Interactive Learning and Probabilistic Retrieval in Remote Sensing Image Archives

Michael Schröder, Hubert Rehrauer, Klaus Seidel, and Mihai Datcu

Abstract

We present a concept of interactive learning and probabilistic retrieval of user-specific cover-types in a content based remote sensing image archive. A cover-type is incrementally defined via user-provided positive and negative examples. From these examples we infer probabilities of the Bayesian network that links the user interests to a pre-extracted content index. Due to the stochastic nature of the cover-type definitions, the database system not only retrieves images according to the estimated coverage, but also according to the accuracy of that estimation given the current state of learning. For the latter, we introduce the concept of separability. We expand on the steps of Bayesian inference to compute the application-free content index using a family of data models, and on the description of the stochastic link using hyper-parameters. In particular, we focus on the interactive nature of our approach, which provides instantaneous feedback to the user in the form of an immediate update of the posterior map, and a very fast, approximate search in the archive. A java-based demonstrator using the presented concept of content-based access to a test archive of Landsat TM, X-SAR, and aerial images is available over the Internet [http://www.vision.ee.ethz.ch/∼rsia/ClickBayes].

Correct page range: 2288–2298. This work is supported by the ETH project "Advanced Query and Retrieval Techniques for Remote Sensing Image Databases".

M. Schröder, H. Rehrauer, and K. Seidel are with the Swiss Federal Institute of Technology ETH Zürich, Communication Technology Lab, CH-8092 Zürich, Switzerland (e-mail: schro@ieee.org). M. Datcu is with the German Aerospace Center DLR, German Remote Sensing Data Center DFD, Oberpfaffenhofen, D-82234 Weßling, Germany.

I. Introduction

How can one find useful images in today's large remote sensing image archives? How can one know whether a certain image is useful for a particular application? Intelligent remote sensing information systems attempt to provide answers to these problems by offering novel access and retrieval techniques. Some approaches provide highly specialized content information that was automatically extracted from the images: analysis of cloud types for meteorology [1], automatic forest inventories [2], sea ice analysis of SAR images [3] or volcanoes on SAR images of Venus [4]. All these approaches have in common that they have been designed with the respective application in mind. Hence it is very difficult to extend their scope of applications.

On the other hand, with the evolving discipline of data mining, a whole family of methods for automatic content and knowledge extraction from large amounts of data is being established. For a review see [5,6,7]. Here various methods are applied to find characteristic structures (that is, content) in large amounts of data without a later application in mind. Several attempts exist to introduce this concept to remote sensing: by linking existing data mining and information systems [8], by combining distributed computing with interactive analysis and collections of algorithms [9], or by using a Bayesian approach in a signal-oriented framework [10,11]. The work presented in this paper has been done within the Bayesian framework.

In this paper we focus on a very important step in providing content-based query techniques: the interaction with the user and the flexible incorporation of application-specific interests. We first apply a family of strong signal models to obtain image features (e.g., spectral models or texture models at various scales), then perform an unsupervised clustering to obtain characteristic signal classes, and finally present these to be utilized by a user in defining his specific interests.

With this hierarchical procedure consisting of a slow, unsupervised clustering step, and a fast, user-interactive learning stage, the setup resembles the FourEyes system [12]. However, whereas in the original idea a self-organizing map is used for modeling a weight-space, we use different Bayesian networks [13] to directly link user interests and signal classes. This provides the user of the image archive with a probabilistic classification of the content of each image. Such a soft classification, which already has been proposed using fuzzy analysis [14], provides very intuitive information on the image content to the remote sensing user.

An advantage of Bayesian techniques is that they are free of ad-hoc assumptions and that they can lead to robust processing of information even for very limited user examples. Furthermore, we can use the same Bayesian network to retrieve images in a probabilistic way. Such techniques for probabilistic retrieval are currently also being proposed for text and multimedia retrieval [15].

Since computationally intensive operations (application of signal models and unsupervised clustering) are performed at the insertion time of the images, and since the actual user-specific labeling requires only very simple computations, it is possible to perform the latter in an interactive fashion on the client computer of the user. Existing network technologies and the capabilities of the Java programming language [16] enabled us to set up an online demonstration [17] that can be accessed by any user on the Internet without special software requirements. Most of the examples shown in this paper can be easily repeated and modified for own applications.

The outline of the paper is as follows. In Sec. II, we give a short summary of the basic concepts of Bayesian inference. Sec. III describes the application-free hierarchical modeling of image content, which is then linked to user-specific semantic labels in Sec. IV. In Sec. V, we present the interactive learning of the corresponding hyper-parameters of the probabilistic link, followed by the probabilistic image retrieval in Sec. VI. After a section on practical applications, Sec. VII, we conclude with a short summary.

II. Principles of Bayesian Inference

In the Bayesian formalism, the concept of probability as the frequency of an event is extended to the concept of probability as the degree of certainty of a particular hypothesis. The basic expressions are statements about conditional probabilities, e.g., $p(H|D)$, which specify the belief in the hypothesis H under the assumption that D (e.g., the data) is known with absolute certainty.

The heart of Bayesian techniques lies in the celebrated inversion formula called "Bayes' Rule"

$$p(H|D) = \frac{p(D|H)\,p(H)}{p(D)}, \tag{1}$$

which relates the posterior probability $p(H|D)$ to the likelihood $p(D|H)$. Here, the prior probability $p(H)$ reflects the prior belief in the hypothesis H and the prior predictive distribution $p(D)$ acts as normalizing constant.

The power of Bayesian techniques in (image) data processing comes from the fact that strong stochastic models are available for the (image) data. These models can be applied as models for the likelihood $p(D|H)$ for direct information extraction (e.g., for texture estimation [18]) or as powerful priors $p(H)$ in stochastic image reconstruction [19].

These stochastic models can be low level data models (e.g., Gibbs-Markov random field models modeling spatial structures [20]), intermediate level models (the semantic level we focus on in this paper; e.g., all examples of 'city' exhibit a certain color and fine texture) or on a very high level of semantics (e.g., a 'skiing scene' exhibits 'snow', 'blue sky' and 'crowds of people' [21]).

Using these models a decision for or against the hypothesis H can be made by considering the *posterior odds*

$$\Lambda = \frac{p(H|D)}{p(\neg H|D)} = \frac{p(D|H)}{p(D|\neg H)}\frac{p(H)}{p(\neg H)}. \tag{2}$$

If H is the hypothesis of D being a realization of what a user thinks is 'city', then the posterior odds allow the user to make a decision of 'city' versus another label, e.g, 'forest', 'lake', or simply 'not city' based on the stochastic models of 'city' and 'not city' and his prior belief.

The final goal of modeling image content is to calculate posterior odds, Eq. (2), for certain cover-types H (e.g., 'city'). To do so we have to develop a data model $p(D|H)$ of the data D for given hypothesis H, which is not trivial in the general case. To justify our approach of hierarchical modeling, we compare it to how humans would define 'city' in an image: maybe as "the cover-type city exhibits fine, regular texture and a certain color; furthermore, it exhibits a characteristic shape". To make the computer able to draw similar conclusions we need two prerequisites: (1) a strong set of vocabulary of characteristic signal properties allowing us to model a similar hierarchic representation of image content and (2) a formalism to pool the evidence of different information sources coherently (e.g., texture and spectral properties) and draw correct conclusions.

In the following we show how both prerequisites can be fulfilled using Bayesian inference: (1) by a hierarchical scheme following the different levels of semantic abstraction, in Sec. III, and (2) by using Bayesian belief networks linking objective signal properties and subject user interest in Sec. IV.

III. Application-free Hierarchical Modeling of Image Content

We arrange the information on five levels of different semantic abstraction as depicted in Fig. 1. The application of a family of signal models and the separation into unsupervised clustering and supervised learning resemble very much the FourEyes system [12]. However, note that from each level to the next we perform a step of Bayesian inference. This allows us to calculate on each level the most probable content description given the descriptions on the levels below.

On the lowest level, the image data (**Level 0**), we apply stochastic models M in order to obtain image features (**Level 1**). In general, these models are given as parametric data models via the likelihood

$p(D|\boldsymbol{\theta}, M)$ that assigns a probability to a given realization of the data D depending on a parameter vector $\boldsymbol{\theta}$. Thereby, different values of the parameter vector correspond to different structures in the image data (individual elements of a texture parameter vector might correspond to, e.g., degree of "strength" and "randomness" of the described texture).

