Knowledge-Assisted Image Analysis Based on Context and Spatial Optimization

G. Th. Papadopoulos, Aristotle University of Thessaloniki and Informatics and Telematics Institute/Centre for Research and Technology, Greece

Ph. Mylonas, National Technical University of Athens, Greece

V. Mezaris, Informatics and Telematics Institute/ Centre for Research and Technology, Greece

Y. Avrithis, National Technical University of Athens, Greece

I. Kompatsiaris, Informatics and Telematics Institute/ Centre for Research and Technology, Greece

ABSTRACT

In this article, an approach to semantic image analysis is presented. Under the proposed approach, ontologies are used to capture general, spatial, and contextual knowledge of a domain, and a genetic algorithm is applied to realize the final annotation. The employed domain knowledge considers high-level information in terms of the concepts of interest of the examined domain, contextual information in the form of fuzzy ontological relations, as well as low-level information in terms of prototypical low-level visual descriptors. To account for the inherent ambiguity in visual information, uncertainty has been introduced in the spatial relations definition. First, an initial hypothesis set of graded annotations is produced for each image region, and then context is exploited to update appropriately the estimated degrees of confidence. Finally, a genetic algorithm is applied to decide the most plausible annotation by utilizing the visual and the spatial concepts definitions included in the domain ontology. Experiments with a collection of photographs belonging to two different domains demonstrate the performance of the proposed approach.

Keywords: *context; knowledge-assisted analysis; multimedia ontologies; semantic annotation; semantic image analysis*

INTRODUCTION

Recent advances in both hardware and software technologies have resulted in an enormous increase of the number of images that are available in multimedia databases or over the Internet. As a consequence, the need for techniques and tools supporting their effective and efficient manipulation has emerged. To this end, several approaches have been proposed in the literature regarding the tasks of indexing, searching, and retrieval of images.

The very first attempts to address these issues concentrated on visual similarity assessment via the definition of appropriate

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

quantitative image descriptions, which could be automatically extracted and suitable metrics in the resulting feature space. Coming one step closer to treating images the way humans do, these were later adapted to a finer granularity level, making use of the output of segmentation techniques applied to the image (Smeulders, Worring, Santini, Gupta, & Jain, 2000). While low-level descriptors, metrics, and segmentation tools are fundamental building blocks of any image manipulation technique, they evidently fail to fully capture the semantics of the visual medium by themselves; achieving the latter is a prerequisite for reaching the desired level of efficiency in image manipulation. To this end, research efforts have concentrated on the semantic analysis of images, combining the aforementioned techniques with *a priori* domain specific knowledge, so as to result in a high-level representation of images (Al-Khatib, Day, Ghafoor, & Berra, 1999). Domain specific knowledge is utilized for guiding low-level feature extraction, higher-level descriptor derivation, and symbolic inference.

Depending on the adopted knowledge acquisition and representation process, two types of approaches can be identified in the relevant literature: implicit, realized by machine learning methods, and explicit, realized by model-based approaches. The usage of machine learning techniques has proven to be a robust methodology for discovering complex relationships and interdependencies between numerical image data and the perceptually higher-level concepts. Moreover, these elegantly handle problems of high dimensionality. Among the most commonly adopted machine learning techniques are Neural Networks (NNs), Hidden Markov Models (HMMs), Bayesian Networks (BNs), Support Vector Machines (SVMs), and Genetic Algorithms (GAs) (Assfalg, Berlini, Del Bimbo, Nunziat, & Pala, 2005; Zhang, Lin, & Zhang, 2001). On the other hand, model-based image analysis approaches make use of prior knowledge in the form of explicitly defined facts, models, and rules (i.e., they provide a coherent semantic domain model to support "visual" inference in the specified context) (Dasiopoulou, Mezaris, Kompatsiaris, Papastathis, & Strintzis, 2005; Hollink, Little, & Hunter, 2005).

