Probabilistic Reasoning on Object Occurrence in Complex Scenes

A. Bauer*

Fraunhofer Institute for Information and Data Processing (IITB), 76131 Karlsruhe, GERMANY

ABSTRACT

The interpretation of complex scenes requires a large amount of prior knowledge and experience. To utilize prior knowledge in a computer vision or a decision support system for image interpretation, a probabilistic scene model for complex scenes is developed. In conjunction with a model of the observer's characteristics (a human interpreter or a computer vision system), it is possible to support bottom-up inference from observations to interpretation as well as to focus the attention of the observer on the most promising classes of objects. The presented Bayesian approach allows rigorous formulation of uncertainty in the models and permits manifold inferences, such as the reasoning on unobserved object occurrences in the scene. Monte-Carlo methods for approximation of expectations from the posterior distribution are presented, permitting the efficient application even for high-dimensional models. The approach is illustrated on the interpretation of airfield scenes.

Keywords: Image Understanding, High-level vision, Bayesian inference, Selective perception, Monte-Carlo estimation

1. INTRODUCTION

Fig. 1 depicts a schematic drawing of an airfield scene. Such complex scenes (e.g. infrastructure and industrial facilities, traffic scenes and other scenes in everyday life) are composed of various kinds of objects, which are related to each other by their functional composition (e.g. a civil airfield requires to contain a runway and a tower). Many applications of remote sensing require the interpretation of images depicting such scenes as a basis for decision making and planning. Despite recent advances in automated image interpretation methods, such as the development of invariant image features and advanced classification methods, computational methods for image interpretation are still lacking far behind the performance of a human interpreter when it comes to complex scenes. One of the reasons for the superior performance of the human interpreter is its ability to incorporate prior knowledge of the scene into the interpretation.

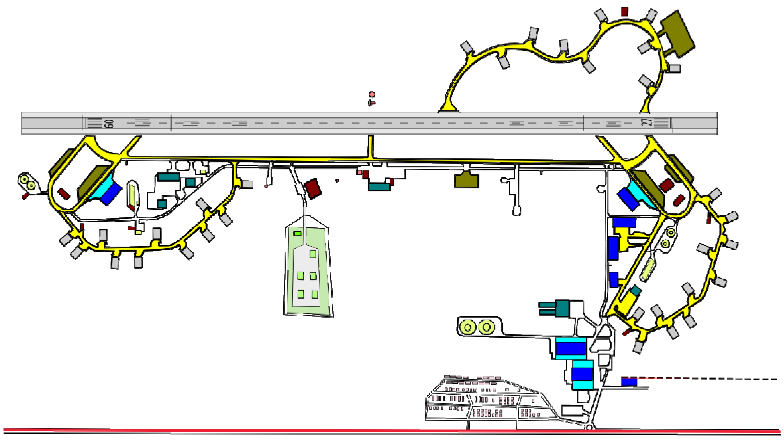

Fig. 1: Schematic drawing of an airfield scene

alexander.bauer@iitb.fraunhofer.de; phone +49 721 6091395; www.iitb.fraunhofer.de

High level vision scene interpretation methods focus on the representation of prior knowledge about objects in a scene and their spatial and taxonomical relations. They allow for reclassification of observed objects in a scene depending on their spatial alignment and provide inference about unobserved objects. Inspired by the advances in artificial intelligence in the 80s, early approaches were build on rule-based inference[1-4].

From the 90s until today, probabilistic approaches and Bayesian inference have drawn attention from cognitive psychology as well as from the computer vision community. Since then, several probabilistic approaches for high level scene interpretation have been proposed, of which only a few can be mentioned here. Rimey and Brown developed a system to control selective attention, using Bayesian Networks and Decision Theory[5]. Lueders used Bayesian Network Fragments to model taxonomy and partonomy relations between scene objects to compute the most probable scene interpretation based on perceptive information[6]. A stochastic graph grammar in conjunction with a Markov Random Field has been used by Lin et al. to recognize objects which are composed of several parts with varying alignment and occurrence[7].

Following this current, the presented approach focuses on the efficient application of probabilistic high level vision methods for the interpretation of complex scenes in remote sensing. It is able to improve results acquired from low-level methods such as automated object recognition algorithms and to control their application. At the same time, the approach is applicable for the support of a human interpreter in a decision support system for image exploitation of complex scenes.

2. BAYESIAN INFERENCE FOR SCENE INTERPRETATION

The Bayes' theorem provides a formalism to update the uncertainty about an unknown variable based on new evidence[8]. To apply the formalism, it is required to define a prior distribution of the unknown variable, which represents the prior knowledge on the variable before any new evidence has been collected. To update the prior distribution to a posterior distribution of the random variable given the new evidence, the conditional distribution of the evidence given the realizations of the unknown variable has to be defined.