We perform the process of information extraction by estimating the maximum a posteriori estimate of the parameter vector

$$\hat{\boldsymbol{\theta}}_M = \arg_{\boldsymbol{\theta}} \max p(\boldsymbol{\theta}|D, M). \tag{3}$$

An example for this kind of information extraction is the application of Gibbs-Markov random field models [18] as textural features of the image data. Note that for multi-spectral features this process of information extraction is not necessary, since these features are directly available (maybe after normalization or atmospheric correction).

The Bayesian formalism can also be applied to find the most evident model M given some data D—in contrast to the model that best describes the data, which will always be the most complex one. In order to calculate the posterior probability of the model, we have to perform marginalization

$$p(M|D) = \frac{p(M)}{p(D)} \int p(D|\boldsymbol{\theta}, M)\, p(\boldsymbol{\theta}|M)\, d\boldsymbol{\theta}, \tag{4}$$

where we have to integrate over the complete parameter space of $\boldsymbol{\theta}$. Through this integration the standard deviation of the prior distribution $p(\boldsymbol{\theta}|M)$ as compared to the width (in $\boldsymbol{\theta}$) of the likelihood $p(D|\boldsymbol{\theta}, M)$ leads to the Occam factor [22], which acts as penalty for models that are too complex to describe the existing data D.

The evidence of a particular model, Eq. (4), and the information which model best describes the data are slightly closer to the actual "meaning" of the data than the features themselves. We therefore introduce the next level of semantics (**Level 2**) and call these features of the features "meta features". Again, for multi-spectral features this complicated procedure is not applicable, whereas for textural features new information on the image content is obtained (e.g., whether the data contains more evidence for a complicated than a simple texture model [18]).

Based on the features on level 1 and the meta features on level 2 we want to derive a set of signal classes describing characteristic groups of points in the parameter spaces of the different models. This "vocabulary" of characteristic signal classes should be valid across all images (to avoid the time-consuming calculation of similarity functions) and should reflect existing structures in the different feature spaces of the data (e.g., 'fine', 'coarse', 'random' for a texture feature).

The individual elements of this "vocabulary" constitute the classes of the image classification (**Level 3**). They are obtained through unsupervised clustering of a subset of data points of the complete image archive and subsequent maximum likelihood classification of all data points using the obtained signal class characteristics. We perform the unsupervised clustering either using Bayesian classification [23], which uses multivariate Gaussian cluster models and which is able to find the optimum number of classes, or using conventional computer vision algorithms, such as k-means [24], using a predefined number of clusters. The latter approach is very similar to vector quantization and has proven to be sufficient in practical experiments.

Currently, the system is purely based on signal characteristics. The notion of "content" is therefore limited to types of ground coverage ("cover-types"). However, since the spatial resolution of remote sensing images continuously increases, geometric features have to be added. This can be either done directly on the image data (Level 0; e.g., to find 'linear structures'), on the signal class level (Level 3; e.g., to find 'circular' patches of 'fine' texture), or on the semantic level (Level 4; e.g., to find all 'lakes' with a characteristic shape).

IV. Application-specific Semantic Labeling

Levels 1 to 3 constitute a completely unsupervised characterization of the image data D on level 0. Based on this objective characterization we now define the interests of the users of the image archive (**Level 4**). We denote these subjective elements by A_ν and link them to the objective elements ω_i through the probabilities $p(\omega_i|A_\nu)$.

Using the image classification (Level 3) of a single feature is in general not sufficient to ensure a robust representation of application-specific cover-types (Level 4). However, since at the time of data insertion we do not know which features have to be combined for the later user applications, we calculate a separate vocabulary for each feature. Therefore the whole vocabulary of signal classes is decomposed into many "sub"-vocabularies

$$\omega_{jk\ldots} = \omega_{\mathrm{sp,j}} \otimes \omega_{\mathrm{tx,k}} \otimes \ldots \tag{5}$$

We exemplify the combination of different features using the combination of one spectral and one textural model (denoted by 'sp' and 'tx', respectively). However, we point out that any combination of any number of models, such as one particular texture model at several different scales, can be used in this framework.

We link the elements of the joint space of signal classes to the user interests. This can be done using several models for the likelihood $p(\omega_{jk}|A_\nu)$: one is to assume full dependence, that is,

$$p(\omega_{jk}|A_\nu) = p(\omega_{\mathrm{sp,j}}, \omega_{\mathrm{tx,k}}|A_\nu), \tag{6}$$

another is to assume conditional independence for fixed A_ν, that is,

$$p(\omega_{jk}|A_\nu) = p(\omega_{\mathrm{sp,j}}|A_\nu)\, p(\omega_{\mathrm{tx,k}}|A_\nu). \tag{7}$$

The latter is also known as "naive Bayes' classifier". Based on user examples it is possible to infer the optimal model for $p(\omega_{jk}|A_\nu)$. In this paper, however, we restrict ourselves to the discussion of the case of conditional independence.

The image classification (Level 3) provides us the posterior probabilities $p(\omega_{jk}|D)$ of the signal classes. This can be either the posterior probability of the (joint) class ω_{jk} for each pixel (e.g., obtained from Bayesian classification [23]) or just the frequency of it in a larger image window (or in the whole image as described in Sec. VI-A). Then we can calculate the posterior probability of the label A_ν as

$$p(A_\nu|D) = \sum_{jk} p(A_\nu|\omega_{jk})\, p(\omega_{jk}|D), \tag{8}$$

where we have assumed $p(A_\nu|\omega_{jk}, D) \equiv p(A_\nu|\omega_{jk})$, that is, that the signal characteristics of the cover-type A_ν are fully described by ω_{jk}. Using the Bayesian inversion formula, Eq. (1), this time between level 3 and 4, we can further write

$$p(A_\nu|D) = p(A_\nu) \sum_{jk} \frac{p(\omega_{jk}|A_\nu)\, p(\omega_{jk}|D)}{p(\omega_{jk})}, \tag{9}$$

with the priors of the signal class ω_{jk} being $p(\omega_{jk}) = \sum_\nu p(\omega_{jk}|A_\nu)\, p(A_\nu)$ and of the cover-type label $p(A_\nu)$ given by the statistics of the training regions.

Since the data D is spatial image data, we can visualize the posterior probability $p(A_\nu|D)$, Eq. (8), as an image. We call the spatial visualization of $p(A_\nu|D)$ "posterior map" and use it to communicate the current state of the cover-type definition to the user. We depict examples of posterior maps in Figs. 2 and 3.

V. Interactive Learning

Before making the inference from the image data D to the cover-type label A_ν, the system has to learn the likelihoods $p(\omega_{jk}|A_\nu)$ based on user-provided examples. This is either done directly for the classes ω_{jk} as combinations of features, or if we assume conditional independence, Eq. (7), for each feature independently. Therefore, for the following description of the learning, we restrict ourselves to denoting the signal classes by a general ω_i.

Bayesian inference provides a clean way of learning probabilities in a Bayesian network from data [13]. For the following derivation of learning the complete set of r probabilities $p(\omega_i|A_\nu)$ we assume a user-provided training data set T given by $\{N_1, \ldots, N_r\}$, with N_i being the number of instances of ω_i in T. Since ω_i is a variable with r states, the vector of N_i has a multinomial distribution [25] if we introduce the parameter vector $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_r\}$ as parametric model of the desired probabilities

$$p(\omega_i|A_\nu, \boldsymbol{\theta}) = \theta_i. \tag{10}$$

We note that for each label A_ν we need a separate parameter vector $\boldsymbol{\theta}$. However, in order to avoid awkward notations like $\theta_{i\nu}$ we assume A_ν fixed (for the moment) and omit any further ν-indices.

