BAYESIAN LABELING OF REMOTE SENSING IMAGE CONTENT

M. SCHRÖDER AND K. SEIDEL
Swiss Federal Institute of Technology ETH Zürich,
Communication Technology Lab, Image Science Group
Gloriastr. 35, CH-8092 Zürich

AND

M. DATCU
German Aerospace Center DLR,
German Remote Sensing Data Center DFD, Oberpfaffenhofen
D-82234 Weßling

Abstract. In this paper we present a multi-level scheme for stochastic description of image content. The different levels are derived from the different degrees of abstraction. On the level of the image data, we use stochastic data models and Bayesian parameter estimation to derive low-level image features. On the next level, we derive meta features that provide both the fit of these models and the actual complexity of the data. The low-level and meta features are combined in an unsupervised clustering scheme to obtain an objective description of the image content. To obtain this objective description we use clustering by melting. The descriptions by several models are then linked to application-oriented, semantic labels using another process of Bayesian inference. We sketch in detail the various processes of inference and give an example for this kind of information on each level of abstraction using satellite images (RESURS-01 and X-SAR).

Key words: Bayesian networks, Bayesian statistics, clustering by melting, Gibbs fields, Markov random fields, remote sensing, texture

1. Introduction

Typical remote sensing images exhibit an enormous amount of information. This information is contained in possibly very many spectral bands or different polarizations. Furthermore, radiometric preprocessing and geometric correction may be necessary. Classical methods of image classification [1] mainly work with multispectral models, possible combined with simple textural features, in order to ob-

1 Email: {schroeder}@vision.ee.ethz.ch
2 Email: mihai@datcu@dlr.de
tain a segmentation based on given training data. The classical approach is highly application-oriented and does not provide an overall description of the image content as it is desired in future remote sensing image archives that offer the possibility of query by content [2].

In recent years, content-based query from databases of pictures and images has been subject of intensive research [3]. However, for the query from complex remote sensing image data these results are of very limited use. To close the gap between traditional methods of remote sensing image interpretation and pictorial information system we use a novel information theoretic concept of Bayesian information extraction [4,5]. We have already applied this concept in deriving a new texture descriptor in remote sensing images [6], that we shortly review in this paper. However, we also intend to apply this concept to the characterization of information on higher levels of abstraction.

The paper is organized as follows. In Sec. 2, we define the different levels of abstraction on which we characterize the image content. The probabilistic nature of each level is discussed in detail in the following sections: the Bayesian parameter estimation of low-level image features in Sec. 3, the derivation of meta features in Sec. 4 and the segmentation of the image and the possibility to incorporate geometrical models in Sec. 5. The final level of semantic content labels is reached in Sec. 6 with the description of a user-adaptive labelling scheme. Finally, we conclude with a short summary in Sec. 7.

2. Levels of Abstraction

Information in remote sensing images exhibits an enormous complexity. Our goal is to capture, describe, and assess this information using stochastic models. However, an overall, completely un-biased description of the information in remote sensing images is not possible. Therefore, we use a certain set of stochastic models and organize the obtained information according to the different levels of abstraction. In this framework, a natural grouping yields six levels of abstraction, which we summarize in Tab. 1.

<table>
<thead>
<tr>
<th>level</th>
<th>information</th>
<th>elements</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>semantics</td>
<td>content labels</td>
<td>snow, ice, mountains</td>
</tr>
<tr>
<td>4</td>
<td>geometrical</td>
<td>region characterization</td>
<td>points, lines, area</td>
</tr>
<tr>
<td>3</td>
<td>image segmentation</td>
<td>image segments</td>
<td>class labels</td>
</tr>
<tr>
<td>2</td>
<td>meta features</td>
<td>properties of models used</td>
<td>evidence, complexity</td>
</tr>
<tr>
<td>1</td>
<td>image features</td>
<td>estimations of signal models</td>
<td>parameters, uncertainty</td>
</tr>
<tr>
<td>0</td>
<td>image data</td>
<td>raw data</td>
<td>pixel values</td>
</tr>
</tbody>
</table>

