

# Image Information Mining System Evaluation Using Information-Theoretic Measures

**Herbert Daschiel**

*German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Oberpfaffenhofen, 82230 Wessling, Germany  
Email: herbert.daschiel@dlr.de*

**Mihai Datcu**

*German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Oberpfaffenhofen, 82230 Wessling, Germany  
Email: mihai.datcu@dlr.de*

*Received 18 December 2003; Revised 2 September 2004*

During the last decade, the exponential increase of multimedia and remote sensing image archives, the fast expansion of the world wide web, and the high diversity of users have yielded concepts and systems for successful content-based image retrieval and image information mining. Image data information systems require both database and visual capabilities, but there is a gap between these systems. Database systems usually do not deal with multidimensional pictorial structures and vision systems do not provide database query functions. In terms of these points, the evaluation of content-based image retrieval systems became a focus of research interest. One can find several system evaluation approaches in literature, however, only few of them go beyond precision-recall graphs and do not allow a detailed evaluation of an interactive image retrieval system. Apart from the existing evaluation methodologies, we aim at the overall validation of our knowledge-driven content-based image information mining system. In this paper, an evaluation approach is demonstrated that is based on information-theoretic quantities to determine the information flow between system levels of different semantic abstraction and to analyze human-computer interactions.

**Keywords and phrases:** image information mining, CBIR, evaluation, information theory, human-computer interaction.

## 1. INTRODUCTION

In recent years, the growth of data collected and stored in archives and the access via the world wide web has greatly exceeded our ability to significantly extract user-relevant information from the data. This has resulted in combined efforts to develop new methods and concepts to manage large volumes of data: content-based image retrieval (CBIR) [1], data mining [2], knowledge discovery in databases [3], and information visualization. A broad range of techniques was developed to deal either with particular data types, such as text and numerical records, and also with heterogeneous data types, for example, combining video and sound. But one of the most complex tasks still remains: the access of image information. Image retrieval systems require both database and machine vision capabilities, but a gap exists between these systems. Until recently, standard databases did not deal with multidimensional pictorial structures and computer vision pattern recognition systems did not provide database query capabilities. Since the size and the information content of imagery keeps increasing, particularly in the remote sensing domain, many new applications in content-based image retrieval are closer to computer vision and require the knowl-

edge and understanding of complicated spatial and structural characteristics among image objects.

As in other fields, however, further development in image information mining depends on the capability to carefully evaluate the image retrieval and image understanding functions and methods. Such an evaluation should include the technical (objective) quality of a system as well as user-related (subjective) concepts. The basic problem in CBIR performance evaluation is the lack of standardization: defining a common image database and relevance judgments. One can already find a considerable literature on evaluating content-based image retrieval systems [4]. Most of the evaluation approaches are based on query by image example and “relevant” versus “nonrelevant” classes of images. They mainly stay on the level of precision-recall graphs and only validate the image retrieval system function. A work very close to our system validation approach is described by Jermyn et al. [5] where a retrieval system that is regarded to consist of different semantic levels is analyzed.

In this paper, the overall validation of a knowledge-driven image information mining system is addressed. Therefore, the evaluation principles are divided in two main categories. The first one can be seen as objective and involves

technical considerations in extracting the information content, the information quality, and system complexity. The second category is based on subjective concepts and includes relevance feedback, user satisfaction, and semantic and semiotic confusions. In [6], we describe an algorithmic evaluation protocol that measures the (objective) quality of semantic cover-types, identifies the user's (subjective) degree of satisfaction, and compares the two reports. In this paper, however, we restrict ourselves to the analysis of the information flow between different levels of our system, namely "image space," "content-index space," and "semantic space." Additionally, we include the user as an inherent part of the retrieval loop and analyze the communication between user and system.

The paper is organized as follows. In Section 2 we present the concept that is behind our knowledge-driven image information mining (KIM)<sup>1</sup> system of hierarchical Bayesian representation of image content. Additionally, the mining system is briefly illuminated from the communication channel view. Section 3 introduces information-theoretic quantities and in Section 4 they are applied to determine the information between different levels of semantic abstraction in the system. Before we conclude the paper in Section 6, we will analyze the information content of human-computer interactions as presented in Section 5.

## 2. KNOWLEDGE-DRIVEN IMAGE INFORMATION MINING

In order to build a system that is free of the application specificity and to enable its open use in almost any scenario, we start from an application-free hierarchical modeling of the image content as demonstrated in Figure 1. The applied concept of information representation on hierarchical levels of different semantic abstraction is based on a 5-level Bayesian learning model [2, 7].

To extract primitive image features (level 1) from the data  $D$  (level 0), we apply various stochastic signal models  $M$ . These models are given as parametric data models  $p(D|\theta, M)$  and assign the probability to a given realization of the image data  $D$  for a particular value of the parameter vector  $\theta$ . The process of information extraction is realized by the maximum a-posteriori estimate of the parameter vector.

The Bayesian formalism is applied to find the most evident model given some data, in contrast to the model that best describes the data, which will always be the most complex one. Therefore, we introduce the next level of semantics and call these features of the features metafeatures (level 2).

Based on the features on level 1 and the metafeatures on level 2, we derive a set of signal classes  $\omega_i$  describing char-

acteristic groups of points in the parameter spaces of the different models  $M$ . This vocabulary of characteristic signal classes should be valid across all images (to avoid the time-consuming calculation of similarity functions) and should reject existing structures in the different feature spaces of the data. The individual elements of this vocabulary constitute the classes  $\omega_i$  at the image classification (level 3). They are obtained by unsupervised clustering of the entire set of data points (features and metafeatures) of the complete image archive. We perform the unsupervised clustering using a dyadic  $k$ -means algorithm with a predefined number of clusters depending on the size of the archive. The approach is very similar to vector quantization and proved to be sufficient in practical experiments.

Levels 1 to 3 of the hierarchical modeling describe the image data  $D$  at level 0 in a completely unsupervised and application-free way. Based on this objective characterization, the users interests  $L_v$  (level 4), that is, semantic interpretations of the image content, are linked to the signal classes  $\omega_i$  by probabilities  $p(\omega_i|L_v)$ . The information at level 4 can be interactively defined by users with a learning paradigm [2].

After presenting the basic concept of the information mining system of hierarchical image content modeling, we point out the system characteristics from the communication channel point of view. Therefore, the image data at the lowest level in the hierarchical abstraction of image content is considered as a message transmitted by an imperfect communication channel as 2D signals to users. The difficulty in understanding the received image information in form of symbols and semantics in a certain semiotic context, and inferring about the original image causes the problem of unsupervised image content modeling. Through the hierarchical image content characterization, the image retrieval system can be viewed as a composed communication channel. The imperfect nature of the system in combination with the well-known statement of information theory, which says that data processing cannot increase information, entails that each level in the hierarchical scheme is associated with a certain loss of information. The accuracy of communication, for example, accessing a target image or a category of images as exploration results, depends on the assumed levels of image modeling.