We now move the discussion from assessing the probability of the signal class ω_i to assessing the probability of the parameter $\boldsymbol{\theta}$, that is, we investigate the probability of "the probability of the signal class ω_i". If a new label is to be defined, then the initial prior distribution is constant

$$p(\boldsymbol{\theta}) = \Gamma(r) = (r-1)!, \tag{11}$$

with the number r of signal classes ω_i and the Gamma function $\Gamma(x)$. Note that Eq. (11) results from the normalization constraint of Eq. (10) and that it is equivalent to the number of combinatorial subsets of independent variables in $\boldsymbol{\theta}$. After observing the instances N_i of the training dataset T the posterior probability is

$$\begin{aligned} p(\boldsymbol{\theta}|T) &= \frac{p(T|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{p(T)} \\ &= \frac{\Gamma(r+N)}{\prod_i \Gamma(1+N_i)} \prod_i \theta_i^{N_i} \\ &= \mathrm{Dir}(\boldsymbol{\theta}|1+N_1, \ldots, 1+N_r), \end{aligned} \tag{12}$$

with the Dirichlet function $\mathrm{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha})$ as defined in App. A-A, the total sum $N = \sum_i N_i$ of the training samples, and the hyper-parameters

$$\alpha_i = 1 + N_i. \tag{13}$$

Note that in Eq. (12) the Gamma functions result again from the normalization constraint, but can also be derived from combinatorial considerations.

After observing another training dataset T', which we assume to be independent from T, we obtain via

$$\begin{aligned} p(\boldsymbol{\theta}|T', T) &= \frac{p(T'|\boldsymbol{\theta}, T)\, p(\boldsymbol{\theta}|T)}{p(T')} \\ &= \mathrm{Dir}(\boldsymbol{\theta}|\alpha_1 + N'_1, \ldots, \alpha_r + N'_r) \end{aligned} \tag{14}$$

a further update of the hyper-parameters by adding the number of times ω'_i in the training data set T':

$$\alpha'_i = \alpha_i + N'_i. \tag{15}$$

Therefore, each observed set of data can be incorporated as an update of the hyper-parameters. The initial state of knowledge, Eq. (11), is represented by

$$\boldsymbol{\alpha}_0 = \{1, \ldots, 1\}. \tag{16}$$

Given a description by hyper-parameters $\boldsymbol{\alpha}$ obtained through some training T, we obtain the desired probabilities as expectation over all possible values of $\boldsymbol{\theta}$. As sketched in App. A-A this results in

$$p(\omega_i|A_\nu, T) = \frac{\alpha_i}{\alpha}. \tag{17}$$

with $\alpha = \sum_i \alpha_i$ being the sum of all individual hyper parameters. In addition, analytical expressions for the variance are available.

The straightforward calculation of $p(\omega_i|A_\nu)$, Eq. (17), and the intuitive updating rule of new training data, Eq. (15), make the hyper-parameters $\boldsymbol{\alpha}$ a very convenient means for describing the probabilistic link between objective and subjective image content description as depicted in Fig. 1.

Altogether, the computational costs of learning the probabilistic definition of a user-specific cover-type and generating a posterior map of it are very moderate. It is therefore possible to run the whole process of Bayesian inference in the java virtual machine of a standard client computer and give instantaneous feedback to the user after giving a new training sample.

We depict our experimental Web interface in Fig. 2. The user defines one label at a time and gives examples for and against it (e.g., 'city' and 'not city'). The examples are given by clicking with the mouse on points in the image using the left mouse button for positive, the right one for negative examples. After each click, the posterior map of the cover-type is updated. This immediate response provides a very intuitive way of training a cover-type. We show an example sequence for the definition of 'city' in Fig. 3.

The user can only train on one image at a time, but he can issue an immediate search in the archive for his cover-type and then continue on one of the result windows. This incremental improvement is a very essential component for a robust definition of the user-interest. We expand on this in detail in Sec. VI-C.

A more severe limitation of the current interface is that training on full resolution data is only possible for selected geographic locations, otherwise quicklooks of the original images are shown to the user. However, the content-based index (Level 3) has been computed on full resolution data. Therefore, our interactive labeling interface enables the user to actually "see" into the archived data without actually getting it. This might be a very attractive feature to promote image data with a very restrictive data policy.

VI. Probabilistic Image Retrieval

In this section, we use an existing cover-type definition to retrieve images from a possibly very large image archive. First, in Sec. VI-A, we explain which information we store as content-based indices in the database system. Then, in Sec. VI-B, we derive two different ways of retrieving images from the archive: (1) by posterior probability and (2) by separability. As very important prerequisite for robust definitions of user interest, we present the concept of iterative learning in Sec. VI-C. Finally, we discuss quantitative measures to assess the quality of the cover-type definitions in Sec. VI-D.

A. Index Organization

At insertion time of each image I_ζ, the system calculates features and meta features of the models used and classifies the image according to the respective "vocabularies" as explained in Sec. III. For each image, this results in as many classification maps as models are used. From these maps, the system calculates the (frequentist) probability

$$p(\omega_{m,j}|I_\zeta) \tag{18}$$

of the jth class of the mth model in the ζth image. These probabilities are stored in a relational database system, which also manages the meta information of the images (such as date, time, or geographic coordinates). The features and the whole classifications are not stored in the database system, but put on a tape archive for later retrieval when using them for user training or further analysis.

This procedure drastically reduces the amount of stored data: e.g, for images of size $1024 \times 1024 \times 6$ channels (Landsat TM without thermal band) and one spectral and five texture models at different scales with 150 classes each, the amount of data is reduced from 6 MBytes to 3576 Bytes!

This might look like an unrealistically high compression of the information when simultaneously claiming to capture all different kinds of cover-types. However, one has to keep in mind that in order to answer queries for images of a particular cover-type (e.g., 'snow') it is sufficient to have one of the unsupervised signal classes being a very good representative of that cover-type and to have the percentage of that signal class stored for each image (which costs 4 Bytes of storage when storing probabilities as floats).

Furthermore, we point out that such high ratios of image data size to index data size will be needed when applying these content-based query techniques to full-size remote sensing archives of future missions, such as the SRTM mission to be flown in autumn 1999. A rough numerical example of the previous two X-SAR missions in 1994 with a total of 37000 images results in a data size of 3 TBytes compared to 27 MBytes of index data size.

B. Calculation of Posterior Probability and Separability

In a similar way as calculating the posterior probability of A_ν given a particular data D by Eq. (9), we can calculate the posterior probability of A_ν given the whole image I_ζ

$$p(A_\nu|I_\zeta) = \sum_i p(A_\nu|\omega_i)\, p(\omega_i|I_\zeta), \tag{19}$$

where we have assumed $p(A_\nu|\omega_i, I_\zeta) \equiv p(A_\nu|\omega_i)$. Since the likelihood $p(\omega_i|A_\nu)$ is known as a probability distribution (that is, as a "probability of a probability" as discussed in detail in Sec. V), we have to calculate the expectation of the posterior $p(A_\nu|\omega_i)$ as described in App. A-D.

The probability $p(\omega_i|I_\zeta)$ is the frequency of ω_i in I_ζ. When combining classes of different signal models, we need to know $p(\omega_{\mathrm{sp,j}}, \omega_{\mathrm{tx,k}}|I_\zeta)$ to calculate the posterior $p(A_\nu|I_\zeta)$. To enable a fast search and to avoid the calculation of the joint histogram of $\omega_{\mathrm{sp,j}}$ and $\omega_{\mathrm{tx,k}}$ of each image I_ζ, we use the approximation

$$p(\omega_{\mathrm{sp,j}}, \omega_{\mathrm{tx,k}}|I_\zeta) \sim p(\omega_{\mathrm{sp,j}}|I_\zeta) \cdot p(\omega_{\mathrm{tx,k}}|I_\zeta), \tag{20}$$

which is identical to the maximum entropy estimation of $p(\omega_{\mathrm{sp,j}}, \omega_{\mathrm{tx,k}}|I_\zeta)$ for a given pair of marginal probabilities $p(\omega_{\mathrm{sp,j}}|I_\zeta)$ and $p(\omega_{\mathrm{tx,k}}|I_\zeta)$. This approximation allows us to calculate a rough estimate of $p(A_\nu|I_\zeta)$ from the index described in the previous subsection.

In Fig. 4, we show the retrieved images according to the posterior probability of the cover-type 'forest' from an archive of aerial images. The posterior probability is a measure of how probable an image "is of" a particular cover-type. Therefore, a value of 0.5 can mean everything from "50 percent of this image is of that cover-type with certainty" up-to "on the whole image it can be either that cover-type or not".