TABLE 1. Levels of abstraction of spatial information. From bottom to top the semantic meaning becomes stronger. Each level is obtained from one or more preceding levels using a process of Bayesian inference.
Starting from the actual image data with bare pixel values (level 0) we first obtain low-level image features such as multi-spectral or textural features (level 1). These features provide additional information such as the evidence or the optimum complexity of a particular model (‘information on information’ or ‘meta information’, level 2). According to these features, an un-supervised segmentation of the image into homogeneous regions is possible (level 3). These regions can be further described using geometrical objects such as points, lines, and areas (level 4). The information on levels 1–4 is—in the frame of the used models—purely objective. On the final level (level 5), semantic labels are assigned. These labels are defined by experts in particular fields of applications and linked to the features at lower levels using a formalism similar to Bayesian networks.

This scheme promises to be a powerful and robust framework for the representation of information on image content on various levels of abstraction. The processes of Bayesian inference from one level to the next are discussed in the following sections.

3. Low-level image features (level 1)

In this section, we demonstrate how to perform stochastic modeling of the observed pixels of an image, that is, how to describe image content on a level very close to the actual image data.

The stochastic model $M$ is defined via the likelihood $p(x_s | M, \Theta)$ of the image pixels $\{x_s\}$ in terms of the parameter vector $\Theta$. In particular, Gibbs Markov random field models are defined via

$$p(x_s | \partial x_s, \Theta) = \frac{1}{Z} e^{-H(x_s; \partial x_s, \Theta)}$$

as the probability distribution function of the pixel $x_s$ given its neighborhood $\partial x_s$.

Here, $H(x_s; \partial x_s, \Theta)$ is the energy function and $Z$ the partition sum of the Gibbs distribution.

The process of information extraction is equivalent with finding the parameter vector $\Theta$ that best explains the given data. The posterior probability of the parameter vector $\Theta$ is obtained using Bayes’ equation

$$p(\Theta | \{x_s\}) = \frac{p(\{x_s\} | \Theta) p(\Theta)}{p(\{x_s\})},$$

with the likelihood $p(\{x_s\} | \Theta)$ of the data and the prior probabilities $p(\Theta)$ and $p(\{x_s\})$. The prior $p(\Theta)$ can be derived from, e.g., group theory or maximum entropy considerations. The maximization of the posterior probability of $\Theta$ can be done in the frame of, e.g., the maximum pseudo-likelihood approximation. For certain models, computational efficient algorithms are available, such as the conditional least squares estimator of the auto-binomial model [6], which provides both the incorporation of the correct prior and a very fast calculation.

In Fig. 1, we give an example for this first level of information extraction using the auto-binomial model. In this model, the energy function is defined as

$$H(x_s; \partial x_s, \Theta) = -\ln \left( \frac{G}{\pi} \right) - x_s \cdot \eta,$$
where we denote the maximum grey value with $G$ and the binomial coefficients with $\binom{m}{n}$. The neighboring pixels $\delta x_j$ have an effective influence via the scalar quantity

$$
\eta = a + \sum_{ij} b_{ij} \frac{x_{ij} + x'^{ij}}{G}.
$$

(4)

Due to the redundancy of the parameter $a$ [6], the auto-binomial model is characterized by the parameter vector

$$
\Theta = [b_{11}, b_{12}, b_{21}, b_{22}, \ldots].
$$

(5)

As the example shows, there is a clear correlation between the visual appearance of the image data and the magnitude of $|\Theta|$. Depending on the choice of the model, different spatial characteristics of the image data can be captured. If a family of different models at possibly many scales of the image is applied, then we obtain a robust description of the actual image data, which incorporates all aspects the models are able to describe.