Applied to scene interpretation, the unknown random variable S is the description of the scene to be interpreted (see section 3). The evidence is collected as observations, described by the random variable O (as explained in detail in section 5). Using the Bayes' theorem, the updated posterior probability is defined as

$$P(S = s_i \mid O = o_k) \propto P(O = o_k \mid S = s_i) \cdot P(S = s_i)\,, \tag{2.1}$$

For clarity, the term for the linear factor which normalizes the distribution to 1 is omitted. Fig. 2 illustrates the required elements of Bayesian inference for scene interpretation.

The prior distribution is defined by the scene model, describing the characteristics of possible scene realizations and their probability of occurrence. The observer model defines the characteristics of the observer, which can be either a human interpreter or a computer vision system. The observer acquires observations of the scene from sensor data. The observations are used to update the scene interpretation posterior distribution by Bayesian inference (also known as bottom-up process). From the posterior distribution, expectations for possible observations are derived to control the selective perception of the observer (top-down process).

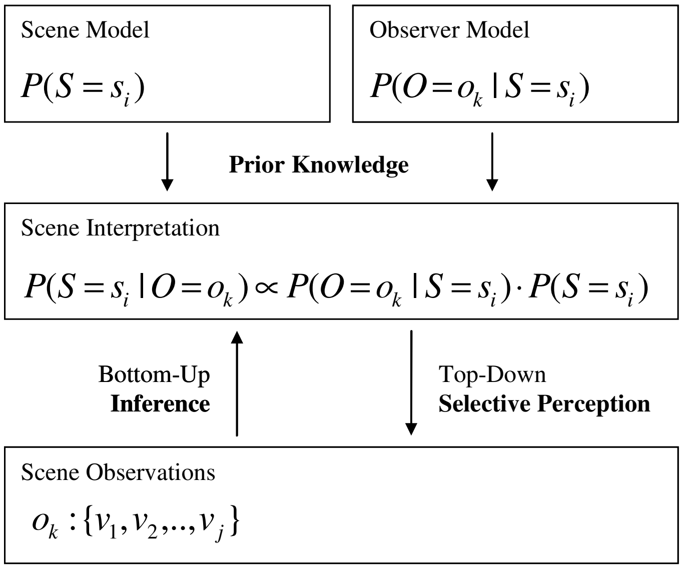

Fig. 2: Elements of Bayesian Inference for Scene Interpretation

Starting with an empty observation set, the proposed approach is able to support a recursive process of collecting evidence, updating the interpretation and deriving expectation for the most useful observation to be made next. This corresponds to the approach usually applied by a human interpreter.

3. SCENE REPRESENTATION

The scene representation formally describes a real-world scene in terms of its parts. Parts of a scene can be either single objects or object compositions. For higher efficiency, functionally related objects in a scene are often arranged in vicinity to each other to form an object composition. For example on an airfield, buildings which are dedicated to the maintenance and repair of aircraft are composed to a "Maintenance & Repair Area". This fact leads to a natural representation of a scene and its parts as a tree of objects, in which the edges define functional relations, as depicted in Fig. 3.

Each object in the real-world scene is represented by an *object instance* ω_i as part of the *scene representation* s. For easier formulation, single objects and object compositions are both generally treated as objects instances. Object instances are associated with a particular *object class* Φ, written as $\omega_i \sim \Phi$. Object classes are concepts such as "Airport", "Runway", etc.

The scene representation s itself is defined as a tree $s=(\Omega,F)$. The set of tree nodes Ω represents the set of object instances ω_i occurring in the scene to be described. The set of edges F represents their functional composition.

