A semantic approach to the efficient integration of interactive and automatic target recognition systems for the analysis of complex infrastructure from aerial imagery

A. Bauer*, E. Peinsipp-Byma
Fraunhofer Institute for Information and Data Processing (IITB), 76131 Karlsruhe, GERMANY

ABSTRACT

The analysis of complex infrastructure from aerial imagery, for instance a detailed analysis of an airfield, requires the interpreter, besides to be familiar with the sensor's imaging characteristics, to have a detailed understanding of the infrastructure domain. The required domain knowledge includes knowledge about the processes and functions involved in the operation of the infrastructure, the potential objects used to provide those functions and their spatial and functional interrelations. Since it is not possible yet to provide reliable automatic object recognition (AOR) for the analysis of such complex scenes, we developed systems to support a human interpreter with either interactive approaches, able to assist the interpreter with previously acquired expert knowledge about the domain in question, or AOR methods, capable of detecting, recognizing or analyzing certain classes of objects for certain sensors. We believe, to achieve an optimal result at the end of an interpretation process in terms of efficiency and effectivity, it is essential to integrate both interactive and automatic approaches to image interpretation. In this paper we present an approach inspired by the advancing semantic web technology to represent domain knowledge, the capabilities of available AOR modules and the image parameters in an explicit way. This enables us to seamlessly extend an interactive image interpretation environment with AOR modules in a way that we can automatically select suitable AOR methods for the current subtask, focus them on an appropriate area of interest and reintegrate their results into the environment.

Keywords: image interpretation, interactive support systems, automatic object recognition, high level vision, ontology

1. INTRODUCTION

The ongoing advances in imaging sensors and aerial sensor platforms, particularly unmanned aerial vehicles (UAV), have made aerial reconnaissance a powerful tool to gain information vital to strategic and tactical decision making. The quality of information still highly depends on image interpretation. In the field of aerial reconnaissance, the most prominent objective of image interpretation is to describe recognized objects, states and behaviors relevant to the reconnaissance task.

The evaluation of a countries critical infrastructure, predestined for the application of aerial reconnaissance, requires a detailed analysis of objects like airfields, harbors, lines of communication and heavy industry. Unlike the recognition of single objects like land vehicles, airplanes or ships, merely requiring exhaustive knowledge of existing object classes and their features, the interpretation of a complex infrastructure demands a deeper understanding of the objects' functionality, their interrelations and the meaning of object arrangements in a specific infrastructure domain.

The task of image interpretation in the field of reconnaissance is primarily performed by humans with their specific skill for visual pattern recognition. Automated object recognition (AOR) algorithms are able to detect, recognize and analyze objects from the image signal. Detection and screening algorithms detect relevant objects in a scene or identify regions of interest (ROIs) potentially containing relevant objects. Object recognition is possible by classification of objects based on features extracted from the image. Algorithms for analysis of objects determine distinct characteristics of an object like their material, geometric proportions or their state of operation. To account for errors of such algorithms introduced by noise or ambiguity, a human interpreter has to review the results generated by automatic algorithms. Assuming a reasonable error rate, it can be shown that the overall performance of an interpretation process is improved by the application of computer vision algorithms[1]. However, the human is required to correctly judge his own ability and the ability of AOR algorithms for optimal performance of assisted image interpretation[2].

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

High level vision methods focus on the representation of knowledge about objects and their spatial and taxonomical relations, allowing for reclassification of observed objects in a scene depending on their spatial alignment and providing inference about unobserved objects. Several methods have been investigated for the application of high level vision in aerial image interpretation based on rule-based inference[3,4], relaxation labeling[5] or bayesian networks[6].

Interactive support systems for aerial image interpretation rely on the human skills in detecting, recognizing and analyzing objects from image data, providing support at making decisions about the image scene in an interactive manner. Basic support is realized by handbooks, containing a catalogue of objects relevant to the interpretation domain[7]. Object characteristics are explained by reference images and text. To support the interpreter at classifying an object, high level vision methods are applied to generate classification hypothesis based on object features[8] and other objects observed in the environment. Similar methods are used to generate hypotheses about undetected objects in the scene and their expected location. The location of undetected objects can be presented to the interpreter as regions of interest in the image.

AOR algorithms and interactive support systems target at different aspects of image interpretation. AOR algorithms work on the signal level and output information about the image on the object level. High level methods applied in interactive support systems are able to process and visualize information on the object level. Therefore we think that the two approaches can complement one another very well and a concept for their combination should be developed.