![Figure 1. Bayesian parameter estimation of a Gibbs field. Remote sensing image of the alpine region (left, RESURS-01 satellite, 1km x 1km per pixel) and resulting parameters $\Theta$ of a 3rd order auto-binomial model (right, ‘black’ to ‘white’ represent low to high values). Here we plot the norm $|\Theta|$ of $\Theta$ as a measure of the strength of the texture. The window size for estimation of the parameters is $32 \times 32$.](image)

4. Meta features as image content information (level 2)

In a second level of Bayesian inference we obtain the model $M_j$ that best describes the image data. We select this model from a family of models $\{M_i\}$ by maximizing the posterior probability

$$
p(M_j | \{x_s\}) = \frac{p(\{x_s\} | M_j) p(M_j)}{p(\{x_s\})},
$$

(6)

with the prior $p(M_i)$ of the model $M_i$ and the evidence

$$
p(\{x_s\} | M_i) = \int p(\{x_s\} | \Theta, M_i) p(\Theta | M_i) d\Theta.
$$

(7)
In Fig. 2, left, we show the evidence of the example in Fig. 1, right. Note that 'strong' textures are accompanied by a low evidence as compared to fairly homogeneous areas. This is in accordance with the fact that homogeneous areas are almost deterministic and can be described by the model with very high probability.

The use of several models with different complexity enables us to estimate the actual complexity of the image data. The decision between model $M_1$ and $M_2$ is done by considering the threshold

$$\frac{p(M_1 | \{x_s\})}{p(M_2 | \{x_s\})} \geq \Lambda.$$  \hspace{1cm} (8)

The example in Fig. 2, right, depicts the result of the model selection between the 3rd and 1st order auto-binomial model with a threshold of $\Lambda = 2.3$. Note that there are regions with a better fit of the more complex model without strong texture being present in the original image (e.g., in the upper right corner). This indicates that the result of the model selection can be used as an additional characterization of image content.

![Example of Bayesian model selection](image)

**Figure 2.** Example of Bayesian model selection. Evidence of the 3rd order auto-binomial model (left, ‘black’ to ‘white’ represent low to high evidence) and result of the model selection between 1st and 3rd order auto-binomial model (right). In the right example, ‘white’ areas correspond to areas with a significantly higher evidence of the 3rd order model (threshold $\Lambda = 2.3$). Note that ‘strong’ texture is accompanied by low evidence as compared to homogeneous areas (left), but decides clearly for the more complex model (right). The original image data is shown in Fig. 1.

5. **Image segmentation and geometrical models (level 3 and 4)**

After obtaining image features on levels 1 and 2 the amount of data to describe the image is still considerably large. To further reduce the amount of data, while keeping the information on the image content, a segmentation into characteristic classes has to be performed. For this task numerous clustering algorithms are available from the fields of pattern recognition and computer vision [7].

However, to have a consistent flow of information from levels with low to high abstraction we abstain from the use of these traditional approaches and apply ‘clustering by melting’ [8]. This algorithms provides an un-biased segmentation, which captures the characteristics present in the data. It postulates an ‘interaction
energy $E_k = (\Theta_k - \tilde{\Theta})^2$ of the parameter vector $\Theta_k$ with the cluster center $\tilde{\Theta}$. Here, the index $k$ enumerates all estimations. Maximization of the entropy

$$ S = -\sum_k p(\Theta_k | \tilde{\Theta}) \log p(\Theta_k | \tilde{\Theta}), \quad (9) $$

under the constraint of conserved expectation value of the energy, leads to the contribution $p(\Theta_k | \tilde{\Theta})$ of the $k$th parameter to the cluster center $\tilde{\Theta}$ of

$$ p(\Theta_k | \tilde{\Theta}) = \frac{1}{Z} e^{-\frac{(\Theta_k - \tilde{\Theta})^2}{\sigma^2}}, \quad (10) $$

with the partition sum $Z$ and the temperature $T$. Starting at $T = 0$ with each estimation being a cluster and slowly increasing $T$, the algorithm produces a tree of clusters in the scale space. By selecting a final temperature the coarseness of the segmentation can be defined.