In this paper, we deal with the measuring of information (association) between different system levels using information-theoretic quantities. Before we calculate these system performance measurements, we will specify the applied measures in the following section.

## 3. MEASURES OF INFORMATION

In this section, we will shortly describe the concept of information and give some properties that are essential for its use. The best known information measures are Shannon's entropy, Kullback-Leibler divergence, and Fisher's information matrix. Here, mainly the concept of accuracy in order to determine the information contained in the observation about a certain parameter will be considered.

<sup>1</sup>KIM was successfully implemented and tested under ESA contract. It aimed at the generation of a prototype system to rapidly support users in gathering relevant information from remote sensing image archives. The tool should further manage and add value to the huge amounts of historically and newly acquired satellite datasets. KIM can be accessed by everyone via internet at <http://www.acsys.it:8080/kim>.

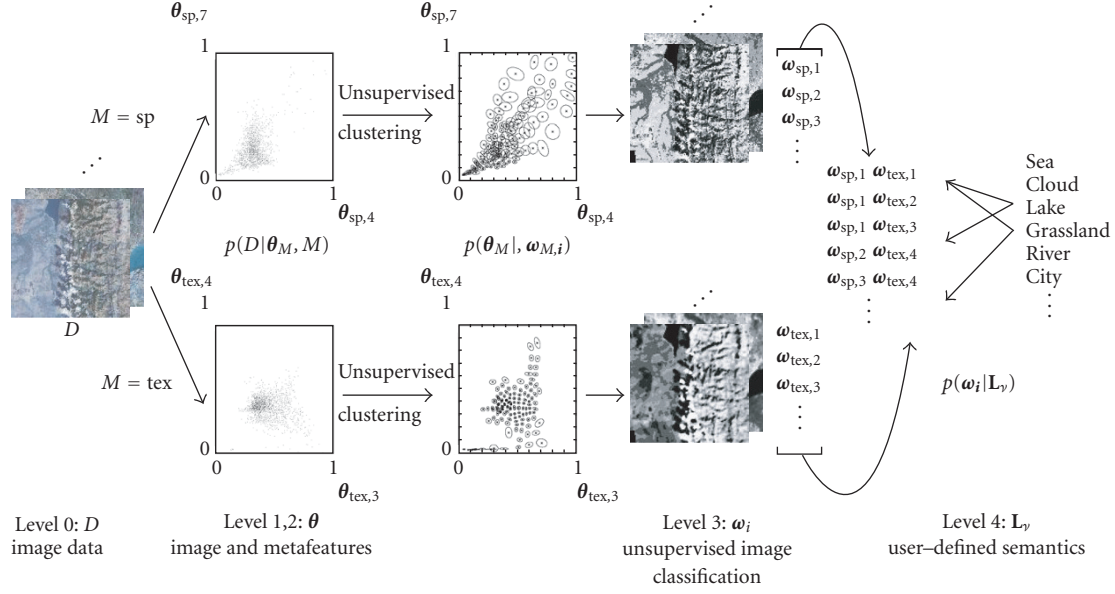


FIGURE 1: Hierarchical modeling of image content. First, primitive features  $\theta$  (level 1) and metafeatures (level 2) are extracted from image data  $D$  (level 0) based on different parametric signal models  $M$ , for example, spectral ( $sp$ ) and texture ( $tex$ ) at a certain scale. Through an unsupervised clustering across all images in the archive, we derive a vocabulary of significant signal classes  $\omega_i$  as content-index. With simple Bayesian networks we link user-defined cover-type labels  $L_\gamma$  (level 4) to the content-index by probabilities  $p(\omega_i | L_\gamma)$ .

### 3.1. Shannon's measure of information

Shannon and Wiener suggested the following measure of entropy [8] contained in a random variable  $X$  with its probability mass function  $p(x)$  as

$$H(X) = - \sum_x p(x) \log p(x). \quad (1)$$

The entropy  $H(X)$  is the uncertainty of a single random variable. Further, conditional entropy can be defined, that is the entropy of a random variable given another random variable. This reduction in uncertainty due to another random variable is called mutual information. For two random variables  $X$  and  $Y$  this reduction is

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \end{aligned} \quad (2)$$

The mutual information  $I(\cdot)$  indicates how much information the random variable  $Y$  has about another random variable  $X$ . If the observation of  $X$  is done under very noisy conditions, for instance, the variables  $X$  and  $Y$  will be statistically independent. No "information" about  $X$  will be recognized in  $Y$  and  $I(X; Y) = 0$ .

Until now we have considered the correlation between two random variables  $X$  and  $Y$  modeled by the causality

$$X \longrightarrow Y. \quad (3)$$

The causality between  $X$  and  $Y$  is also said to form a Markov chain in the order  $X \rightarrow Y$  if the conditional distribution of  $Y$

only depends on  $X$ . With the more general causality

$$U \longrightarrow X \longrightarrow Y \longrightarrow V, \quad (4)$$

described by the joint probability distribution

$$p(u, x, y, v) = p(v|y)p(y|x)p(x|u)p(u), \quad (5)$$

the following inequality (data processing theorem) can be obtained:

$$I(U; V) \leq I(X; Y). \quad (6)$$

This equation states that data processing never increases information, or, no clever manipulation of the data can improve the inferences that can be made from the data. To make (6) more evident, we can interpret the model as the causality between different levels of semantic abstraction in the hierarchical image content representation implemented in the mining system: image data ( $U$ ), extracted primitive image attributes ( $X$ ), content-index ( $Y$ ), and finally the semantic interpretation of the image content by the user ( $V$ ).

### 3.2. Kullback-Leibler divergence

Kullback-Leibler divergence is a generalization of Shannon's measure of information. The divergence is a function of two probability mass functions  $p(x)$  and  $q(x)$  that potentially characterize the same random variable  $X$ :

$$D(p; q) = \sum_x p(x) \log \frac{p(x)}{q(x)}. \quad (7)$$

We want to mention that the divergence is nothing else but the Shannon measure of uncertainty for a random variable  $X$  if  $q(x)$  is a uniform probability mass function. Thus, Shannon's entropy can be interpreted as the amount of information in a model  $p(x)$  of  $X$  compared to the maximum uncertainty model—the uniform distribution. The uniform distribution is the one with maximum entropy.

#### 4. INFORMATION BETWEEN DIFFERENT SYSTEM LEVELS

In the last section, we summarized basic measures used in information theory and statistics. Now, we will apply them to determine the information flow in our knowledge-driven mining system. Therefore, we define three basic levels of different semantic abstraction:

- (i) image space  $\mathbf{I}$ ,
- (ii) content-index space or class space  $\omega$ ,
- (iii) semantic label space  $\mathbf{L}$ .