Since the distribution of $p(\omega_i|A_\nu)$—resulting from limited training data—is known in detail, we can specify not only how probable a label is in a particular image, but also how much variation is expected. We do this by calculating the expected variance of the posterior

$$\delta^2 p(A_\nu|I_\zeta) = \sum_i \delta^2 p(A_\nu|\omega_i)\, p(\omega_i|I_\zeta), \tag{21}$$

with the δ^2-symbol denoting the variance of the following quantity. We expand on the calculation of the variance $\delta^2 p(A_\nu|\omega_i)$ of the posterior in App. A-D. Note that in Eq. (21) we make use of the same assumption as in Eq. (19).

As measure of how well A_ν is separated from $\neg A_\nu$ in a particular image I_ζ we introduce the separability

$$S(A_\nu|I_\zeta) = \frac{\delta^2 p(A_\nu|I_\zeta)}{p(A_\nu|I_\zeta)\,(1 - p(A_\nu|I_\zeta))}, \tag{22}$$

which is the variance in units of the maximal possible variance (if a normalized distribution $\in [0,1]$ has mean p then the maximal possible variance is $p(1-p)$). The smaller S the "better" we call the separability. Both our retrieval examples, Fig. 4 and 5, list also the images of best and worst separability.

Retrieval according to the separability of a particular cover-type promises to be a very valuable tool for image retrieval, in particular from remote sensing image archives with many different sensors. Furthermore, it is very helpful for the cover-type definition itself as we explain in the following section.

C. Iterative, incremental Learning

For a robust cover-type definition it is necessary to train the stochastic link of the cover-type using examples on several images. Therefore, our system provides the possibility to continue training on the results of a search. After an initial training on a particular image—maybe on an image of a known location obtained via conventional meta search—the user invokes a first search and obtains the search results, which are ranked according to posterior probability, Eq. (19) and according to separability, Eq. (22). Of these result images, the user can now select an image for continuation of training that is—from his point of view—at the wrong place of the ranking (e.g., an image displaying large areas of his cover-type is among the images with bad separability).

We sketch this iterative learning in Fig. 6. The user can repeat the loop of interactive training and probabilistic retrieval as often as it is necessary for the results to converge to a satisfying result—or as much as necessary to find out that the selected content index is not capable of representing his cover-type.

If the process of definition of a particular cover-type is successful, then the user can insert "his" label into the inventory of user-interests. There the cover-type name and its definition via hyper-parameters are stored. We show part of this inventory together with the signal models used in Tab. I. All labels have been defined by users on the Internet and can be used for probabilistic retrieval.

An existing label can also serve as the basis for the definition of a new label. In this case, the hyper-parameters of the existing label are transferred to the user's client computer and modified by him with additional examples (e.g., a forest expert might take an existing rough definition of 'forest' to define several sub-classes, such as 'burnt forest' or 'young forest').

D. Analysis of the Defined Labels

Once the stochastic link $p(\omega_i|A_\nu)$ between subjective user interests (Level 4) and objective feature classes (Level 3) is established, the question of the quality of the search results arises. We have to differentiate between two aspects of quality: (1) the quality of the stochastic link and (2) the quality of the resulting search results.

The quality of the stochastic link is directly accessible using information-theoretic principles. Kullback [27] defines the "divergence" between two complete sets of probabilities $\mathcal{A}_\nu = \{p(\omega_1|A_\nu), \ldots, p(\omega_r|A_\nu)\}$ and $\neg\mathcal{A}_\nu = \{p(\omega_1|\neg A_\nu), \ldots, p(\omega_r|\neg A_\nu)\}$ as

$$D(\mathcal{A}_\nu, \neg\mathcal{A}_\nu) = \sum_{i=1}^{r} (p(\omega_i|A_\nu) - p(\omega_i|\neg A_\nu)) \ln \frac{p(\omega_i|A_\nu)}{p(\omega_i|\neg A_\nu)}, \tag{23}$$

which is the symmetric sum of two relative entropies. The divergence $D(\mathcal{A}_\nu, \neg\mathcal{A}_\nu)$ can be viewed as distance between the two probability distributions $\mathcal{A}_\nu$ and $\neg\mathcal{A}_\nu$. This distance can be calculated for the complete link to the combination of signal classes or for individual signal models. It can serve as a guide to keep strong and remove weak features for training of a particular label (e.g., when training a particular kind of 'clouds' it might turn out that the link to spectral classes is 'weak' and to texture classes on one scale is 'strong'. Then it might be a good idea to experiment with texture classes on two different scales.).

Another easily accessible quality measure is the variance of the posterior probability of a cover-type in the whole image, Eq. (21). This variance arises from the limited amount of available training data and will become smaller the more training examples are given. We have discussed details on how to calculate and interpret the variance of the posterior in the two preceding sections.

A quantity that is very difficult to obtain is the actual quality of the search results. This is due to the fact that, in addition to the quality of the stochastic link and its uncertainty, the "quality of the user examples" and the "user's judgment of a completely defined label" have to be taken into account. Both items can also be seen as the "honesty" and "care" of the user when defining the label (e.g., when a user defines the label 'forest' and gives example of 'lake' or when a label is trained on a single image and the search result is not thoroughly examined for badly ranked images).

Therefore, a concept of validation of the defined labels has to be developed. In the current system the validation is the task of the user defining his label. But how to detect "dishonest" or just "lazy" cover-type definitions? A possible approach could be a validation by selected experts of different fields of applications, who do not need to have remote sensing experience, but only significant knowledge in their application field or of a particular geographic area.

VII. Practical Applications

Our concept of unsupervised image content indexing and user-specific semantic labeling is currently being extensively tested using large volumes of data from different sensor types and resolutions. We show the current content of our test archive together with the signal models used in Tab. II. The complete archive can be accessed using the content-based approach presented in this paper and user-specific cover-types can be interactively defined for each sensor via the Web and added to the public inventory.

In addition to the Web interface presented in this paper, we have also developed an interface for direct analysis of the content-based image classifications on level 3 using the full image resolution. It uses exactly the same formalism for learning the stochastic link and providing posterior maps of cover-types. This tool can be used for the fast and automatic generation of thematic maps based on user-provided training regions.

In order to demonstrate the accuracy of the content-based indexing and the semantic inference process, we use the stochastic confusion matrix

$$M_{\nu\mu} = \sum_{jk} p(A_\nu|\omega_{jk})\, p(\omega_{jk}|A_\mu) \tag{24}$$

as a measure of how well the defined labels are represented in the space of signal classes. This matrix is a rough measure of the error matrix to be expected when using the inference process for classification.

For the set of panchromatic aerial images in our test archive various land cover types have been defined and large parts of the images have been manually segmented. In Tab. III, we show the stochastic confusion matrix $M_{\nu\mu}$ for the land cover classes based on the manually provided training regions and the indices derived from Gibbs-Markov Random Fields models at scales 1m and 2m. Since, for that particular application, no label should be preferred against another, a uniform prior $p(A_\nu)$ was used. The average percentage of correctly reproduced land cover types is larger than 90%. Given

the fact that even humans might have problems in differentiating certain cover-types (e.g., 'forest' and 'young forest'), these results are quite promising for future applications.

The extension with additional data, both optical (IRS, SPOT, NOAA/AVHRR) and SAR (images of the SRTM mission), are currently being prepared. Furthermore, the extension with additional available information, such as height and slope derived from digital elevation models, is straightforward.

VIII. Summary

In this paper, we have presented a concept to bridge the gap between modern approaches of information mining in large remote sensing information systems and the actual user interests. The use of Bayesian networks enabled us to build a very intuitive user interface that provides almost instantaneous feedback to the user, thus making the process of cover-type definition genuinely interactive. The concept of probabilistic retrieval opens a completely new perspective for the design of future content-based image archives: queries like "show me all images on which snow and ice can be separated" (without specifying a particular sensor) might be answered in such archives.

The concept presented in this paper can help new users of remote sensing to understand the data and to find appropriate images and methods for further analysis. In combination with strong signal models and modern data-mining techniques, it can help experts in exploring the potentials and capabilities of data coming from new sensors types. Altogether, interactive labeling and probabilistic retrieval promise to be very valuable tools for future remote sensing image archives.