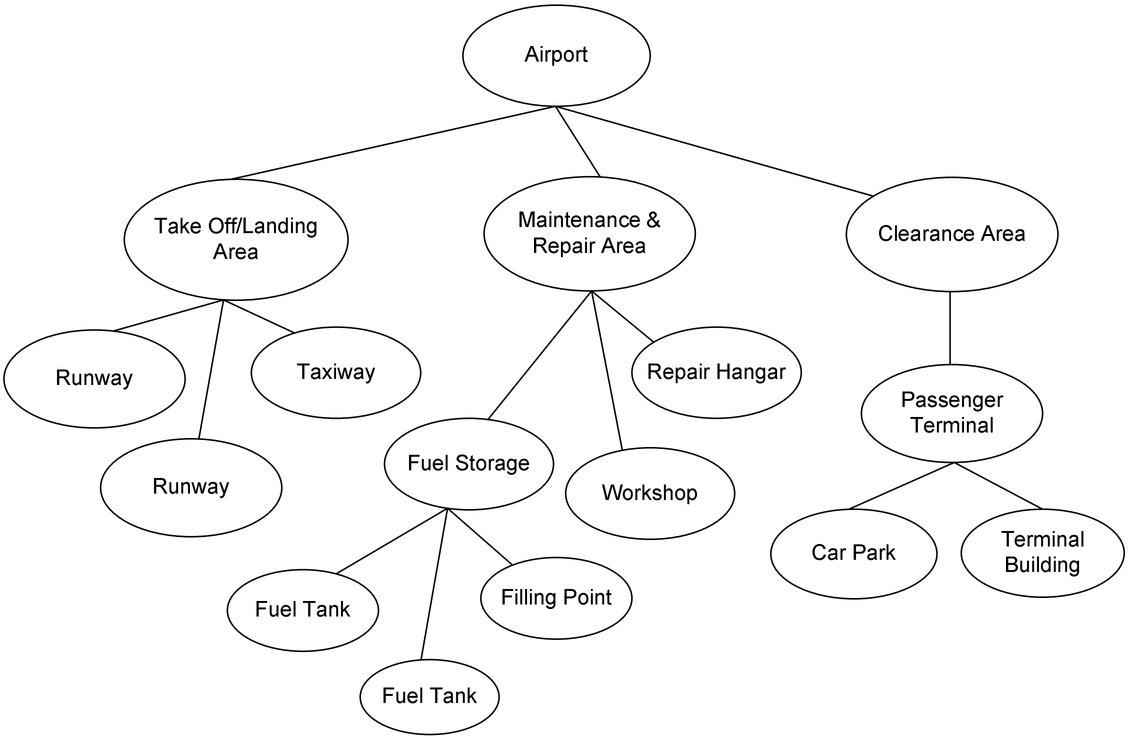

Fig. 3: Example of a scene representation tree

The number of scene representations possible to occur is very large. This is due to the high variability of real-world scenes. First, multiple object instances of the same object are likely to occur, e.g. for redundancy and extended capacity reasons. Adding or removing a single object instance results in a new scene representation, therefore a high number of variations is possible only by changing the number of occurrences of a single object class. Second, variations in the structure of real-world scenes are very likely, even if the objects in the scene serve a similar function. For example in the case of airfields, regional variations or the advances in design over time can result in very different realizations of an airfield.

4. SCENE MODEL

Observations extracted from images usually do not contain enough information to find a unique interpretation of a scene from the object signatures alone. A human interpreter therefore relies on his internal model of the real-world to resolve ambiguity. That means prior knowledge about the objects and structure of the scene to be observed is a very useful tool in image interpretation and should be also used in computer vision.

How to model prior knowledge about scenes? As explained in section 2, the probabilistic approach requires modeling prior knowledge as a prior distribution about each possible scene representation. For the reasoning about object occurrences, the scene representation has been introduced in section 3 as a scene representation tree. Therefore, the scene model must provide a probability for each possible scene representation tree, i.e. a distribution of the random variable *S,* representing the actual scene which is to be analyzed.

A second requirement for the scene model is that the acquisition of the model parameter must be tractable and comprehensible. As sufficient training data is hardly available for a complex scene domain, in most cases it is necessary to consult a human expert to establish a comprehensive and useful scene model. Therefore a straight-forward representation, based on the verbal description of an object class by a human expert, is chosen.

The scene model is defined as the set of *object class models M(Φ)*, from which all possible scene representation trees can be generated. For easier understanding, Fig. 4 shows some of the object class models necessary to model the scene representation trees of an airfield. There are three different kinds of object class models:

- *Composition models* (C-models $M_C(\Phi)$) describe the object class in terms of its possible parts. These object classes occur in parent nodes of the scene representation tree (such as "Runway Area", see Fig. 3). To represent the probability of all possible compositions, the distributions of the number of occurrences of each part object model is defined in the compositional model. Assuming independence on the occurrence of different object models, it is sufficient to define the distribution for each single object class and to establish the joint probability distribution by multiplication. In Fig. 4 the distributions are chosen to be uniform inside a reasonable interval, which simplifies the acquisition process in cooperation with an expert by using statements such as: "Airports have at least one runway, up to a maximum of 5 runways". However, more complex distributions can be used to incorporate more informative prior knowledge, also taking into account dependencies between the occurrence probabilities of different object classes.

- *Taxonomy models* (T-models $M_T(\Phi)$) define abstract object class models, which can not be used for the association in the scene representation tree, but are used for the distinction between different subtypes of an object class. For example the object class "Airfield" is further specified by the discrimination of disjunctive subtypes of that object class, as depicted in Fig. 4. For each subtype, a probability is defined which represents the conditional probability $P(\omega_i \sim \Phi_j | \omega_i \sim \Phi_T)$ of an object instance ω_i to be associated with the discriminations Φ_j given it is associated with the abstract object class Φ_T.

- *Atomic models* (A-models $M_A(\Phi)$) define object classes which can be neither further discriminated by more specific object classes nor divided into sub-parts. As depicted in Fig. 4., all objects which are not described by a box are represented by an A-model.