2. METHODOLOGY

To benefit from the integration of AOR algorithms and an interactive support system into an image interpretation environment, it is necessary to establish interoperability between different heterogeneous components. Interfaces have to be available to transmit information from one system to another and a common data model must be defined. To achieve this, a software architecture for an integrated image interpretation environment is defined. An ontology is developed to specify the data model and to represent domain knowledge in a formal way. Assuming that interoperability between the components is established, an algorithm for alternating employment of interactive and automatic interpretation methods is proposed.

2.1 System Architecture

Fig. 1 shows the system architecture for an extended image interpretation environment, which is separated into the image viewer, the interpretation support system and additional resources. One essential component of an image interpretation environment is the image viewer, allowing the image interpreter to visualize the image data according to his needs. The interactive support system is modeled as an independent component. AOR algorithms, explaining texts and images and other useful resources for interpretation are modeled as interpretation resources, which are either present inside or outside the interactive support system.

Many different image viewers are used in existing image interpretation environments. Moreover, it is difficult to integrate an interactive software component with a complex graphical user interface (GUI) into proprietary software. However, most image viewers provide an application programming interface (API) allowing to access and transfer image data and image annotations to external applications. A simple interface can be developed to synchronize the interactive support system with the image viewer. This way it is possible to visualize results from AOR algorithms on the image display and at the same time to feed back user inputs made inside the image viewer into the interactive support system.

The GUI of the interactive support system is displayed on the same physical screen as the image viewer or displayed on a second monitor. The inference engine of the interpretation support system implements high level vision methods for reasoning about objects. The information about relevant objects and their relations are stored in the domain model. Explaining texts and images provided by the interactive support system are stored in an internal resource database. The description of interpretation resources is represented in a resource metadata database, establishing the link between the domain model and either internally or externally stored resources. Objects recognized either by the human interpreter or by AOR algorithms are instantiated and stored inside the scene description, representing the current state of the interpretation process. An interface exists to exchange object and image information with AOR algorithms.