IX. Acknowledgment

We thank all users of our system, and are grateful to them for reporting feedback to us. The Landsat TM data has been provided by the Swiss National Point of Contact (NPOC) and the X-SAR data by DLR/DFD. We are thankful to Dominic Brander (ETHZ) for providing the set of aerial photographs and his manual segmentation. The X-SAR images have been processed by Gintautas Palubinskas and Marc Walessa (DLR/DFD). Furthermore, we thank the two anonymous reviewers for their fruitful comments.

Appendix

I. Uncertain Probabilities in the Naive Bayes Classifier

In this appendix, we summarize the results of some lengthy calculations that lead to closed-form expressions for the most important uncertainties in the naive Bayesian classifier.

A. Mean and Variance of a Single Likelihood

As introduced in Sec. V, we use $\boldsymbol{\theta}$ to model the likelihood as $p(\omega_i|A_\nu, \boldsymbol{\theta}) = \theta_i$. As also mentioned in the main text, the posterior distribution of $\boldsymbol{\theta}$ after training T is a Dirichlet distribution

$$\begin{aligned} p(\boldsymbol{\theta}|T) &= \mathrm{Dir}(\boldsymbol{\theta}|\alpha_1, \ldots, \alpha_r) \\ &= \frac{\Gamma(\alpha)}{\prod_{i=1}^{r} \Gamma(\alpha_i)} \prod_{i=1}^{r} \theta_i^{\alpha_i - 1}, \end{aligned} \tag{25}$$

with the hyper-parameters $\boldsymbol{\alpha}$. Therefore, we can obtain the likelihood via integration over all $\boldsymbol{\theta}$ as

$$\begin{aligned} p(\omega_i|A_\nu, T) &= \mathrm{E}\left[\theta_i\right] \\ &= \int \theta_i \, p(\boldsymbol{\theta}|T) \, d\boldsymbol{\theta} = \frac{\alpha_i}{\alpha}, \end{aligned} \tag{26}$$

where the integration extends over the allowed parameter space of $\boldsymbol{\theta}$ as given by the normalization constraint of the likelihood. Furthermore, we obtain the variance of the likelihood as

$$\begin{aligned}\delta^2 p(\omega_i|A_\nu,T) &= \mathrm{Var}\left[\theta_i\right] = \mathrm{E}\left[\theta_i^2\right] - \mathrm{E}\left[\theta_i\right]^2 \\ &= \frac{\mathrm{E}\left[\theta_i\right]\left(1-\mathrm{E}\left[\theta_i\right]\right)}{\alpha+1}.\end{aligned} \tag{27}$$

Note, that for a strict notation, we always have to specify the state of knowledge T on the right side of the conditioned probability. However, if the state of knowledge is clear from the context, we omit T in order to avoid awkward notations.

B. Joint Likelihood

The naive Bayes classifier is based on the assumption of conditional independence of the likelihood of A_ν

$$p(\omega_{jk}|A_\nu) = p(\omega_{\mathrm{sp,j}}|A_\nu)\, p(\omega_{\mathrm{tx,k}}|A_\nu). \tag{28}$$

For each term on the right, we have a distribution in terms of hyper-parameters of the Dirichlet function, for $p(\omega_{\mathrm{sp,j}}|A_\nu)$ by $\boldsymbol{\alpha}$ and for $p(\omega_{\mathrm{tx,k}}|A_\nu)$ by $\boldsymbol{\beta}$. Since these distributions are independent, we directly obtain the mean value as product of the individual means

$$\mathrm{E}\left[p(\omega_{jk}|A_\nu)\right] = \frac{\alpha_j}{\alpha}\cdot\frac{\beta_k}{\beta}, \tag{29}$$

where we have twice made use of Eq. (26).

To calculate the variance of Eq. (28), we use

$$\mathrm{Var}\left[X\cdot Y\right] = \mu_y^2\mathrm{Var}\left[X\right] + \mu_x^2\mathrm{Var}\left[Y\right] + \mathrm{Var}\left[X\right]\mathrm{Var}\left[Y\right], \tag{30}$$

which relates the variance of the product of the two independent stochastic variables X and Y to their individual means and variances (with the short notation $\mu_x = \mathrm{E}\left[x\right]$). By using this rule and Eqns. (26) and (27) we obtain

$$\begin{aligned}\mathrm{Var}\left[p(\omega_{jk}|A_\nu)\right] &= \frac{\alpha_j\beta_k}{\alpha\beta(\alpha+1)(\beta+1)} \\ &\cdot\left[(1+\alpha_j+\beta_k) - \frac{\alpha_j\beta_k}{\alpha\beta}(1+\alpha+\beta)\right].\end{aligned} \tag{31}$$

C. Joint Prior

Since the likelihood exhibits some uncertainty, also the prior $p(\omega_{jk})$

$$p(\omega_{jk}) = p(\omega_{jk}|A_\nu)\, p(A_\nu) + p(\omega_{jk}|\neg A_\nu)\, p(\neg A_\nu) \tag{32}$$

has some nonzero variance. However, since the likelihoods of A_ν and $\neg A_\nu$ are independent, mean and variance of the prior are obtained by merely summing the corresponding terms using Eqns. (29) and (31), respectively.

D. Joint Posterior

To obtain the mean and variance of the posterior

$$p(A_\nu|\omega_{jk}) = \frac{p(\omega_{jk}|A_\nu)\, p(A)}{p(\omega_{jk})}, \tag{33}$$

we have to calculate

$$\mathrm{E}\left[\frac{X}{X+Y}\right] \quad \text{and} \quad \mathrm{Var}\left[\frac{X}{X+Y}\right] \tag{34}$$

with $X = p(\omega_{jk}|A_\nu)\,p(A_\nu)$ and $Y = p(\omega_{jk}|\neg A_\nu)\,p(\neg A_\nu)$ being two independent stochastic variables. We obtain an approximation formula for the mean by performing a Taylor series expansion of $X/(X+Y)$ around $X = \mu_x$ and $Y = \mu_y$, keeping only terms of order two or less and then taking the expectation. This results in

$$\mathrm{E}\left[\frac{X}{X+Y}\right] \sim \frac{\mu_x}{\mu_x+\mu_y} + \frac{\mu_x \mathrm{Var}\,[Y] - \mu_y \mathrm{Var}\,[X]}{(\mu_x+\mu_y)^3}. \tag{35}$$

Through the same procedure for the variance, we obtain a similar, rather lengthy expression for the variance. These approximations allow us to calculate both the expected posterior and its variance in the presence of uncertainties due to limited training samples. In the main text, in Sec. VI, we present the concept of probabilistic retrieval and separability, which are based on the uncertainties presented in this appendix.

References

[1] J. Van der Lubbe, E. Backer, and R. Goedhart, "An expert system for the automatic analysis of cloud types in satellite imagery: The METEODES project," in *Proc. of the 11th Int. Conf. on Expert Systems and Their Applications*, 1991, pp. 37–48.

[2] S. Matwin, D. Charlebois, D. Goodenough, and P. Bhogal, "Machine learning and planning for data management in forestry," *IEEE-Expert.*, vol. 10, no. 6, pp. 35–41, 1995.

[3] C. Bertoia and B. Ramsay, "Sea ice analysis and products: cooperative work at the US and Canadian national ice centers," in *Proc. IGARSS'98*, 1998, vol. 4, pp. 1944–1947.

[4] M. Burl, L. Asker, P. Smith, U. Fayyad, P. Perona, L. Crumpler, and J. Aubele, "Learning to recognize volcanoes on Venus," *Machine Learning*, vol. 30, pp. 165–195, 1998.

[5] H. Liu and H. Motoda, Eds., *Feature Extraction, Construction and Selection: A Data Mining Perspective*, Kluwer, 1998.

[6] R. S. Michalski, I. Bratko, and M. Kubat, Eds., *Machine Learning and Data Mining*, John Wiley, 1998.

[7] G. Nakhaeizadeh, Ed., *Data Mining—Theoretische Aspekte und Anwendungen*, Physica-Verlag, 1998.

[8] G. Marchisio and J. Cornelison, "Content-based search and clustering of remote sensing imagery," in *Proc. IGARSS'99*, 1999, vol. 1, pp. 290–292.

[9] M. Burl, Ch. Fowlkes, J. Roden, A. Stechert, and S. Mukhtar, "Diamond Eye: A distributed architecture for image data mining," in *SPIE DMKD, Orlando*, April 1999.