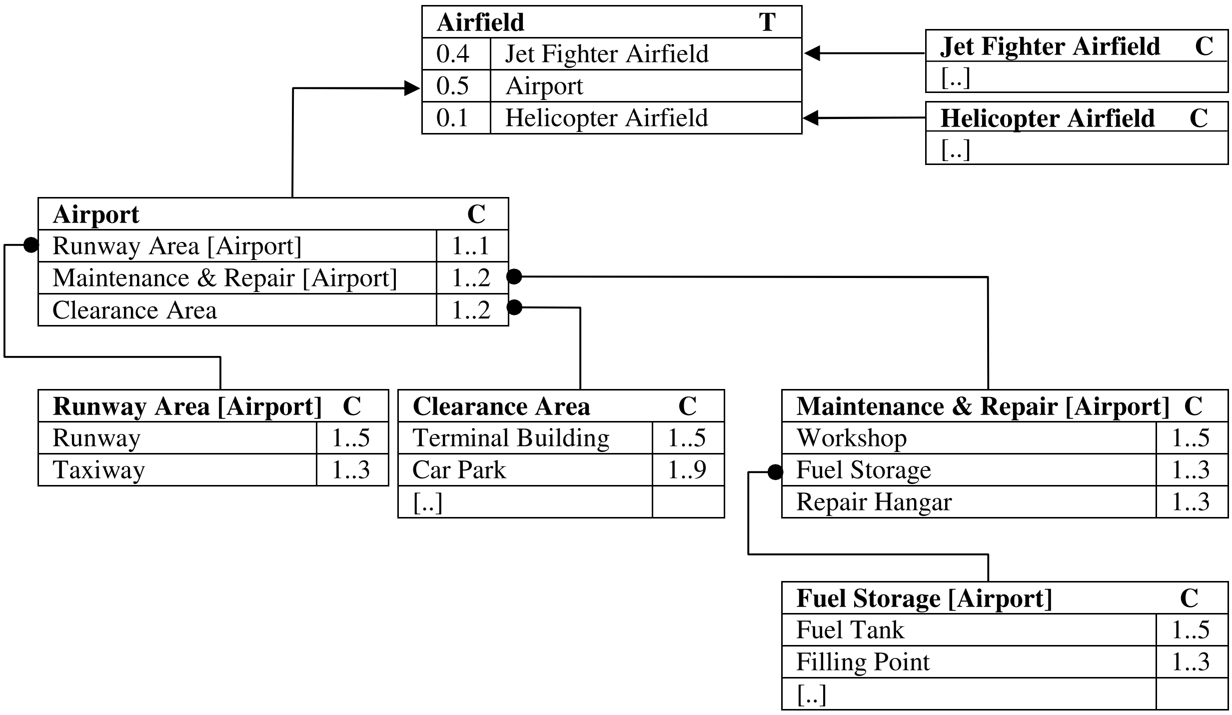

Fig. 4: Exemplary Scene Model Definition

Using the described object class models, the *scene model M(S)* is defined by the tupel $M(S)=\langle M_C, M_T, M_A, \Phi_0 \rangle$ in terms of the sets of the three kinds of object class models and the root object class model Φ_0. From the scene model, all scene representation trees $s \in S$ and their corresponding prior probability $P(S = s)$ can be generated using the following algorithm.

1. Create object instance ω_0 as the root node of the scene representation tree s, associate it with object class Φ_0. Initialize $P(S = s) = 1$, $\omega_i = \omega_0$
2. As long as object instance ω_i is associated with a T-model, choose a subtype object class Φ_j, update $P(S = s)$ with $P(S = s)\cdot P(\omega_i \sim \Phi_j|\omega_i \sim \Phi_T)$ as defined in the T-model.
3. If ω_i is a associated with a C-model:
    a. Choose a composition according to the C-model description.
    b. Create the object instances of the composition and associate them to the corresponding object classes.
    c. Update the scene prior probability $P(S = s)$ by multiplying with the composition probability according to the C-model.
4. Repeat step 2 and 3 for each newly created object instance.

Using the algorithm, for any given scene representation tree s, it is possible to determine the corresponding prior probability $P(S = s)$ by choosing the respective subtype classes in step 2 and the respective compositions in step 3. If the decisions in step 2 and 3 on the T-model subtype or the C-model composition are drawn randomly from the corresponding discrete probability distribution defined in the models, the scene model generates samples according to the prior probability $P(S = s)$. This fact is exploited for Monte-Carlo approximation in section 7.

5. OBSERVER MODEL

Information about the scene acquired from remote sensing or other information sources is summarized under the term *scene observations*. Scene observations can be either made by a human interpreter or by a computer vision system. Similar to a measurement in metrology, observations are always subject to error, especially in remote sensing. In Bayesian inference theory, observations are used to update the probability for a hypothesis to be true. In the case of scene interpretation, the scene observations are used to update the posterior distribution of the scene representation variable S (as explained in section 3). To achieve this, the conditional distribution of possible observations given a certain scene realization has to be defined.

Each single scene observation v_j contributes to a set of scene observations $o_k=\{v_1, v_2, v_3, .., v_n)$ of which all possible sets define the random variable O. The information acquired by a scene observation can be manifold, such as the position, shape, color, etc. of a scene object. In the case of reasoning about object occurrence, the information about previously detected objects and their associated object class Φ already provides relevant evidence for the estimation of the scene. Therefore, in a first step, the possible scene observations are limited to the observation of an object instance and its associated object class. An observation of the occurrence of an object of a specific object class Φ_j is therefore defined by a 1-tuple $v_k=<\Phi_j>$.