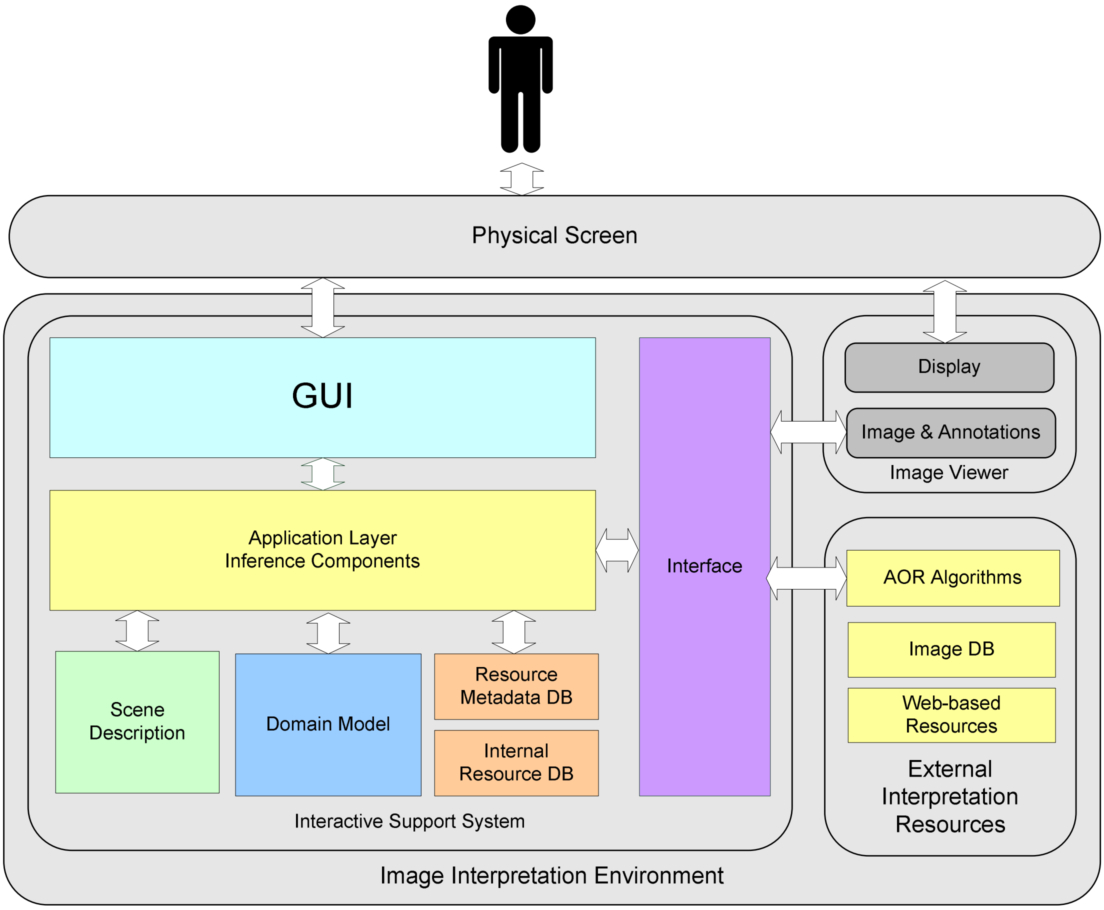


Fig. 1. Architecture of the Image Interpretation Environment

2.2 Why using an ontology?

In computer science, an ontology defines a representation with which it is possible to model a domain of knowledge or discourse[9]. The representation is made of a few primitives:

- **Individuals** (Objects) are used to represent concrete objects which are either physical such as people, animals, molecules, or virtual such as units, functions etc.
- **Concepts** (Classes) represent an abstract subset of individuals or other concepts, like for example the class of all people or the class of all cars.
- **Properties** (Attributes) are introduced to describe individuals in terms of a property name and value, for example the age of a person.
- **Relations** are a special kind of properties, used to model relationships between individuals.
- **Restrictions** are introduced to describe concepts in terms of properties which have to be satisfied by all individuals of a concept.

The commitment of all actors (people and computer programs) involved in a system to an ontology as a formal specification of the vocabulary of a domain enables easy sharing and reuse of knowledge. In contrast to a database scheme or a data model, ontologies are formulated in a representation language having high expressive power. That means, semantic constraints like for example “A is a subclass of B” can be made explicit, whereas in a database scheme such a relation has to be represented implicitly by foreign keys in a table column. The additional expressiveness of ontology languages enables the use of generic reasoning mechanisms which are able, for instance, to decide if an individual belongs to a certain concept based on its properties and relations.

An ontology for the domain of image interpretation provides a powerful language for sharing and reusing knowledge between components defined in the system architecture of an image interpretation environment (see 2.1).

2.3 Ontology Development

Ontology Development is a challenging task and still a topic of research[10]. There is no single correct way to model a domain and the feasibility of an ontology for a specific task can only be evaluated in application. Therefore ontology development is an inherently iterative process. For the development of an image interpretation ontology a four-step development methodology has been used:

1. Definition of Domain and Scope
2. Reuse of existing ontologies
3. Ontology Design
4. Evaluation and Application

3. ONTOLOGY FOR AN IMAGE INTERPRETATION ENVIRONMENT

To simplify the understanding of the ontology developed, it is divided into three parts: an infrastructure ontology describing the infrastructure domain model, an image ontology to represent images and regions of interest, and a resource ontology to describe interpretation resources so that they can be used by the interactive support system.

3.1 Infrastructure Ontology

The infrastructure ontology is defined to provide the basic concepts to model objects and relations relevant for an infrastructure domain. Fig. 2 shows the ontology in UML notation. The following concepts have been introduced:

- **Domain** is used to distinguish different domains of infrastructures, for example airfields and harbors.
- **EntryCategory** represents categories of objects which can be easily classified by their unique appearance in the image, without taking into account the context and surrounding of the object. A few examples are stated in the ontology like buildings, tarmacs and antennas. The set of categories can be extended and refined by defining additional subclasses.
- **Geometry** and **Projection** (tagged with the prefix "geo") are necessary to express the geometric shape and position in geographic coordinates. A geometry can be any closed shape, expressed in coordinates of a specific projection. A generic representation for two-dimensional geometry in world coordinates has been already defined by the Simple Features Specification of the Open Geospatial Consortium[11]. This way, geometries and projections can be represented as a string which is stored with every individual of a geometry and a projection.
- **InfrastructureObject** is the basic concept to model classes of objects of a domain. Each object is assigned to a domain and an entry category. The geometry of the object is described by an individual of the concept geometry.