[10] K. Seidel, M. Datcu, G. Schwarz, and L. Van Gool, "Advanced remote sensing information systems at ETH Zürich and DLR," in *Proc. IGARSS'99*, 1999, vol. 5, pp. 2363–2365.

[11] 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, no. 5, pp. 1431–1445, Sept. 1998.

[12] T. P. Minka and R. W. Picard, "Interactive learning with a 'society of models'," *Pattern Recognition*, vol. 30, no. 4, pp. 565–581, 1997.

[13] D. Heckerman, "A tutorial on learning with Bayesian networks," Tech. Rep. MSR-TR-95-06, Microsoft Research, Advanced Technology Division, 1996, Available on the WWW.

[14] U. Benz, "Supervised fuzzy analysis of single- and multichannel SAR data," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 2, pp. 1023–1037, March 1999.

[15] Peter Schäuble, *Multimedia Information Retrieval—Content-Based Information Retrieval from Large Text and Audio Databases*, Kluwer Academic Publishers, 1997.

[16] SUN Microsystems, "Java," http://www.javasoft.com/.

[17] M. Schröder and H. Rehrauer, "WWW online demonstration: Interactive learning and probabilistic retrieval," http://www.vision.ee.ethz.ch/∼rsia/ClickBayes, 1999, At least Netscape Communicator 4.06 or Internet Explorer 4.0 required.

[18] 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, no. 5, pp. 1446–1455, Sept. 1998.

[19] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[20] Gerhard Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer, Berlin, 1995.

[21] N. Vasconcelos and A. Lippman, "A Bayesian framework for semantic content characterization," in *Proc. CVPR'98*, 1998, pp. 566–571.

[22] D. Sivia, *Data Analysis—A Bayesian Tutorial*, Clarendon Press, Oxford, 1996.

[23] P. Cheeseman and J. Stutz, "Bayesian classification (AutoClass): Theory and results," in *Advances in Knowledge Discovery and Data Mining*. MIT Press, 1995.

[24] R. Dubes and A. Jain, "Clustering methodologies in exploratory data analysis," in *Advances in Computers*, M. Yovits, Ed., vol. 19, pp. 113–228. Academic Press, 1980.

[25] J. Bernardo and A. Smith, *Bayesian Theory*, Wiley, 1994.

[26] M. Walessa, "Texture preserving despeckling of SAR images using GMRFs," in *Proc. IGARSS'99*, 1999, vol. 3, pp. 1552–1554.

[27] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, pp. 79–86, 1951.

[28] R. M. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

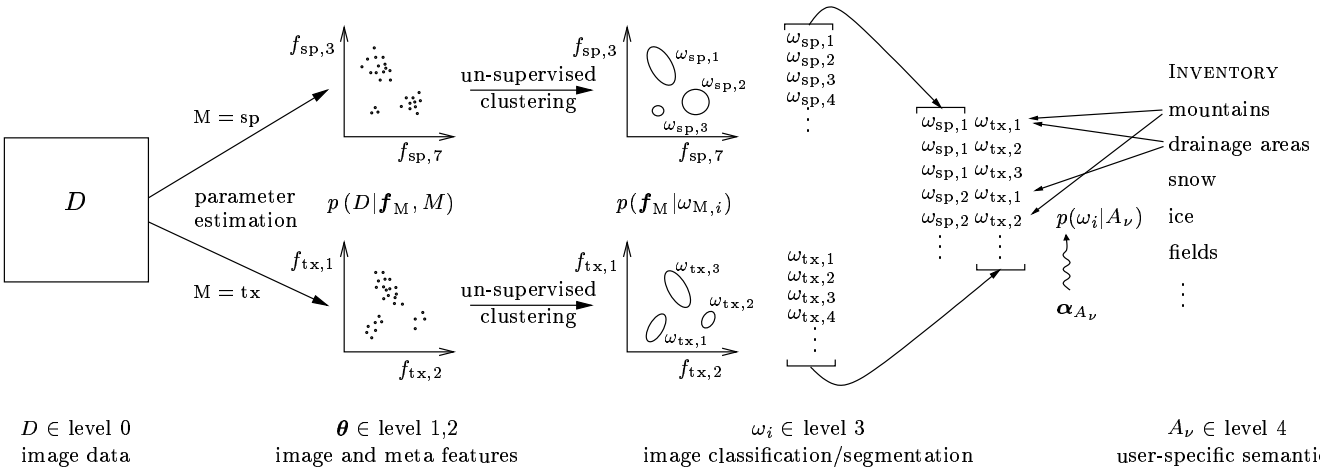

Fig. 1

Hierarchical scheme of image content description. After extracting features and meta features (Levels 1 and 2) of the image data (Level 0), we obtain a "characteristic vocabulary" of signal classes ω_i (Level 3) for each signal model M. User-specific interests, that is, cover-type labels A_ν, are linked to combinations of these vocabularies using simple Bayesian networks, characterized by the probabilities $p(\omega_i|A_\nu)$. Levels 1 to 3 are obtained in a completely unsupervised and application-free way during data insertion. Elements of Level 4 can be interactively defined by users using the posterior map $p(A_\nu|D)$, Eq. (8), and the interface depicted in Fig. 2. In an additional step of stochastic modeling, we describe the stochastic link $p(\omega_i|A_\nu)$ using a vector of hyper-parameters $\boldsymbol{\alpha}_{A_\nu}$.

TABLE I

Some Cover-Types in the Public Inventory WWW Link: view the current content of the invenory!

Name A_ν	Sensor	Model 1	Model 2
lake	TM	**spectral**	**GRF, 25m**
clouds	TM	**spectral**	GRF, 25m
cumulus	TM	spectral	**GRF, 25m**
forest	TM	**spectral**	GRF, 25m
fields	TM	**GRF, 25m**	**GRF, 50m**
snow	TM	**spectral**	GRF, 25m
village	aerial	intensity	**GRF, 1m**
forest	aerial	**GRF, 1m**	**GRF, 2m**
mountains	X-SAR	**MBD, 50m**	**MBD, 100m**
jura (mt.)	X-SAR	intensity	**MBD, 25m**

GRF, MBD see Tab. II
bold=strong link with distance $D(\mathcal{A}_\nu, \neg\mathcal{A}_\nu) > 3$