##### 4.1. Image space versus class space

First, we will evaluate the correlation between image space  $\mathbf{I}$  and class space  $\omega$ . For this verification, we prepared two datasets each consisting of 438 geocoded and coregistered Landsat TM and ERS1 images. All images cover Mozambique and have a size of  $2000 \times 2000$  pixels. The complexity of Landsat TM images is rather high, both from the point of view of image content and subjective understanding by users. The images indicate a huge diversity of spectral signatures and a very broad variety of structural information at different scales. Contrary to optical image data, the information content of ERS1 SAR images of Mozambique is quite small. Due to the SAR sensitivity to the surface geometry, however, large scale structures like rivers and geomorphology are well visible in the images.

From optical Landsat TM images we use spectral properties and textural properties at different scales. The latter are related to the auto-binomial model of the Gibbs random field (GRF) family [9]. On ERS1 radar data we apply a filter and obtain the model-based despeckled (MBD) intensity image [10]. Additionally, structural details are extracted based on a Gauss-Markov random field (GMRF) prior texture model.

From the generated content-index, the mutual information between image space  $\mathbf{I}$  and class space  $\omega$  can be computed as

$$I(\mathbf{I}; \omega) = \sum_{\zeta, i} p(\omega_i | \mathbf{I}_\zeta) p(\mathbf{I}_\zeta) \log \frac{p(\omega_i | \mathbf{I}_\zeta)}{p(\omega_i)}, \quad (8)$$

where  $p(\omega_i | \mathbf{I}_\zeta)$  indicates the posterior probabilities of signal classes  $\omega_i$  given a certain image  $\mathbf{I}_\zeta$  from the archive. Prior probabilities for signal classes and images are given by  $p(\omega_i)$  and  $p(\mathbf{I}_\zeta)$ , respectively.

In Table 1, we summarize the calculations between image and class space. The measures indicate the information transmitted from image data through feature extraction and unsupervised content-index generation (clustering) to the class

TABLE 1: Mutual information  $I(\mathbf{I}; \omega)$  between image space and class (content-index) space. The class space  $\omega$  was separately generated for 438 coregistered Landsat TM and ERS1 images based on different signal models.

	Sensor	Signal models	Scale	$I(\mathbf{I}; \omega)$
A	Landsat TM	Spectral	30 m	1.41
B	Landsat TM	GRF	30 m	0.92
C	Landsat TM	GRF	60 m	1.23
D	Landsat TM	GRF	120 m	1.39
E	ERS1	MBD	60 m	0.53
F	ERS1	MBD	120 m	0.59
G	ERS1	GMRF	60 m	0.43
H	ERS1	GMRF	120 m	0.56

space. Note that for radar data, the computed mutual information  $I(\mathbf{I}; \omega)$  is much lower than for Landsat TM reflecting the low information content of ERS1 scenes. For Landsat TM images, the minimum information is given for texture at lowest scale.

##### Measuring image database complexity

The association between image space and class space can further be used to measure the complexity of images in the archive. Since the query performance of content-based image retrieval systems depends on the complexity of the data, analyzing the image database that is used for testing is rather important for evaluation. Similar to the method of [11] that applies image database statistics and information theory to determine the complexity of image databases, we measure the information between image space and class space based on Kullback-Leibler divergence.

In comparison to mutual information, (8), Kullback-Leibler divergence can be applied to determine the complexity of a single image in the entire archive. Thus, we define the prior probability  $p_i$  as the probability  $p(\omega_i)$  of a particular class  $\omega_i$  in the global (across image) class space and the posterior probability  $q_i$  as the probability  $p(\omega_i | \mathbf{I}_\zeta)$  of a class  $\omega_i$  given a particular image  $\mathbf{I}_\zeta$ . For these two quantities, Kullback-Leibler divergence is given according to (7) as

$$D(p_i; q_i) = \sum_i p(\omega_i | \mathbf{I}_\zeta) \log \frac{p(\omega_i | \mathbf{I}_\zeta)}{p(\omega_i)}, \quad (9)$$

and can be interpreted as how much a single image  $\mathbf{I}_\zeta$  is a typical mixture of the complete image content of the whole dataset or if the image is rather simple. Of course, the image complexity expressed by (9) highly depends on the ability of the applied signal model to describe the image content and to capture characteristic image structures. We show the image complexity on five images as given in Figure 2. These images belong to the Landsat TM and ERS1 archives that are composed of 438 scenes, respectively. For each of the depicted images, we calculated the relative entropy  $D(\cdot)$  for the applied signal models and compared the results (Figure 3). Although there are differences between the signal models, a correlation along the signal models is visible. An interesting fact is that the GRF texture model at a scale of 30 m (original resolution) delivers the smallest entropies.



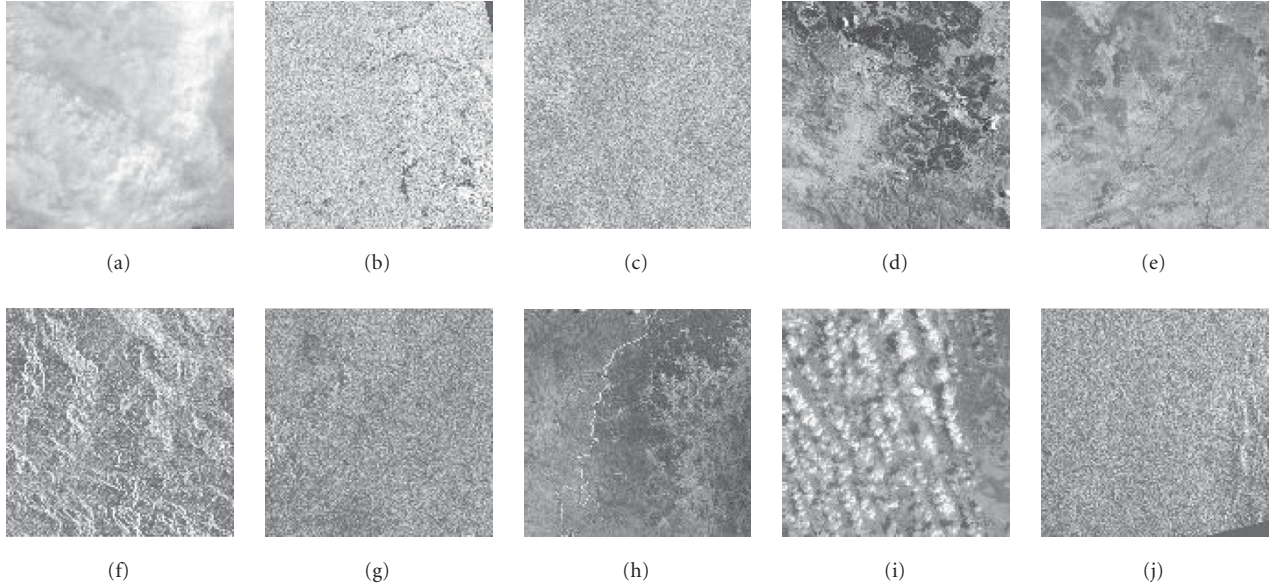


FIGURE 2: Example of multimission dataset containing five coregistered Landsat TM (upper plots) and ERS1 (lower plots) images. For each image we computed the complexity in the archive using different signal models as outlined in Figure 3. (a) Lan096\_cut19, (b) lan107\_cut1, (c) lan104\_cut6, (d) lan100\_cut55, (e) lan101\_cut81, (f) ers096\_cut19, (g) ers107\_cut1, (h) ers104\_cut6, (i) ers100\_cut55, and (j) ers101\_cut81.

