation Division, Advanced Computer Systems, Rome, Italy. She has authored several papers on infinite group theory.

Andrea Colapicchioni was born in Rome, Italy, in 1969. He finished his high school with the specialization in informatics in 1988.

He started as a Freelance Programmer, developing an internet portal and several applications in Windows environments. He has developed and tested computer-aided design software for roads and railroads on Hewlett–Packard UNIX platforms. From 1992–1997, he was a Information Specialist and Lecturer at the Italian Electric Company (ENEL), Rome, Italy, where he was involved in the user requirement analysis, maintenance, testing, and user training for the internal information system. He is currently a Project Manager with the Advanced Computer Systems, Rome, Italy, where he collaborates on many international projects and is involved in database design, object-oriented architecture design, and development of the HyBro family products, a catalogue for georeferenced objects, based on the Informix database management system, C++ programming language, and QT portable widget library, installed at th most of the European Space Agency satellite ground stations. He has also participated as a Technical Coordinator and as a Support Developer for the KIM project. In 2002, he was a Speaker at the joint EUSC–ESA Seminar on Knowledge Driven Information Management in Earth Observation data.

Marco Pastori was born in Rome, Italy, in 1965. He finished his high school with the specialization in electronics and telecommunication in 1984.

He received a Fellowship for Olivetti SpA, Rome, Italy, as a Young Programmer in 1984. He was a Senior Programmer on UNIX for a Hewlett Packard Software House and developed a computer-aided roads and railways design application, the first commercial product in the world using clothoid and hyperclothoid curves. He has worked in the remote sensing application field since 1991, first as a Development Engineer in a military application for quality control, then as a Development Manager of a signal exploitation application for the Helios project. Since 1994, he his been with the Space Division, Advanced Computer Systems (ACS), Rome, Italy, and is currently a Senior Development Manager, where he collaborates with many international projects as a Consultant or Head of Development. At ACS, he has developed many specialized image processing products and algorithms. His recent work with different involvment levels include the following projects: CryoSAT, LIVIA, KES, KIM, Persia, HyBRO, Ramses, UMARF, Elettra, RAI-Trascrizioni. His research interests are in development processes, applied development models and techniques for distributed computing, distributed systems, and object-oriented programming.

Klaus Seidel received the B.S. degree in experimental physics in 1965 and the Ph.D. degree in 1971, both from the Swiss Federal Institute of Technology ETH Zurich.

He was with the Computer Vision Laboratory, ETH Zurich, and was Head of the remote sensing group until 2002. Since 1987, he has been a Swiss Delegate and Expert in various European Space Agency (ESA) Working Groups and at the same time functioning as the National Point of Contact for Switzerland. He is currently a Consultant for ESA projects that specialize in image information mining related to remote sensing archives. He also teaches courses in digital processing of satellite images and has published several papers concerning snow cover monitoring, geometric correction, and multispectral analysis of satellite images, as well as on remote sensing image archivation.

Pier Giorgio Marchetti received the laurea degree in electronic engineering from the University of Rome "La Sapienza," Rome, Italy, in 1977.

He is currently working in the Research and Technology Development Section, Earth Observation Programme, European Space Agency (ESA), Frascati, Italy, where he runs projects on information mining and feature extraction from earth observation products and archives, service-oriented infrastructures, and the grid for the earth observation ground segments. From 1986 to 1998, he was with the ESA's Information Retrieval Service, participating with the definition of the ANSI Z39.50 standard and running technology research projects on hypermedia and intelligent interfaces. From 1979 to 1985, he was a Researcher in the Fondazione Ugo Bordoni, Rome, Italy, where he carried out research projects on microwave atmospheric propagation, digital signal processing, digital radio communication, and cochannel interference for radio relay and satellite links. He has participated in the cochannel interference experiment between the SIRO and the ESA's OTS satellite with small receiving stations in Italy and China. In 1978, he was part of the R&D team of the Radio Laboratory of Autovox Spa (subsidiary of Motorola Inc.), Rome, Italy. He has coauthored more than 30 referred conference and journal papers.

Sergio D'Elia started his career in November 1966 as an Electronic Hardware Designer and then became a Software Analyst and, finally, Head of technical and team management. Since 1999, he has managed the Research and Technology Development Section, Earth Observation Programme, European Space Agency (ESA), Frascati, Italy, where he contributes to the definition of the European RTD requirements for the earth observation ground segment and manages a team in charge of implementing the related plan (concerned areas include user interfaces, information systems, payload planning, information mining, feature extraction, archiving, processing, distribution, automation, service support, GRID, etc.). From 1975 (when he joined ESA) until 1986, he has served in the ESA's Information Retrieval Service, managing part of the software development of QUEST, the ESA's information retrieval system. From 1966 to 1975, he was employed in the research and development branches of various leading Italian research centers and industries (CNR, CSM, Contraves, and Selenia), with responsibilities focused on electronic design of subsystems in support of various research topics, radar applications, or scientific satellite exploitation.