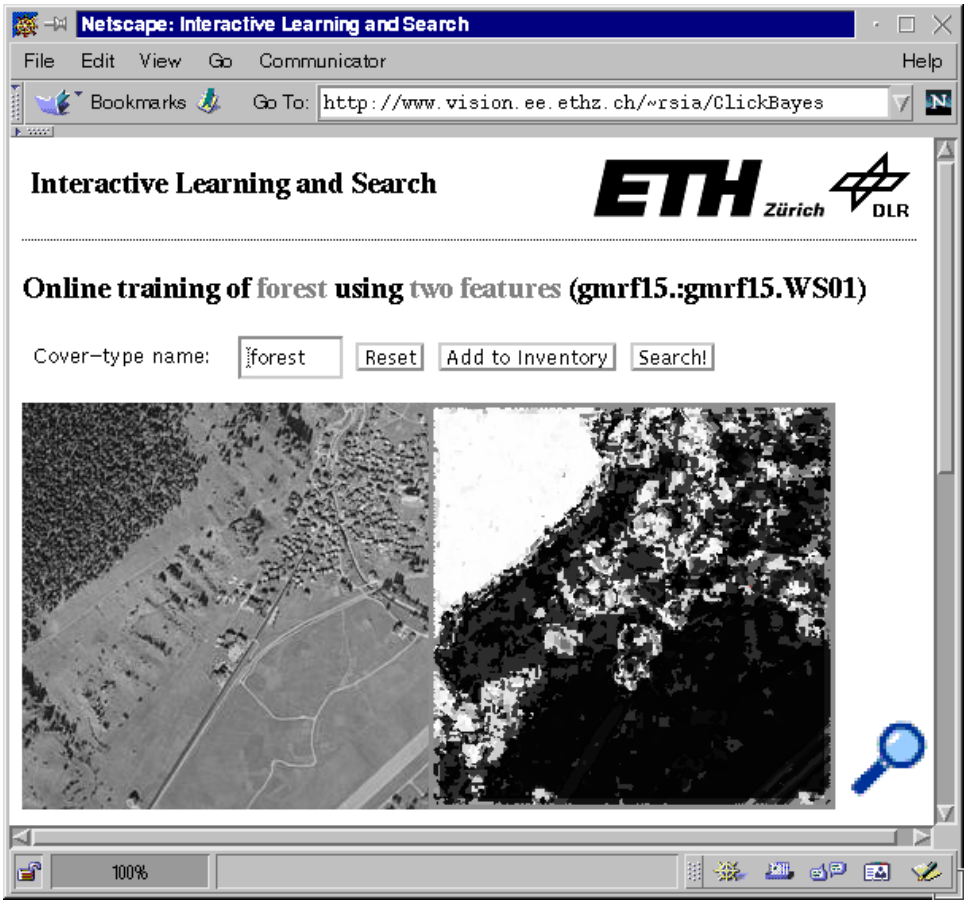

Fig. 2

ONLINE INTERFACE FOR INTERACTIVE LEARNING OF REMOTE SENSING IMAGE CONTENT. ON THE LEFT SIDE THE USER CAN SEE THE ORIGINAL IMAGE (THE EXAMPLE SHOWS AN AERIAL IMAGE OF A VILLAGE IN SWITZERLAND), IN WHICH HE CAN GIVE POSITIVE OR NEGATIVE EXAMPLES FOR HIS COVER-TYPE OF INTEREST BY CLICKING INTO IT WITH THE LEFT OR RIGHT MOUSE BUTTON, RESPECTIVELY. AFTER EACH CLICK, THE HYPER-PARAMETERS OF THE PROBABILISTIC LINK ARE UPDATED, EQ. (15), AND THE POSTERIOR MAP ON THE RIGHT SIDE OF THE DISPLAY IS RECOMPUTED, EQ. (9), AND REDISPLAYED (BLACK TO WHITE DENOTE CONTINUOUSLY 0 TO 1—HERE, WE HAVE REMOVED AN ADDITIONAL COLOR CODING OF THE PROBABILITY). WE SHOW AN EXAMPLE SEQUENCE OF THIS KIND OF LEARNING IN FIG. 3. THE USER PROCEEDS UNTIL HE IS SATISFIED WITH THE POSTERIOR MAP. THEN HE CAN USE THE NEWLY DEFINED COVER-TYPE FOR PROBABILISTIC RETRIEVAL AS SHOWN IN FIG. 4. WWW LINK: TRY IT OUT NOW!

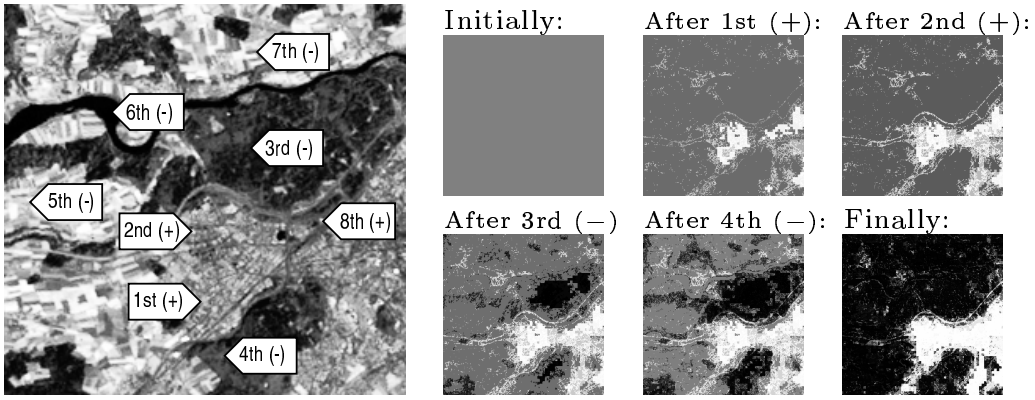

Fig. 3

Example sequence of interactive learning using the interface depicted in Fig. 2. In the left image we show the original image (Landsat TM, band 5) with the sequence of example points given by the user (denoted by the tips of the arrows and marked by (+) and (−) for positive and negative examples, respectively). The small images on the right show how the posterior map, Eq. (8), for the cover-type 'city' evolves. After two positive examples for 'city', the user has to give several negative examples, until he might be satisfied with his definition. Then he can use the newly defined cover-type label for a search in the whole image archive. In this example we have used the signal class indices resulting from normalized spectral intensities and from a Gibbs random field model at scale 25m. WWW Link: Redo this experiment now!

TABLE II
Images in the Test Archive and Precomputed Signal Models

Sensor	Landsat TM	X-SAR	aerial
no. of images	184	110	66
channels, size	7, 1024^2	1, 2048^2	1, 1024^2
resolution	25m, geo.	12.5m	1m, geo.
signal models (levels 1–3)	spectral, GRF 5 scales, GLCM 5 scales	intensity, MBD 3 scales,	intensity, GRF 5 scales

GRF='Gibbs Random Fields', [18]
GLCM='Gray Level Cooccurrence Matrix', [28]
MBD='Model Based Despeckling', [26]
geo.=geocoded

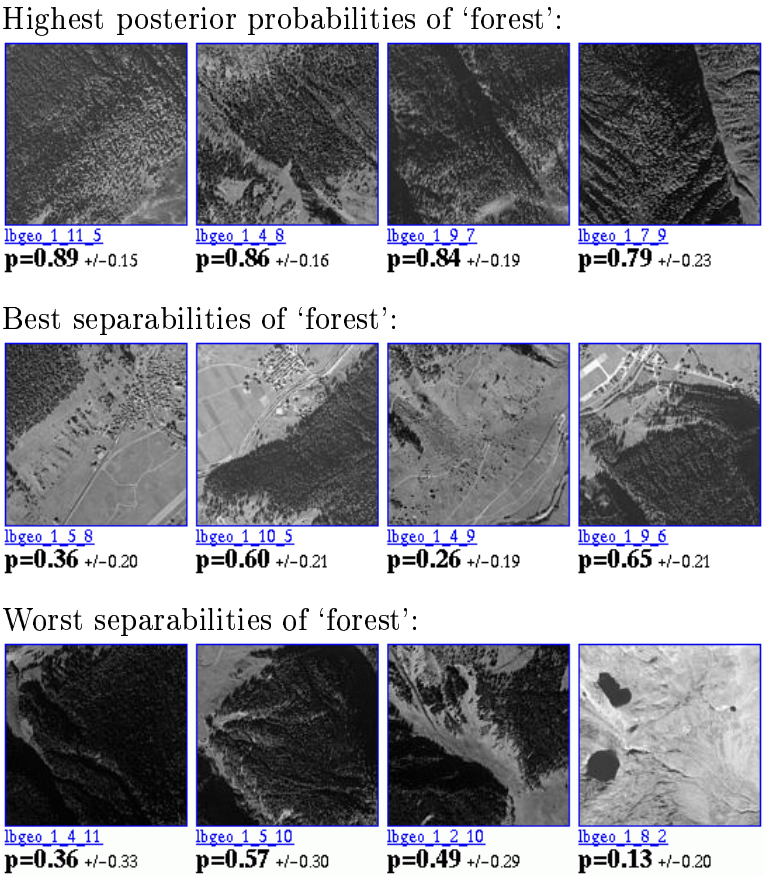

Fig. 4

PROBABILISTIC RETRIEVAL OF 'FOREST' FROM A COLLECTION OF AERIAL IMAGES. IN THE FIRST ROW, WE DISPLAY THE IMAGES WITH HIGHEST POSTERIOR PROBABILITY p OF 'FOREST', EQ. (19). THESE IMAGES ARE THE MOST LIKELY IMAGES IN THE DATABASE TO CONTAIN 'FOREST'. SINCE WE KNOW THE PROBABILISTIC LINK INCLUDING ITS UNCERTAINTIES, WE CAN, IN ADDITION, SPECIFY THE POSSIBLE RANGE OF THE POSTERIOR USING THE STANDARD DEVIATION OBTAINED FROM THE VARIANCE $\delta^2 p$, EQ. (21). WE DISPLAY THE IMAGES WITH BEST AND WORST SEPARABILITY $S = \delta^2 p/p(1-p)$ (SEE EQ. (22)) IN THE MIDDLE AND LOWER ROW, RESPECTIVELY. THROUGH THE RANKING ACCORDING TO THE SEPARABILITY, THE USER OF THE ARCHIVE CAN ASSESS THE QUALITY OF HIS LABEL DEFINITION AND DECIDE ON WHICH IMAGE HE SHOULD CONTINUE TRAINING. THE VALUES OF S ARE 0.17, 0.18, 0.19, AND 0.20 (MIDDLE ROW) AND 0.49, 0.36, 0.35, AND 0.34 (LOWER ROW). FOR THIS EXAMPLE, WE USED THE GIBBS TEXTURE INDEX AT SCALES 1M AND 2M AND WE GAVE 27 POSITIVE AND 16 NEGATIVE EXAMPLES ON FIVE TRAINING IMAGES. THE ICONS DISPLAY THE IMAGES AT SCALE 8M. WWW LINK: REDO THIS SEARCH NOW!