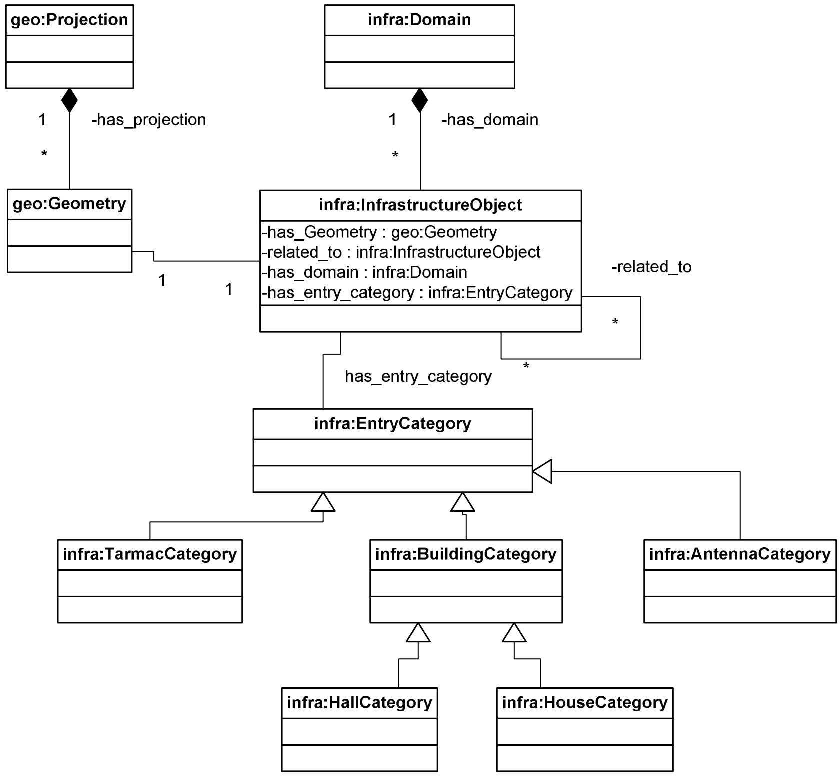


Fig. 2. Infrastructure Ontology

Relations between objects are summarized under the property “related_to”. To distinguish different types of relations, a relation hierarchy is defined as shown in table 1.

Table 1. Object relation hierarchy

Relation Hierarchy	Description
1. **related_to**	Abstract object relation
1.1. **spatial_related_to**	Spatial relation
1.1.2. **contained_in**	Spatial containment
1.1.3. **near_to**	Nearness relation
1.1.3.1. **adjacent_to**	Adjacency relation
1.2. **functional_related_to**	Functional relation
1.2.1. **functional_part_of**	Functional partonomy relation

Spatial relations are relations which can be inferred from the geometry of objects. The perception of spatial relationship for objects of a specific domain has been studied by Abella[12]. Based on the geometric alignment of objects in an image, semantic relations can be derived automatically. The mapping between geometry and semantic relations has to be calibrated to the users of a specific domain. Spatial relations defined here are assumed to be used according to a suitable calibration.

Functional relations are relations which describe functional dependence between objects. Functional parts of an object, forming objects themselves, are mandatory to make the whole object fulfill its function. This relation is especially inherent to objects in infrastructure where larger areas populated by different objects are dedicated to fulfill a specific function. The refueling system of an airport which is assembled from very different objects (fuel tanks, fuel pipes, fuel stations, roads, etc.) is a typical example.

The hierarchical formation of relations simplifies the extension by additional relations without compromising existing relations. Super ordinate relations always apply if at least one subordinate relation applies.

3.2 Image Ontology

To exchange images involved in the interpretation task an image ontology has to be defined. Fig. 3 sketches the concepts (tagged with the prefix "img") necessary to represent images and their parameters.