In Fig. 3, we show the result of such a clustering based on two features, one from level 1 and one from level 2. Note that we only show a small cut of the image data of which we estimated and clustered the features. After this segmentation, the image is characterized by regions of class labels $\omega_i$. These regions can now be further described using geometric models with basic elements such as points, lines, and polygons.

![Image of clustering result](image.png)

**Figure 3.** Un-supervised clustering using clustering by melting. Original X-SAR scene (left, scale 0–1) and result of clustering by melting (right). As features we chose the norm of the parameter vector of a 5th order Gibbs model (level 1) and the evidence of the estimation (level 2).

6. **Semantic labeling (level 5)**

On the highest level of abstraction, semantic labels are assigned to the image data. This is the first time in the process of information extraction where training data, that is, human knowledge, is incorporated. However, we note that through the selection of particular models and the selection of certain algorithms an implicit "human bias" has already been introduced on lower, "objective" levels.

The relationship between the elements $\omega_i$ of lower levels in the hierarchy and content labels $A_\nu$ is learned through the interaction with experts of the particular application area who sketch training areas for each cover type. This provides the likelihood $p(\omega_i | A_\nu)$ of each class label $\omega_i$ given the cover type $A_\nu$. 
Given some image data \( \{x_s\} \), we now calculate \( p(A_v|\{x_s\}) \), the probability of the cover type \( A_v \). In terms of the objective segmentation results \( \omega_i \) (level 3 and 4), we obtain

\[
p(A_v|\{x_s\}) = \sum_i p(A_v|\omega_i, \{x_s\}) p(\omega_i|\{x_s\}),
\]

where the sum over \( i \) extends over all class labels \( \omega_i \). Using Bayes’ equation yields

\[
p(A_v|\omega_i) = \frac{p(\omega_i|A_v) p(A_v)}{p(\omega_i)}.
\]

The prior probabilities \( p(A_v) \) enable us to incorporate prior knowledge on the particular cover type, whereas the prior probabilities \( p(\omega_i) \) are obtained by considering a large set of images. Altogether, we obtain

\[
p(A_v|\{x_s\}) = p(A_v) \sum_i \frac{p(\omega_i|A_v) p(\omega_i|\{x_s\})}{p(\omega_i)}
\]

as posterior probabilities of the cover-type \( A_v \). This way of labeling is easily extended to the case of several models. Then the descriptions on level 3 are combined to a joint set of labels \( \omega_i \). However, with many models the number of resulting elements increases very fast and the use of Bayesian prototype-trees [9] might be considered.

We give an example of semantic labeling in Fig. 4. The general information that is concentrated in the four segmentations (level 3) is transformed into maps
indicating the evidence of user-specific content characteristics (in the example, mountains and drainage areas). For the application in information system indices are easily derived from such maps. Furthermore, information theoretic methods, such as the cross-entropy, can be used to evaluate the quality of these maps and in this way help to reconfigure the chain of information down to the low-level image features on level 1.

7. Summary and conclusion

We have presented an hierarchical scheme of stochastic image analysis that reflects the nature of the information content of large images. The different levels of this scheme are arranged according to their different degree of abstraction and semantics. Inference between two levels is performed in a Bayesian framework providing both a clean way of incorporating prior knowledge and measures of evidence of the features on the higher semantic level.

This scheme promises to be an important ingredient in the process of efficient extraction and storage of information from large image data sets. Extensive tests with a large number of remote sensing images of various kinds (optical, SAR, and aerial photos) are currently being prepared. However, this scheme and the concept of stochastic image analysis may also be applicable in other areas of computer vision, such as medical image analysis or astronomical image interpretation.

Acknowledgments

We thank Prof. J. Nievergelt for an inspiring discussion. The RESURS-01 scene has been provided by NPOC Switzerland, ETH Zurich, and the X-SAR scene by DFD/DLR. Oberpfaffenhofen, Germany.

References