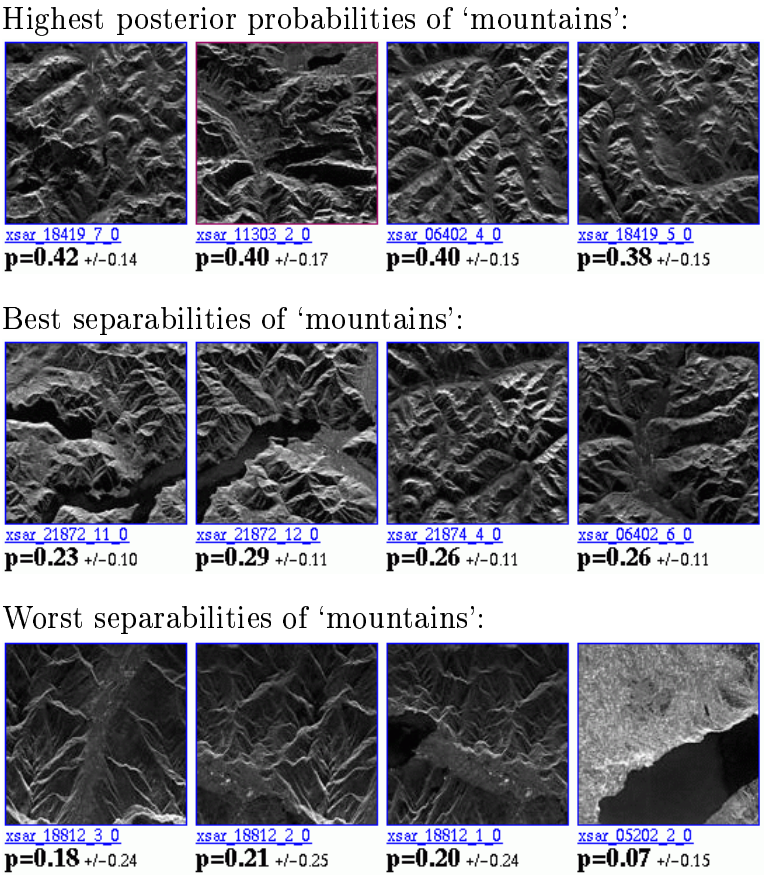

Fig. 5

PROBABILISTIC RETRIEVAL OF A PARTICULAR TYPE OF 'MOUNTAINS' FROM A COLLECTION OF X-SAR IMAGES. WE DISPLAY THE RESULTS IN THE SAME WAY AS IN FIG. 4. THE VALUES OF THE SEPARABILITY S ARE 0.05, 0.06, 0.07, AND 0.07 (MIDDLE ROW) AND 0.38, 0.38, 0.37, AND 0.35 (LOWER ROW). THE IMAGES WITH WORST SEPARABILITY EXHIBIT IMAGE CONTENT NOT WELL KNOWN TO THE SYSTEM SO FAR (PROBABLY 'MOUNTAINS WITH LAYOVER AND FORESHORTENING' AND 'LAKE WITH SURFACE ROUGHNESS') RESULTING IN A BAD SEPARABILITY. THEREFORE, THE USER SHOULD CONTINUE THE TRAINING PROCESS OF 'MOUNTAINS' ON ONE OF THESE IMAGES. IN THIS EXAMPLE, WE HAVE DEFINED 'MOUNTAINS' USING THE SIGNAL CLASS INDICES DERIVED FROM THE ESTIMATED PARAMETERS OF "MODEL BASED DESPECKLING" [26] AT SCALES 50M AND 100M. WWW LINK: REDO THIS SEARCH NOW!

TABLE III

STOCHASTIC CONFUSION MATRIX USING MANUAL SEGMENTATION AND GIBBS TEXTURE INDICES AT SCALES 1M AND 2M

	$M_{\nu\mu}$, Eq. (24), in %					
A_μ	A_1	A_2	A_3	A_4	A_5	A_6
A_1: 'forest'	87	1	1	2	7	1
A_2: 'village'	1	91	0	4	2	1
A_3: 'fields'	1	0	95	0	1	2
A_4: 'shade'	2	4	0	91	1	2
A_5: 'young forest'	7	2	1	1	89	1
A_6: 'stony ground'	1	1	2	2	1	93

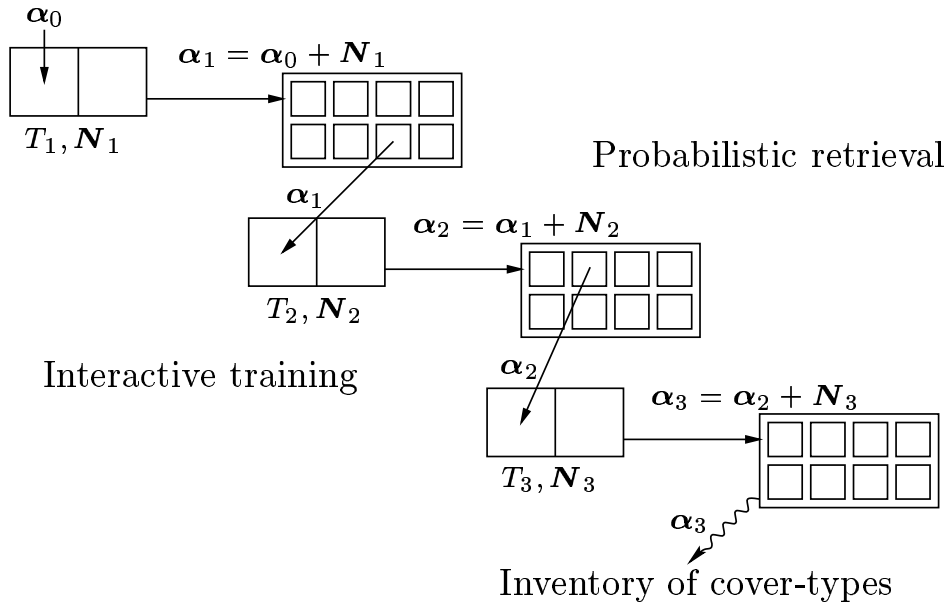

Fig. 6

SCHEME OF ITERATIVE LEARNING USING SEVERAL LOOPS OF INTERACTIVE TRAINING (LEFT BOXES, SEE FIG. 2) AND PROBABILISTIC RETRIEVAL (RIGHT BOXES, SEE FIG. 4). BASED ON AN IMAGE TO START WITH AND AN INITIAL DESCRIPTION $\boldsymbol{\alpha}_0$ OF THE STOCHASTIC LINK BETWEEN LEVEL 3 AND 4, THE USER PROVIDES ADDITIONAL TRAINING EXAMPLES T_1 THAT RESULT IN AN UPDATED HYPER-PARAMETER $\boldsymbol{\alpha}_1$. HE CAN CONTINUE THE INTERACTIVE TRAINING ON THE RESULT IMAGE OF HIS CHOICE AND REPEAT THAT LOOP UNTIL THE RETRIEVAL RESULTS CONVERGE FROM THE HIS POINT OF VIEW. THEN THE USER CAN PUT THE DEFINITION OF THE NEWLY DEFINED LABEL IN THE INVENTORY OF COVER-TYPES (SEE TAB. I). HERE WE DEPICT ONLY THREE TRAINING LOOPS, BUT MANY MORE MIGHT BE NECESSARY.