How to match scene observations and object instances of the scene interpretation tree? As only the number of occurrences of a specific object class is modeled in the scene model, it is sufficient to associate each observation with an object instance of the scene representation tree which matches the following constraints:

- The object instance is associated with the same object class as described by the observation.
- The object instance has not been associated yet with another object observation.

That means, each observation can only associated to one object instance and vice versa. Based on the match of o_k and s_j two counts are defined:

- $p(o_k,s_i)$ represents the number of object instances in s_i which could not be associated to an observation in o_k during the preceding matching procedure. In other words, it represents the number of missing observations in o_k, assuming s_i represents the real scene.

- $n(o_k,s_i)$ represents the number of observations from o_k which could not be associated to an object instance in s_i during the preceding matching procedure. It represents the number of wrong observations or false positive detections in o_k, assuming s_i represents the real scene.

Using those counts, the conditional probability of o_k given s_i is defined as being proportional to the exponential function of the linear combination of both counts:

$$P(O = o_k \mid S = s_i) \propto \exp(-[\lambda_n \cdot n(o_k, s_i) + \lambda_p \cdot p(o_k, s_i)]) \quad (5.1)$$

The parameters λ_n and λ_p control the overall influence and the balance of the both counts on the inference result. The higher the parameter values are chosen, the stronger is the influence of the observations on the posterior distribution in relation to the prior distribution. Using the parameters, the characteristics of the observer is modeled on an abstract level. The parameters are best established from statistics of training data, to fit the proposed parametric probability distribution. The model can of course be extended to take account for different imaging sensors, different characteristics of the observer with respect to the object class, etc., permitting more accurate results in inference and control of selective perception.

6. INFERENCE AND CONTROL OF SELECTIVE ATTENTION

If the posterior distribution $P(S = s_i \mid O = o_k)$ is established, inference about the scene interpretation and control of selective attention based on the current set of scene observations o_k can be provided in a number of ways:

- *Determination of the best matching scene representation* – The scene representation s_{max} which maximizes the posterior distribution $P(S = s_i \mid O = o_k)$ resembles the scene representation which best explains the observations based on the prior knowledge on possible scenes modeled by $P(S = s_i)$, and the observer model $P(O = o_k \mid S = s_i)$. If the posterior probability of s_{max} is sufficiently close to 1, s_{max} can be used as final scene interpretation result. However, to find s_{max} and to calculate the absolute posterior probability, it is necessary to evaluate all posterior probabilities of $s_i \in S$ which is not tractable for complex models. Therefore methods for approximations have to be applied, as presented in section 7.

- *Determination of the most probable unobserved object class occurrence* – To establish a complete scene representation efficiently, the attention should be focused on the most probable undetected object class occurrence. From the posterior distribution $P(S = s_i \mid O = o_k)$ the probability of occurrence of an unobserved object of a certain object class can be easily established by summing the posterior probabilities of all scene representations s_i which contain unmatched object instances associated with the object class in question. The probability of occurrence of an object class Φ is equal to the conditioned expectation

$$P(\exists \omega \in S, \omega \sim \Phi, \neg \exists v = \Phi \mid O = o_k) = E_{S|o_k}\{I_\Phi(S)\} \quad (6.1)$$

  with the indicator function $I_\Phi(s_i)$. $I_\Phi(s_i)$ takes the value 1 if an unmatched object instance of object class Φ exists in the scene representation s_i, otherwise it takes the value 0. This measure drives one of the evaluated strategies for scene interpretation presented in section 8.

- *Determination of the object class which provides the maximum information gain/effort ratio* – If there is special interest in the occurrence of a specific object class, entropy measures can be used to focus the attention on other object classes which are easy to detect and still provide high information on the object class in question. For example, if the most important goal of the scene interpretation is to discover the type of airfield, some objects such as a "weapon storage" will quickly lead to a discrimination of military and civil airfields. The expected information gain[9] (reduction of information entropy) on the probability of occurrence of the object class in question (determined by 6.1) with respect to the detection of another object class is used to determine such discriminative object classes. To calculate the information gain, the entropy of the probability of occurrence (6.1) is determined for every possible observation of other object classes and its expectation is compared to the current entropy. The observation which promises the highest change in entropy provides the maximum information gain. If additionally an detection effort can be defined (which depends for example on the average size and salience of the object class), the ratio of information gain and detection effort further provides a helpful tool to optimize the required effort to decide on the occurrence of a certain object class based on other object classes.

7. APPROXIMATIVE INFERENCE USING MONTE-CARLO ESTIMATION

Monte Carlo methods are used to calculate approximations of integrals using random samples drawn from a known distribution. Since the availability of computers, they have evolved to a powerful tool for numerical approximations of integrals. Besides in statistics and Bayesian inference, they are often used in physics for the study of systems with a high degree of freedom. Based on the definition of the Monte-Carlo estimate[10], its application for the reasoning on object occurrence is derived.