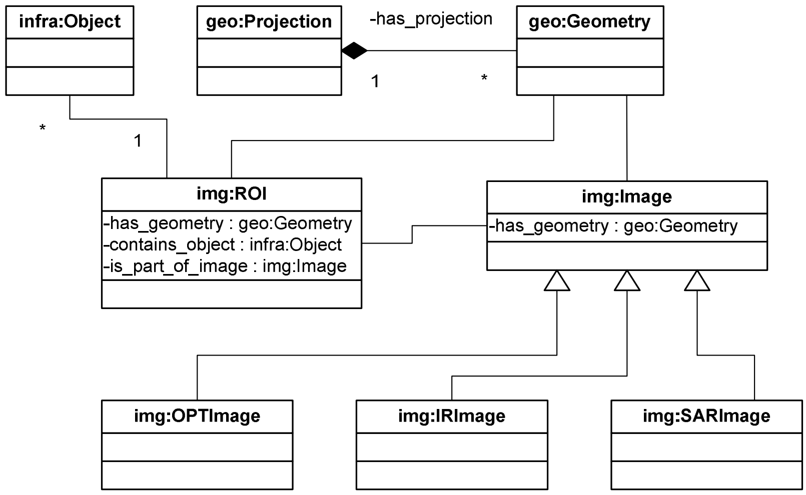


Fig. 3. Image Ontology

The concepts geometry and projection introduced in section 3.1 are reused to represent georeference of an image. The footprint of an image can be represented as a polygon, providing reference points for edge pixels. Images can be further distinguished by the type of sensor used for acquisition, allowing for representation of distinctive parameters for a specific sensor type. Three concepts representing images acquired by SAR (Synthetic Aperture Radar) sensors, optical sensors and infrared sensors are introduced. A region of interest (ROIs) is a region in an image which is supposed to contain specific classes of objects. Properties of ROIs are their geometry, the containing objects and the corresponding image.

3.3 Resource Ontology

As explained in the section 2.1, AOR algorithms, images and documents helpful for an interpretation task are summarized under the general concept resource. Every resource has a subject, describing the scope of a document or reference image, or the object classes an AOR algorithm is able to recognize. A resource ontology (Fig. 4) is defined to represent those concepts, relating to previously defined concepts for the representation of images and infrastructure. AOR algorithms are described by a relation to entry categories, objects and image types they are suitable for. For the description of reference images, showing examples of object appearances in different sensors, the image concept introduced earlier can be reused. The concepts of the resource ontology are tagged with the prefix "res".

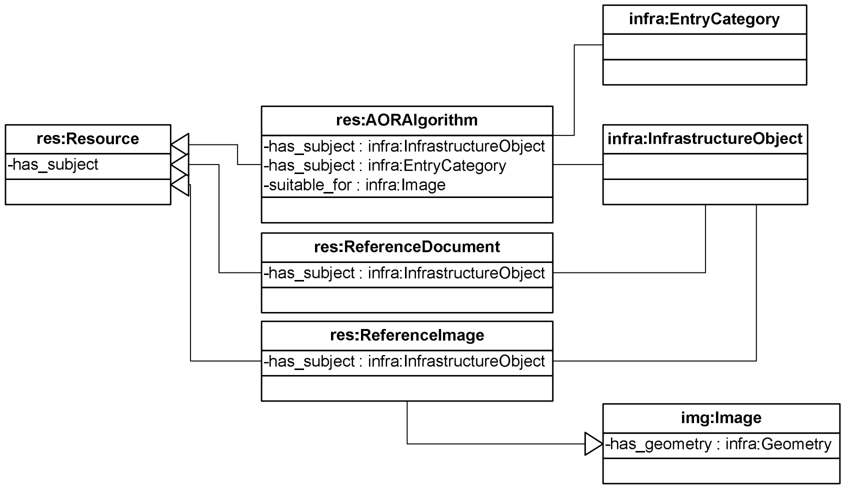
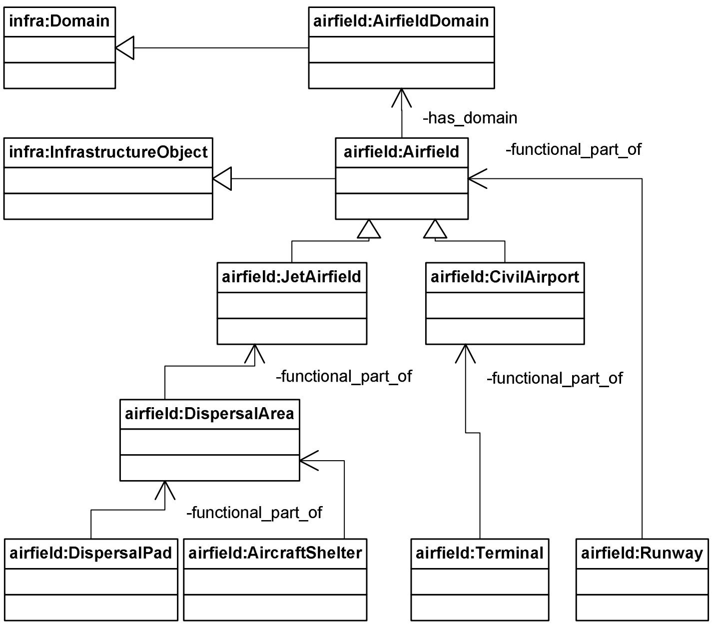


Fig. 4. Reference Ontology

3.4 Airfield Ontology

As an example for modeling a specific domain, a small section of an airfield domain model is shown, expressed by the concepts of the infrastructure ontology. Relations holding true for every individual of an object class are expressed as restriction on the respective relation (see section 2.2). The relation "functional_part_of" is used to assemble different kinds of airfields from their functional parts. Objects which can be found in every airfield are related to the root concept of the airfield domain (denoted as “airfield:Airfield”). Using concept inheritance, existing object concepts can be reused and specialized by the extension of their relational restrictions.