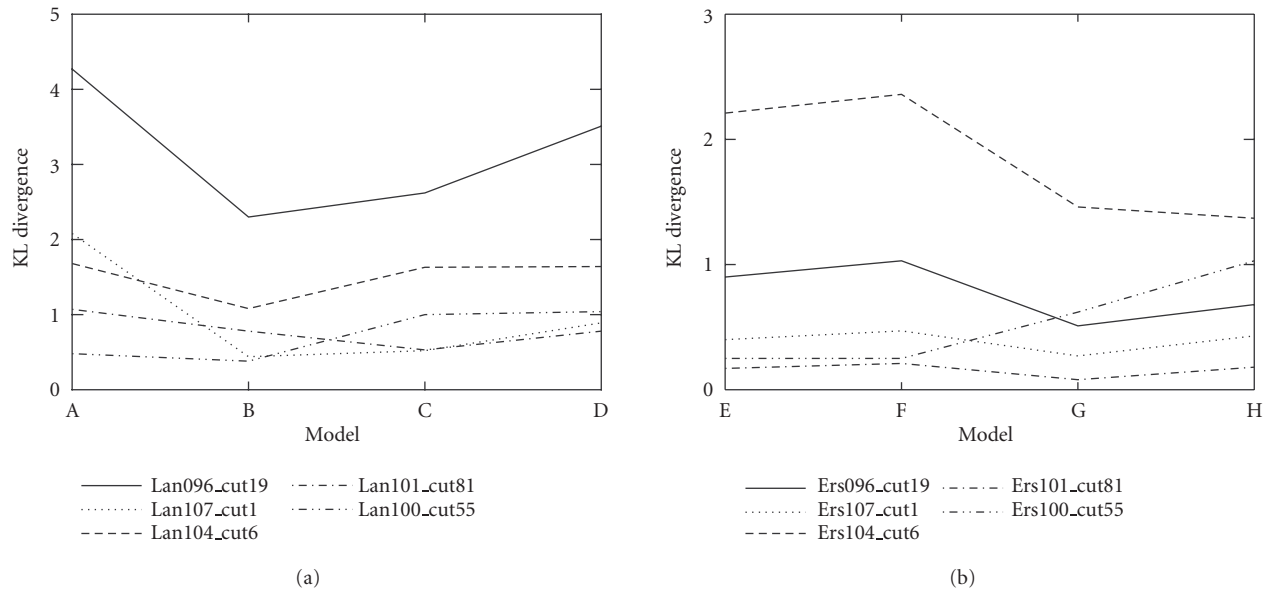


FIGURE 3: Information-theoretic complexity of (a) Landsat TM and (b) ERS1 images as depicted in Figure 2. To describe the content of both datasets, the models from Table 1 are applied. The complexity measures show that “lan096\_cut19” is a very simple image in the Landsat TM archive whereas “lan107\_cut1” and “lan101\_cut55” are quite complex. For ERS1 data, “ers104\_cut6” is rather simple and “ers101\_cut55” is the image with the highest complexity.

It can be seen as the model that captures most of the significant image structures in the archive.

#### 4.2. Content-index space versus semantic label space

In the same way as we calculated the mutual information between image space and class space, we can compute the mutual information between the next levels in the hierarchical

image content representation: class space  $\omega$  and semantic label space  $L$ . The first three levels in the hierarchy are obtained in a complete unsupervised and application-free way. Consequently, the information between image and class space can be seen as a complete objective measure. Subjective user-related concepts neither have an influence on  $I(\cdot)$ , (8), nor on  $D(\cdot)$ , (9).

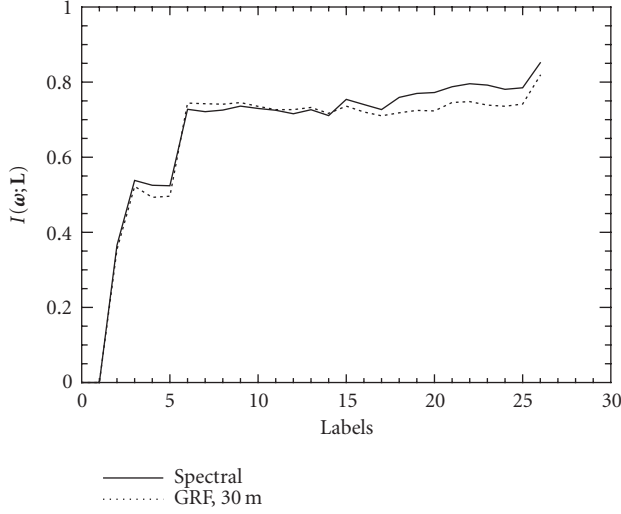


FIGURE 4: Mutual information  $I(\omega; \mathbf{L})$  for a sequence of 26 semantic cover-type labels.  $I(\omega; \mathbf{L})$  was separately computed for each signal model, this time for spectral and GRF, scale 30 m. The increase of the mutual information depends on the diversity of the defined labels.