Let X be a discrete random variable, and $f_X(x) = P(X = x)$ the probability mass function of X. Then, the expectation of a function of the random variable g(X) can be expressed as:

$$E\{g(X)\} = \sum_{x \in X} g(x) f_X(x) \tag{7.1}$$

If it is possible to generate samples $x_1,..., x_n$ from X according to $f_X(x)$, it can be shown that the average

$$\tilde{g}_n(X) = \frac{1}{n}\sum_{i=1}^{n} g(x_i) \tag{7.2}$$

will almost surely converge to (**7**.1) when n goes to +∞ according to the law of large numbers. This is called a *Monte-Carlo estimate*. If it is not possible to draw samples from the distribution of X itself, using

$$E\{g(X)\} = \sum_{x \in X} g(x) f_X(x) \frac{h(x)}{h(x)} \tag{7.3}$$

the estimator can be expressed alternatively for samples generated from $h(x)$:

$$\tilde{g}_n^h(X) = \frac{1}{n}\sum_{i=1}^{n} \omega(x_i) g(x_i) \tag{7.4}$$

with the weights $\omega(x_i) = f_X(x_i) / h(x_i)$. This is called *Importance Sampling*. The variance of the estimate is minimized if the proposal distribution $h(x)$ is equal to the actual distribution of X. Therefore it is important to choose a proposal distribution which gives a good estimate on the actual distribution. If only the normalized probability mass function of X is available, as for the posterior distribution in Bayesian inference, the following approximation can be used:

$$\tilde{g}_n(X) = \frac{\sum_{i=1}^{n} \omega(x_i) g(x_i)}{\sum_{i=1}^{n} \omega(x_i)} \tag{7.5}$$

This is due to the fact that the denominator of (7.5) approximates the normalization factor. In the case of a normalized probability mass function the denominator converges to 1.

Using the Monte-Carlo Estimate with Importance Sampling (7.5) the calculations necessary for inference and control of selective attention described in section 6 become tractable by approximation. The challenge in the application of Importance Sampling is to find a good proposal distribution which well enough approximates the actual posterior distribution, and to ensure that fast sampling from the proposal distribution is possible.

The scene model described in section 4 is able to generate scene representation tree samples according to the prior distribution. Therefore the prior distribution is used as proposal distribution for the Monte-Carlo Estimation of expectations on the posterior distribution. According to (7.5), the estimator for the expectation of a function *g(S)* given the posterior distribution is defined as

$$\tilde{g}_n(S) = \frac{\sum_{i=1}^{n} \omega(s_i) g(s_i)}{\sum_{i=1}^{n} \omega(s_i)}. \tag{7.6}$$

If f(x) and h(x) are replaced with the posterior and the prior distribution, the weights are determined by

$$\omega(s_i) = \frac{P(S = s_i \mid O = o_k)}{P(S = s_i)} = P(O = o_k \mid S = s_i). \tag{7.7}$$

That means, if samples are drawn from the prior distribution, the weights are determined by the likelihood distribution of the observations given the scene representation. The estimation (7.5) is independent of the normalization of the weights, so the definition of the observation likelihood probability (6.1) does not need to be normalized.

To evaluate the accuracy of the Monte-Carlo Estimate, the mean and standard deviation of the estimate of a selected object class occurrence probability according to (6.1) is plotted for different sample sizes in Fig. 5. The selected object class was "Compass Swinging Base". The observation set contained 4 previous observations of object classes in the scene.