4. CONTROL ALGORITHM

The image interpretation ontology ensures that all components of the system can exchange image and object information and is used to represent the domain model. To benefit from the integration of interactive and automatic approaches, the interplay between the interactive support system and AOR algorithms has to be controlled by an algorithm.

The output of the inference engine generates hypothesis about expected objects and corresponding regions of interest. Both the human interpreter and AOR algorithms are able to confirm those hypotheses and newly detected objects or confirmed hypotheses can again lead to updated hypothesis generated by the inference engine. This configuration suggests an iterative employment of inference engine and AOR algorithms, until scene description has stabilized and no additional hypotheses are generated. The scene description will then be presented to the human interpreter, who is now able to verify and correct the scene description. Objects which have not been detected by AOR algorithms are added to the scene description. Hypotheses can be evaluated and confirmed, also resulting in a change of the scene description. If the scene description has changed, a new iteration of alternating invocation of the inference engine and AOR algorithms is started and the result is again presented to the user for verification. The proposed control algorithm can be sketched by seven steps:

Step 1. Select all object concepts and entry categories which are part of the root object concept of the domain.

Step 2. Select AOR algorithms available to detect selected objects concepts and entry categories.

Step 3. Run selected algorithms and collect results as object individuals or ROI individuals.

Step 4. Compute spatial relations.

Step 5. Run inference engine to reclassify objects according to their object relations and generate object hypotheses and ROIs.

Step 6. If the scene description has changed, select AOR algorithms able to confirm generated object hypotheses and continue at Step 3.

Step 7. Present detected objects and ROIs to the human interpreter and collect input from the interpreter represented as object individuals. Continue at Step 4.

5. CONCLUSION

The combination of interactive and automatic approaches to the interpretation of complex infrastructure is promising, as the two approaches assist the human interpreter at different aspects of the image interpretation problem. A system architecture is proposed for the integration of interactive support systems and AOR algorithms into the image interpretation environment. An image interpretation ontology has been developed to allow the communication between system components and to represent domain knowledge necessary for reasoning about objects and relations. A control algorithm ensures that AOR algorithms and inference methods are efficiently selected and the results can be presented and verified by the image interpreter.

To realize the proposed approach, existing high level vision algorithms will be evaluated and enhanced for the application in interactive support systems.

REFERENCES

1. Eckstein, B. A. and Irivine, J. M., "Evaluating the Benefits of Assisted Target Recognition," Proceedings 30th Applied Imagery Pattern Recognition Workshop (AIPR), 39-45 (2001).
2. Entin, E. B., Entin, E. E. and Servaty, D., "Optimizing Aided Target-Recognition Performance", Proceedings of the Human Factors and Ergonomics Society, 40th Annual Meeting, 233-237 (1996).
3. Matsuyama, T. and Hwang, V., *SIGMA: A Knowledge-Based Aerial Image Understanding System*, Plenum Press, (1990).
4. 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).
5. Dillon, C., and Caelli, T., "Learning image annotation: the CITE system," Journal of Computer Vision Research, 1(2), 90-121(1998).
6. 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).
7. Endres, G., Gething, M., *Jane's Aircraft Recognition Guide*, Collins, 3rd edition, (2001).
8. Geisler, J., Pallmer, D., Littfass, M., Kerker, R., Schoenbein, R. and Schumacher, W., "RecceMan - Interactive Interface for Target Recognition Support and Training," Proceedings of the fourth joint international military sensing symposium 45 (1), Veridian ERIM International, 347-360 (2000).
9. Gruber, T. R., "A translation approach to portable ontologies," Knowledge Acquisition 5(2), 199-220 (1993).
10. Gomez-Perez, A., Fernández-Lopez, M. and Corcho, O., *Ontological Engineering*, Springer Press, (2004).
11. Open Geospatial Consortium, Inc., OpenGIS® Simple Features Implementation Specification for SQL 1.1, http://www.opengeospatial.org/standards/sfs, (1999).
12. Abella, A., *From Imagery to Salience: Locative Expressions in Context*, Ph.D. thesis, Columbia University, (1995).