Since a user-defined semantic cover-type label  $\mathbf{L}_v$  is the result of several human-machine interactions, the information between  $\omega$  and  $\mathbf{L}$  can be seen as subjective and objective. More precisely, the stochastic link  $p(\omega_i | \mathbf{L}_v)$  derived from the user's feedback connects objective signal classes  $\omega_i$  to the user-specific interpretation of the image content in form of semantic cover-types  $\mathbf{L}_v$ . Therefore, the set of probabilities  $p(\omega_i | \mathbf{L}_v)$  is the central element of this analysis. With  $p(\omega_i | \mathbf{L}_v)$  as the likelihoods and the priors  $p(\omega_i)$  and  $p(\mathbf{L}_v)$ , the mutual information between signal class space and semantic label space can be computed as

$$I(\omega; \mathbf{L}) = \sum_{i,v} p(\omega_i | \mathbf{L}_v) p(\mathbf{L}_v) \log \frac{p(\omega_i | \mathbf{L}_v)}{p(\omega_i)}. \quad (10)$$

Note that  $I(\omega; \mathbf{L})$  is separately computed for each signal model that the user selected to learn a cover-type of his interest. We can interpret this measure as the quality of semantic cover-types  $\mathbf{L}_v$  to capture the entire diversity of structures and patterns represented by the content-index. In Figure 4, the behaviour of the mutual information is shown for a sequence of semantic labels defined by various users. If semantic labels are inserted that differ from the existing ones in terms of association to signal classes  $\omega_i$  of the different feature models, mutual information increases. Contrary, the mutual information decreases if cover-type labels are similar to the existing ones in the DBMS inventory. We can interpret an increase of  $I(\cdot)$  as the ingestion of a new semantic label in the database system that differs from the existing cover-types in terms of signal classes  $\omega_i$ . Contrary, the mutual information decreases if cover-type labels are similar to existing semantics.

The mutual information between class space and semantic space can further be applied to determine a dynamic index

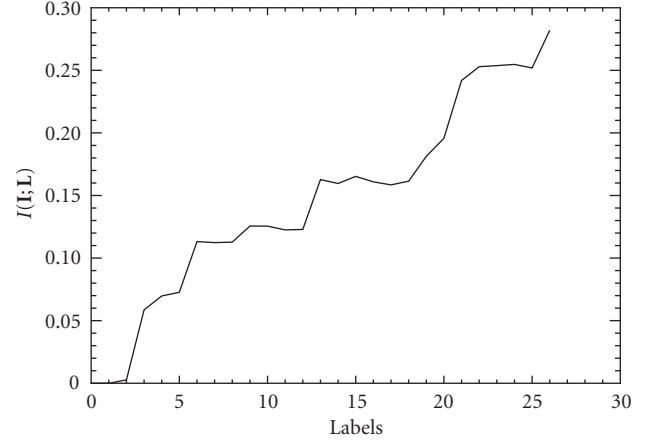


FIGURE 5: Mutual information between image space  $\mathbf{I}$  and semantic label space  $\mathbf{L}$ . If new cover-types that are dissimilar from the existing ones are trained and ingested in the DBMS inventory, the mutual information  $I(\mathbf{I}; \mathbf{L})$  increases.

control function. This function enables to filter out labels at real time that are overlapping with other cover-types or to indicate to the user that similar labels have been already defined (see Section 5).

#### 4.3. Image space versus semantic label space

After computing the mutual information between image-class space and class-semantic space, we can directly make the connection between cover-type labels  $\mathbf{L}_v$  and images  $\mathbf{I}_\zeta$ . Therefore, we start with Bayes' formula

$$p(\mathbf{L}_v | \omega_i) = \frac{p(\omega_i | \mathbf{L}_v) p(\mathbf{L}_v)}{\sum_v p(\omega_i | \mathbf{L}_v) p(\mathbf{L}_v)} \quad (11)$$

to obtain the posterior probabilities  $p(\mathbf{L}_v | \omega_i)$  from the likelihoods  $p(\omega_i | \mathbf{L}_v)$  and the priors  $p(\mathbf{L}_v)$ . Having defined the posterior probabilities  $p(\mathbf{L}_v | \omega_i)$ , we can infer the probability of a semantic label  $\mathbf{L}_v$  given a certain image  $\mathbf{I}_\zeta$  as

$$p(\mathbf{L}_v | \mathbf{I}_\zeta) = \sum_i p(\mathbf{L}_v | \omega_i) p(\omega_i | \mathbf{I}_\zeta). \quad (12)$$

These probabilities assign each semantic label  $\mathbf{L}_v$  to the images in the archive. The images with the highest probabilities  $p(\mathbf{L}_v | \mathbf{I}_\zeta)$  are retrieved by the system and displayed to the system operator.

In a similar way as the mutual information between  $\omega_i$  and  $\mathbf{L}_v$  is calculated, we obtain this information by

$$I(\mathbf{I}; \mathbf{L}) = \sum_{\zeta,v} p(\mathbf{L}_v | \mathbf{I}_\zeta) p(\mathbf{I}_\zeta) \log \frac{p(\mathbf{L}_v | \mathbf{I}_\zeta)}{p(\mathbf{L}_v)} \quad (13)$$

with posterior probabilities  $p(\mathbf{L}_v | \mathbf{I}_\zeta)$  from (12) and priors  $p(\mathbf{I}_\zeta)$  and  $p(\mathbf{L}_v)$ . In Figure 5, we display the computational results based on the same semantic cover-types that are used to analyze the association between class space and semantic

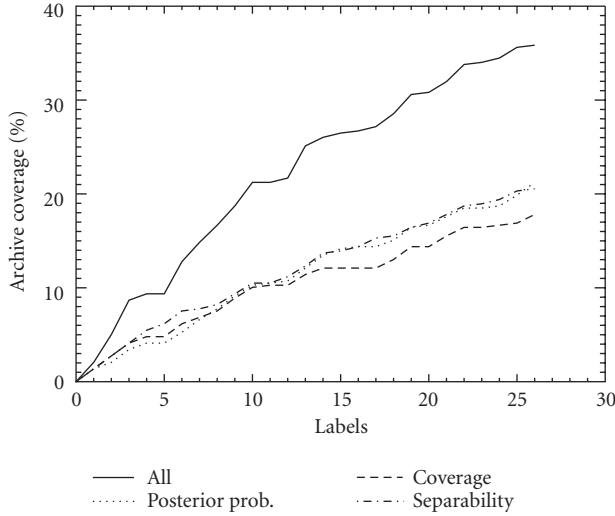


FIGURE 6: Exploration image archive for 26 defined semantic labels. The plot shows the archive coverage using the first 6 top-ranked images in the search results. With all 26 labels, about 35% of the entire archive is assigned to the semantic content.

space. Whereas  $I(\omega; \mathbf{L})$  indicates how much information the clusters contain about semantic labels,  $I(\mathbf{I}; \mathbf{L})$  directly shows the association between images and semantic labels (the information flow from the archive to users). Consequently, we can—at least qualitatively—infer the amount of images in the archive that are connected to cover-types. And if a new semantic label is inserted in the DBMS, its novelty can be assessed to the existing ones.

Notice the similarity between Figures 6 and 5. While the first figure shows the coverage of the database according to retrieved top-ranked images based on different search criteria [2], the second shows the information-theoretic association between image and semantic label space.