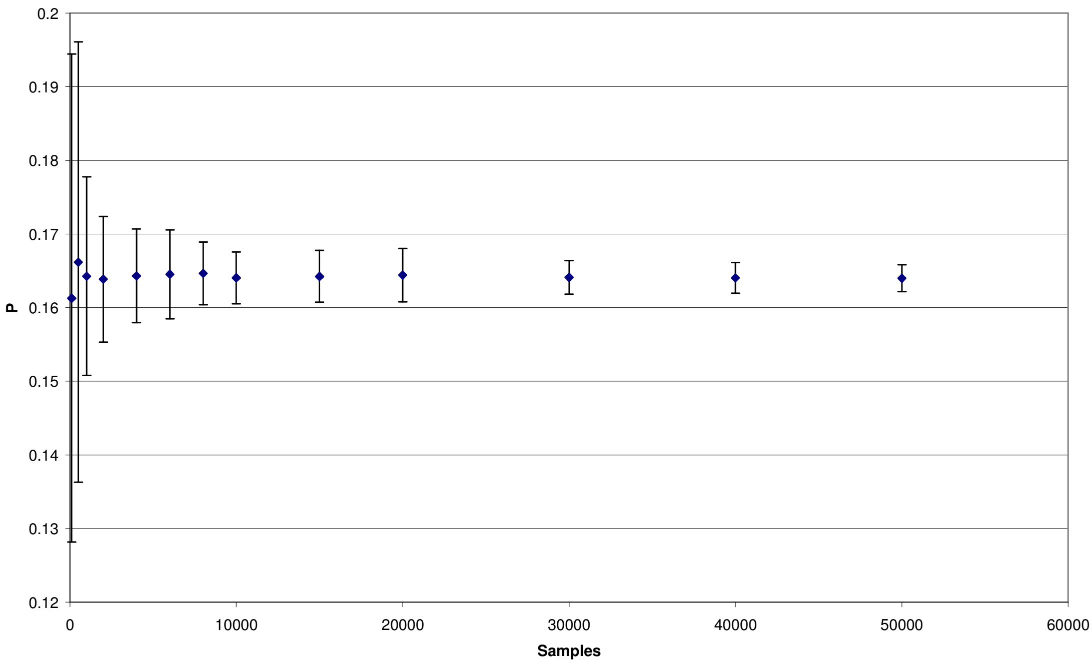

Fig. 5: Plot of mean and standard deviation of a monte-carlo estimated probability

The plot is calculated based on 100 runs for each sampling size, each run using a different random generator seed. The evaluation is based on a scene model for airfield scenes containing 150 object classes, providing a very high variability of possible scene representations. Still, the standard deviation is quickly reduced with increasing sample size. The data confirms the theoretical relation between estimate variance and sample size:

$$V(\tilde{g}) \sim \frac{1}{n} \tag{8.8}$$

The plot provides evidence that a sample size of 10.000 is sufficient to estimate the probability at a reasonable accuracy. The Java™ implementation of the estimator is able to generate and process 10.000 samples per second on an Intel™ 2.1 GHz Core 2 CPU. As the sampling distribution is not dependent on the observations, samples can be reused for different sets of observations and do not have to be redrawn for each recursion step. Therefore the calculation time for 10.000 samples is well below 1 second for a reasonable sample size.

8. EVALUATION

The practical benefit of the presented approach for image interpretation in remote sensing has to be evaluated on real-life scene interpretations. Unfortunately, a public benchmark dataset for high-level image understanding methods is not available and the acquisition of reference data for the domain of airfields is still in progress. However, from simulation of scenes and their interpretation, expectations for the benefit of the presented approach can be derived.

To compare different image interpretation strategies, a cost model is defined: It is assumed that the observer creates a scene description by assessing the scene on the existence of possible object classes. Each assessment has a cost of 1, regardless whether an object of the object class is found in the scene or not. If a matching object has been found, the observer can decide to continue the search for another object of the same class, which causes again a cost of 1, or he can switch to another object class.

To simulate the scene to be interpreted, a scene representation tree is generated according to the scene model described in section 4, sampling from the prior distribution. Three strategies are evaluated regarding their average interpretation costs:

- **Strategy 1:** Selective Perception using random object class selection
  1. Select a random object class from the scene model.
  2. Search in the scene for corresponding objects, until no more objects of that class are found.
  3. Repeat from 1, until the scene is fully described.

- **Strategy 2:** Selective Perception using prior probability
  1. Calculate the prior occurrence probability according to the scene model and arrange them in a list sorted by the occurrence probability in descending order.
  2. Select the first object class from the list.
  3. Search in the scene for corresponding objects, until no more objects of that class are found.
  4. Select the next object class on the list, repeat step 3.
  5. Repeat step 4, until the scene is fully described.

- **Strategy 3:** Selective Perception using updated posterior probability
  1. Calculate the posterior occurrence probability for each object class. Select the object class with the highest occurrence probability and which has not been unsuccessfully searched for before.
  2. Search in the scene for a corresponding object.