## 5. EVALUATION OF HUMAN-MACHINE INTERACTIONS

The development of intelligent human-machine interfaces for information mining applications is a difficult task since no well-established guidelines and models of the functions that such systems should have are available [12]. Despite this difficulty, we designed and implemented graphical, intuitive, and powerful visual interfaces that aim at controlling an underlying mining system, directly interacting with the user and enabling him to retrieve relevant images without the support of a human intermediary.

The experiments reported in this section focus on the evaluation of the performance of human-machine interactions. First, the user's target structures are classified and identified by tracing the man-machine interactions. Then, the convergence of the learning process is analyzed using information-theoretic measurements, and finally the training feedback is applied to predict which cover-types the user might be interested in.

For the evaluation task, we ingested several datasets in the KIM system that had been used for studying a flooding disaster in Mozambique. Typical tasks for image analysts were to identify flooded land and road networks, to look for airstrips to provide humanitarian air-borne supply, and to detect areas that can be used for building refugee camps. The detailed study of all scenes and the identification of relevant objects with “classical” tools for interpreting remote sensing image data required several months of work. With KIM, we provided the involved image analysts an easy-to-learn tool to efficiently explore large archives and to discriminate and analyze the relevant image content. After a short introduction in the KIM system and its functions, the evaluators could define semantic cover-type labels, query the archive for relevant data, and interpret them.

### 5.1. Target structure classification and identification

In order to follow a system operator's training iterations, we use the information stored in the user-log Table 2. With this information, the following details about target structures from the traced human-computer interactions are extracted: the time of performed training samples, the action type, and the location of samples in the image. In order to exemplify the classification of target structures, we superimpose the training samples on the images with indication of training iteration and the kind of training (positive or negative). We show the man-machine interactions for training using five Landsat images in Figure 7.

The user both performed 15 samples on these five images to learn the semantic cover-type “road” and commented (off-line) on each iteration. With the positive examples, the operator tried to include many label relevant (linear) structures. Negative ones are supposed to cover a high diversity of irrelevant objects and structures reflected by different feature models Table 3.

### 5.2. Convergence of learning process

How efficient is the mining system learning convergence? To answer this question, we analyze the human-computer interactions and measure the learning progress using Kullback-Leibler divergence. In [13], the divergence between two complete sets of probabilities  $\mathcal{L}_\gamma = \{p(\omega_1|\mathbf{L}_\gamma), \dots, p(\omega_r|\mathbf{L}_\gamma)\}$  and  $\neg\mathcal{L}_\gamma = \{p(\omega_1|\neg\mathbf{L}_\gamma), \dots, p(\omega_r|\neg\mathbf{L}_\gamma)\}$  is defined as

$$D(\mathcal{L}_\gamma, \neg\mathcal{L}_\gamma) = \sum_{i=1}^r [p(\omega_i|\mathbf{L}_\gamma) - p(\omega_i|\neg\mathbf{L}_\gamma)] \ln \frac{p(\omega_i|\mathbf{L}_\gamma)}{p(\omega_i|\neg\mathbf{L}_\gamma)}, \quad (14)$$

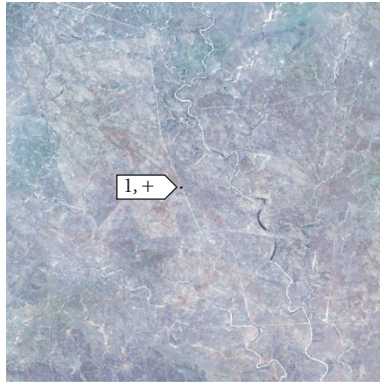
which can be seen as the distance between the two probability distributions  $\mathcal{L}_\gamma$  and  $\neg\mathcal{L}_\gamma$ . The divergence  $D(\mathcal{L}_\gamma, \neg\mathcal{L}_\gamma)$  can be calculated either for a combination of signal models or separately for each signal class.

Each positive and negative training iteration (mouse click) implicitly causes the update of the probabilistic link  $p(\omega_i|\mathbf{L}_\gamma)$  between content-index and semantic cover-type label. The amount of information in each learning iteration is reflected by the increase or decrease of the divergence for each signal model. The questions that arise are (1) how the

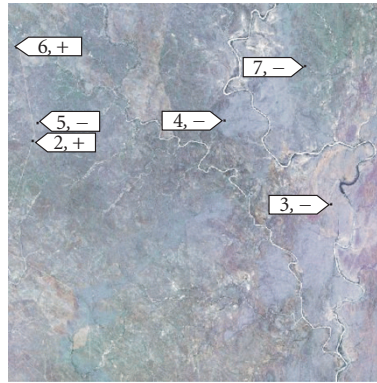


TABLE 2: User-log information stored in the image information mining DBMS. After selecting a certain combination of feature models and an image from the initial gallery, the user started to analyze the image data and to give positive and negative training samples (left and right mouse clicks). Note that each performed action is assigned to a certain action type and time.

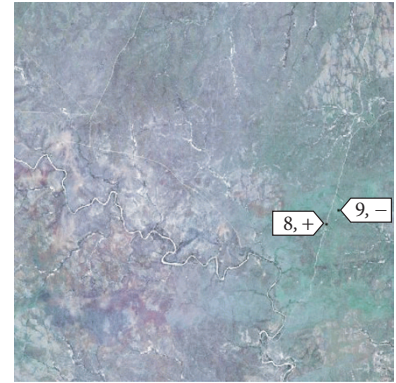
Time	Action type	Parameters
11:30:06	Model	1, 2, Switzerland, spectral, 256, 256, 151
11:30:06	Model	2, 2, Switzerland, texture, 256, 256, 51
11:30:06	Image	0, tmgeo_402_4_6, new label
11:30:38	Applet loaded	—
11:30:48	Click left	Zoom Panel, [x = 195, y = 85]
11:30:52	Click left	Zoom Panel, [x = 179, y = 30]
11:30:54	Click right	Zoom Panel, [x = 172, y = 37]
11:30:59	Click right	Zoom Panel, [x = 106, y = 13]
11:31:02	Click right	Zoom Panel, [x = 95, y = 104]
11:31:07	Click left	Zoom Panel, [x = 196, y = 85]
11:31:10	Search	—
11:31:28	Order	Separability_value ASC, lowest
11:31:28	Image	1, tmgeo_402_3_5, swiss river
11:31:35	Tab change	Learn
11:31:41	Applet loaded	—
11:31:47	Click left	Zoom panel, [x = 87, y = 177]
11:31:52	Click left	Zoom panel, [x = 87, y = 158]
...	...	...



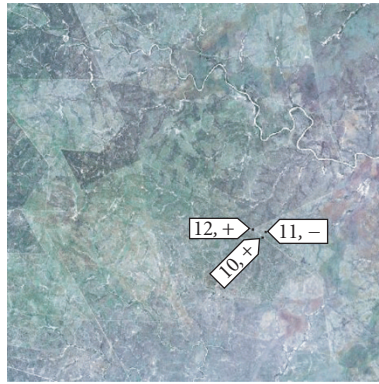
(a)



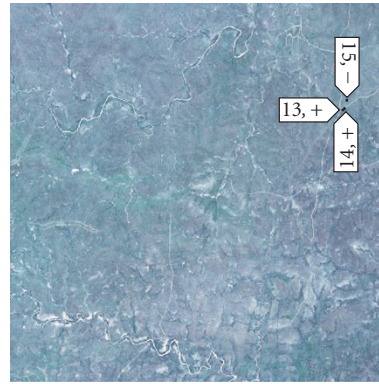
(b)



(c)



(d)



(e)

FIGURE 7: Definition of a semantic label “road” based on spectral and textural feature models (Landsat TM) with 15 training iterations. Altogether, the user performed the training samples (8 positive and 7 negative) with 5 different images. The commented reasons for individual training samples are summarized in Table 3.



TABLE 3: Information about the training samples shown in Figure 7. The user’s aim was to include many similar structures in his positive training whereas the negative training shows a high range of contrary patterns.

Iteration	Training (+/−)	User’s comment
1	+	Starting sample, road in flat terrain
Search archive, selected image from coverage retrieval set		
2	+	Road in another image
3	−	Negative training sample for smooth terrain
4	−	Negative training sample, exclude flat terrain
5	−	Negative training sample close to a road
6	+	Include road
7	−	Exclude rough terrain
Search archive, selected image from separability retrieval set		
8	+	Include highway road on a new image
9	−	Specify road by clicking next to road
Search archive, selected image from coverage retrieval set		
10	+	Include strong road
11	−	Exclude “city”
12	+	Include road crossing
Search archive, selected image from separability retrieval set		
13	+	Include road on another image
14	+	Include road
15	−	Exclude road

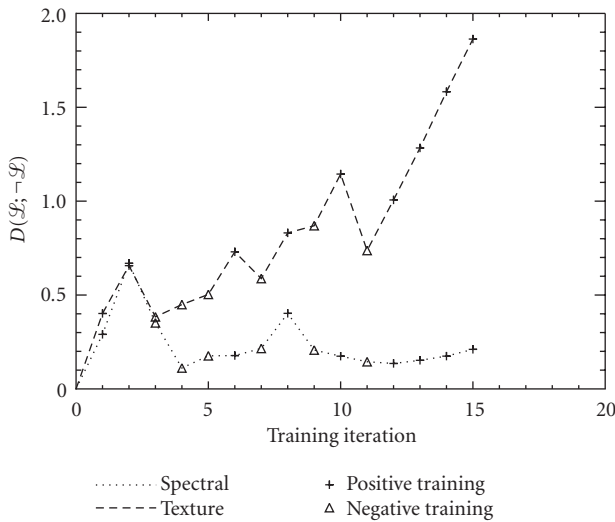


FIGURE 8: Kullback-Leibler divergence for interactive learning. The two graphs, each one for a certain feature model, represent the increase or decrease of divergence after each user-performed training sample.

divergence for the different feature models behaves and (2) whether there is convergence.

In Figure 8, we analyze the Kullback-Leibler divergence for the traced interactions as displayed in Figure 7. After the first two positive training samples, the divergence for both feature models almost equally increases. The third and all following interactions show that the defined semantic label is well represented by texture, the spectral model is of minor importance. Although the last three positive training iterations lead to a linear increase of the divergence for texture, the sum of both models as the average quality of the training

shows convergence. This example shows the real behaviour of the interaction between human and system: with only a few training samples, an operator can define a cover-type label and query the database for relevant images. Of course, the number of iterations necessary to train a label depends on the selected feature models and the complexity of the label.

### 5.3. Matching user-specific semantic labels

The example in Figure 8 demonstrates that a user is able to define a particular semantic cover-type with just a few training iterations. Now, we want to predict the user’s intentions. By monitoring the interactions with the database system, it might be possible to somehow predict which semantic label or category of labels the user is interested in. This kind of forecast is called “matching of user interests” in literature and tries to identify the user’s target by analyzing his actions [14]. In a similar way as we computed the Kullback-Leibler divergence  $D(\mathcal{L}_v, \neg \mathcal{L}_v)$  between the two probability distributions  $\mathcal{L}_v$  (positive training) and  $\neg \mathcal{L}_v$  (negative training) in (14), we can extend this formula and determine the “similarity” between a certain label and other labels in the inventory. Denoting by  $\mathbf{L}_v$  the label a user is training, the similarity to any other label  $\mathbf{L}_\mu$  can be assessed as

$$D(\mathcal{L}_v, \mathcal{L}_\mu) = \sum_i p(\omega_i | \mathbf{L}_v) \ln \frac{p(\omega_i | \mathbf{L}_v)}{p(\omega_i | \mathbf{L}_\mu)}. \quad (15)$$

Labels that are “close”<sup>2</sup> to  $\mathbf{L}_v$  are characterized by a very low divergence  $D(\cdot)$  and labels dissimilar to  $\mathbf{L}_v$  show a high divergence. In Figure 9, we depict the similarity of a particular label to others during the interactive learning process.

<sup>2</sup>“Close” in the semantic sense.