3. If an object is found, add the observation to the observation set. Repeat from 1. until the scene is fully described.

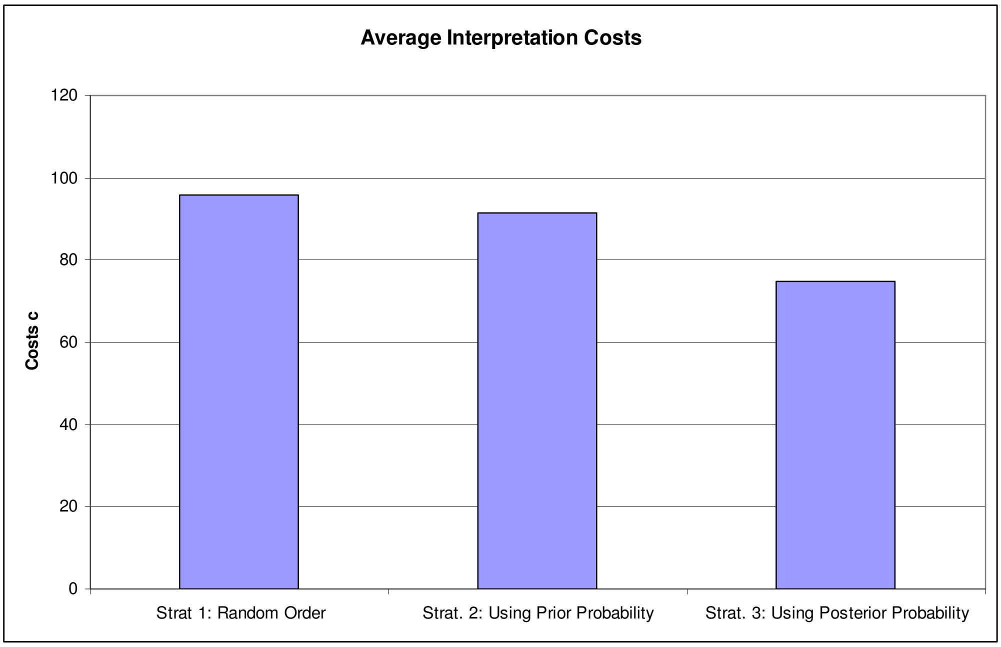

Fig. 6: Average interpretation costs of different strategies of selective perception

The different strategies were evaluated for 250 scene representation trees drawn randomly from the scene model. The average number of objects in the simulated scene representation trees was 43. Only the object classes represented by an atomic model were considered. Fig. 6 shows a diagram of the average interpretation costs for the presented strategies. The use of the prior occurrence probability in strategy 2 reduces the average interpretation costs by ca. 5 %. If the posterior occurrence probability is applied in strategy 3, a significant average improvement of 21 % is achieved. The diagram illustrates the striking benefit of combining prior knowledge and new evidence collected from observations to form optimal expectations for the most likely object class occurrences.

9. CONCLUSION AND OUTLOOK

The proposed probabilistic approach for high-level scene interpretation permits inference on undetected objects in a scene based on previous observations taken from the scene. The scene model is defined in a human understandable and natural way; such that even in the absence of training data, a human expert is able determine the required parameters using composition and taxonomic models. The observer model characterizes the observer to take account for its specific error rate, whereas the observer can be either a human interpreter or low-level object-recognition algorithms. Several applications of the posterior distribution to control the interpretation process were proposed.

The probability of occurrence of an object class can be calculated at reasonable accuracy and efficiency even for very complex models using Monte-Carlo Importance Sampling, which has been illustrated exemplarily for a single object class occurrence probability. A simulation based evaluation of different image interpretation strategies shows, that the application of selective perception using the inferred posterior probability has significant benefit on the efficiency.

In the future, the proposed framework for Bayesian inference shall be extended to incorporate spatial relations between objects. More advances approximation methods such as Markov Chain Monte-Carlo will be examined, to compensate for the increase in complexity due to the spatial modeling. The feasibility of probabilistic modeling for dynamic situations will be examined[11]. The benefit in integrating the model in a decision-support system[12] for image interpretation will be evaluated in a remote sensing scenario, by comparison of the performance of two groups of users, which either use the decision support system or rely on their experience.

REFERENCES

[1] Matsuyama, T. and Hwang, V., SIGMA: A Knowledge-Based Aerial Image Understanding System, Plenum Press, (1990).

[2] Russ, T. A., MacGregor, R. M., Salemi, B., Price, K. and Nevatia, R., "VEIL: Combining Semantic Knowledge with Image Understanding," ARPA Image Understanding Workshop, (1996).

[3] Dillon, C., and Caelli, T., "Learning image annotation: the CITE system," Journal of Computer Vision Research, 1(2), 90-121(1998).

[4] Hanson, A., Marengoni, M., Schultz, H., Stolle, F., Riseman, E. and Jaynes, C., "Ascender II: a framework for reconstruction of scenes from aerial images," Workshop Ascona 2001: Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), 25-34 (2001).

[5] Rimey, R. D. and Brown, C. M., "Control of Selective Perception Using Bayes Nets and Decision Theory", International Journal of Computer Vision, Vol 17, 173-109 (1994)

[6] Lueders, P., "Scene Interpretation Using Bayesian Network Fragments", Lecture Notes in Economics and Mathematical Systems, Volume 581, 119-130, doi:10.1007/3-540-35262-7_7, (2006).

[7] Lin, L., Wu, T., Porway, J. and Xu, Z., "A stochastic graph grammar for compositional object representation and recognition", Pattern Recognition, Volume 42, Issue 7, Pages 1297-1307, ISSN 0031-3203, doi: 10.1016/j.patcog.2008.10.033, (2009).

[8] Robert, P. R., The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, Springer (2nd ed. 2001, 2nd printing, 2007).

[9] Mitchell, T., Machine Learning, McGraw-Hill Education (1997).

[10] Koch, K. R., Introduction to Bayesian Statistics, Second Edition, Springer (2007).

[11] Bauer, A., Emter, T., Vagts, H. Beyerer, J., "Object-Oriented World Model for Surveillance Systems", Conference Proceedings: Future Security, 4th Security Research Conference, Karlsruhe (2009) (accepted).

[12] Bauer, A., "Assisted Interpretation of Infrastructure Facilities from Aerial Imagery", Proceedings of SPIE, Vol 7481 (2009) (accepted).