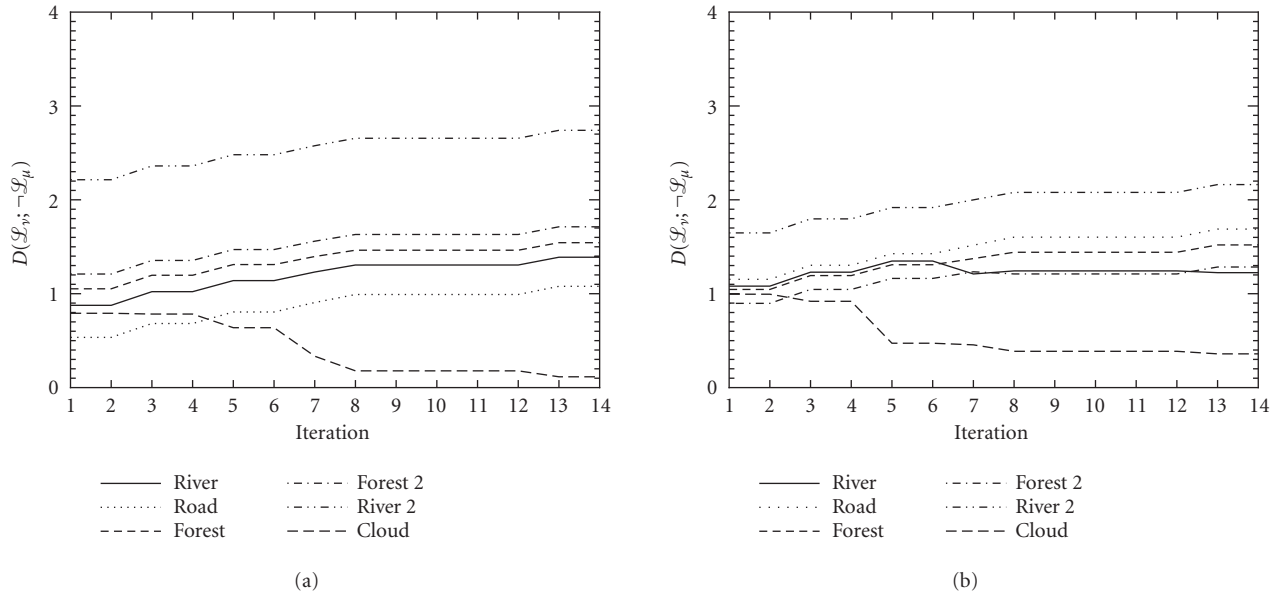


FIGURE 9: Results of matching a trained cover-type label “cloud” using (a) spectral and (b) textural feature model.

After just a few feedback samples, one semantic label close to “cloud” is visible for both the spectral and texture feature model. Of course, the performance of this method depends on the applied signal models for interactive learning and the capability of the user to learn the system.

## 6. CONCLUSIONS

In this paper we demonstrated the evaluation of a probabilistic knowledge-driven image information mining system using information-theoretic measures. We started the presentation with the basic system concept of hierarchical image content modeling and briefly illuminated the system from the communication channel view. Having explained information measures settled in information theory, namely Shannon’s entropy, mutual information, and Kullback-Leibler divergence, we applied them to evaluate the mining system. Therefore, the system was separated in three main levels: image space, content-index space, and semantic label space. We computed the information between image space and class space, class space and semantic space, and finally between image space and semantic space as the direct connection between image archive data and users. Mutual information between image and class space can be viewed as a pure objective measure and indicates the ability of the system to represent the image content. The other measures are related to subjective user-defined labels and enable an analysis of the trained cover-types in the system. To explore the interactions between human and mining system, an operator was traced during several learning sessions. We classified and analyzed the target structures that are given by positive and negative training samples, analyzed the learning progress, and tried to match user-specific semantic labels.

In future, we will further develop the system and its graphical user interface. In order to optimally adjust the user interface on the user’s needs, we will deal with the registration of the user’s behaviour and preferences, typical queries, and actions. The implementation and evaluation of personalizer modules for user profiling, content targeting, and user tracking functionality will be improved to deliver the target specific content to a particular group of users. Additionally, different viewing techniques, for example, 2D/3D visualization, will support the user in analyzing different data sources and will give him an immediate feedback.

## REFERENCES

- [1] Y. Rui, T. S. Huang, and S. Chang, “Image retrieval: past, present, and future,” in *Proc. International Symposium on Multimedia Information Processing (ISMIP ’97)*, Taipei, Taiwan, December 1997.
- [2] M. Datcu, H. Daschiel, A. Pelizzari, et al., “Information mining in remote sensing image archives: system concepts,” *IEEE Trans. Geosci. Remote Sensing*, vol. 41, no. 12, pp. 2923–2936, 2003.
- [3] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, MIT Press, New York, NY, USA, 1996.
- [4] H. Müller, W. Müller, D. McG. Squire, S. Marchand-Maillet, and T. Pun, “Performance evaluation in content-based image retrieval: overview and proposals,” *Pattern Recognition Letters*, vol. 22, no. 5, pp. 593–601, 2001.
- [5] I. H. Jermyn, C. W. Shaffrey, and N. G. Kingsbury, “Evaluation methodologies for image retrieval systems,” in *Proc. Advanced Concepts for Intelligent Visual Systems (ACIVS ’02)*, Ghent, Belgium, September 2002.
- [6] H. Daschiel and M. Datcu, “Information mining in remote sensing image archives: system evaluation,” *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 1, pp. 188–199, 2005.

- [7] M. Schröder, H. Rehrauer, K. Seidel, and M. Datcu, "Interactive learning and probabilistic retrieval in remote sensing image archives," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 5, pp. 2288–2298, 2000.
- [8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.
- [9] M. Schröder, H. Rehrauer, K. Seidel, and M. Datcu, "Spatial information retrieval from remote-sensing images. II. Gibbs-Markov random fields," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 5, pp. 1446–1455, 1998.
- [10] M. Walessa and M. Datcu, "Model-based despeckling and information extraction from SAR images," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, no. 5, pp. 2258–2269, 2000.
- [11] A. Rao, R. K. Srihari, L. Zhu, and A. Zhang, "A method for measuring the complexity of image databases," *IEEE Trans. Multimedia*, vol. 40, no. 2, pp. 160–173, 2002.
- [12] G. Guida and C. Tassa, *Design and Development of Knowledge-Based Systems—from Life Cycle to Methodology*, Wiley, Chichester, UK, 1994.
- [13] S. Kullback, *Information Theory and Statistics*, Dover Publications, Mineola, NY, USA, 1997.
- [14] W. Koh and L. Mui, "An information-theoretic approach for ontology-based interest matching," in *Proc. Workshop on Ontology Learning (OL '01)*, Seattle, Wash, USA, August 2001.

**Herbert Daschiel** received the Diploma degree in geodetic engineering from the Technical University of Munich, Munich, Germany, in 2001, and the Ph.D. degree in computer engineering from the Technical University of Berlin, Berlin, Germany, in 2004. From 2001 to 2004 he was with the Remote Sensing Technology Institute (IMF) of the German Aerospace Center (DLR), Oberpfaffenhofen, Germany, where he carried out algorithms for image information mining system evaluation and enhancement. Since 2005 he has been with the Bayerische Vermessungsverwaltung, Munich, Germany. His interests are in knowledge-based multimedia information retrieval, image processing, human-machine communication, and OO programming.



**Mihai Datcu** received the M.S. and Ph.D. degrees in electronics and telecommunications from the University "Politehnica" of Bucharest (UPB), Romania, in 1978 and 1986. In 1999 he received the title "Habilitation à 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 at the German Aerospace Center (DLR), Oberpfaffenhofen. He held Invited Professor appointments with the Department of Mathematics, University of Oviedo, Spain, the Université Louis Pasteur, and the International Space University, both in Strasbourg, France, and the University of Siegen, Germany. From 1992 to 2002, he was appointed as an Invited Professor at the Swiss Federal Institute of Technology (ETH), Zürich. Currently he is a Senior Scientist and the Image Analysis Research Group Leader at the Remote Sensing Technology Institute (IMF), DLR, Oberpfaffenhofen, and Professor at École Nationale Supérieure des Télécommunications (ENST) Paris, France. His 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 earth observation data.