Regardless of the adopted approach to knowledge representation, the inclusion of spatial information in the knowledge exploited during the analysis process makes necessary the definition and extraction of spatial relations from the visual medium. The relevant literature considers two categories of approaches for the latter task: angle-based and projection-based approaches. Angle-based approaches include Wang, Makedon, Ford, Shen, and Golding (2004), where a pair of fuzzy k-NN classifiers are trained to differentiate between the *Above-Below* and *Left-Right* relations, and the work of Millet, Bloch, Hede, and Moellic (2005) where an individual fuzzy membership function is defined for every relation and applied directly to the estimated angle-histogram. Projection-based approaches include Hollink et al. (2004), where qualitative directional relations in terms of the centre and the sides of the corresponding objects' MBRs were defined, and Skiadopoulos et al. (2005), where the use of a representative polygon was introduced.

Furthermore, in the real world, objects exist in a context. Representing context is a research issue of great importance (Edmonds, 1999) affecting the quality of the produced results, especially in the field of multimedia analysis in general and knowledge-assisted image analysis in particular. The latter can be defined as a tightly coupled and constant interaction between low-level image analysis algorithms and higher-level knowledge representation (Athanasiadis et al., 2005), an area where the role of context is crucial. In recent years, a number of different context aspects related to image analysis have been studied, and a number of different approaches to model context representation have been proposed (Zhao, Shimazu, Ohta, Hayasaka, & Matsushita, 1996).

In this article, an approach to knowledge-assisted image analysis based on coupling explicit prior knowledge in the form of prototypical instances, spatial relations, and contextual information is described. This ap-

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

proach is part of the aceMedia[1] EC-IST project dealing with efficient multimedia content access and personalized delivery. More specifically, a novel ontological representation for context is utilized combining fuzzy theory and fuzzy algebra (Klir & Yuan, 1995) with characteristics derived from the Semantic Web, like the statement's reification technique (W3C, RDF Reification, 2004). In this process, confidence values of labels assigned to regions on the basis of low-level visual information similarity are optimized according to a context-based confidence value readjustment (CCVR) algorithm (Mylonas, Athanasiadis, & Avrithis, 2006). This is followed by a second optimization process utilizing the output of the former as well as spatial information as input to a genetic algorithm, which decides on the optimal semantic interpretation of the image.

The article is organized as follows: "System Overview" presents the aceMedia system architecture. "Low-Level Visual Information Processing" discusses low-level visual information processing, whereas "Knowledge Infrastructure" describes the employed knowledge infrastructure. "Context and Spatial Optimization" addresses the issues of context and spatial optimization making use of the previously defined processing methods and knowledge representations. Experimental results for a collection of photographs belonging to two different domains are presented in the sixth section and conclusions are drawn in the final section.

SYSTEM OVERVIEW

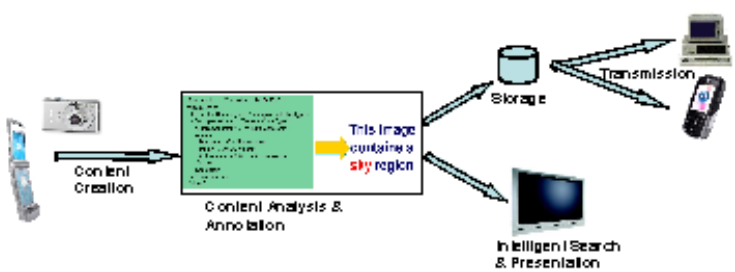
Overall Architecture

The current approach was developed in the aceMedia project (aceMedia) and addresses the issues of efficient multimedia content access and personalized delivery by integration of multimedia analysis technologies with Semantic Web tools and techniques (Figure 1). More specifically, aceMedia develops tools to automatically analyze content, generate semantic metadata and annotation, and support personalized and intelligent content search and retrieval services (Figure 2).

A key component of the aceMedia system is its Knowledge Assisted Analysis module (KAA), which creates automatic multimedia annotations using an ontology driven approach. In KAA, low-level image features are extracted from the multimedia content using tools such as segmentation to atom regions and MPEG-7 descriptors extraction. Conversion of the MPEG-7 descriptors into an RDF representation enables reasoning to be applied such that objects and areas in the scene can be identified with reference to the appropriate domain ontology. Subsequently, the KAA module, using a methodology detailed in the sequel, decides on the labeling of the atom regions with a set of concepts from the domain ontology. The approach that is followed is generic and applicable to any domain as long as appropriate domain ontologies are designed and made available.

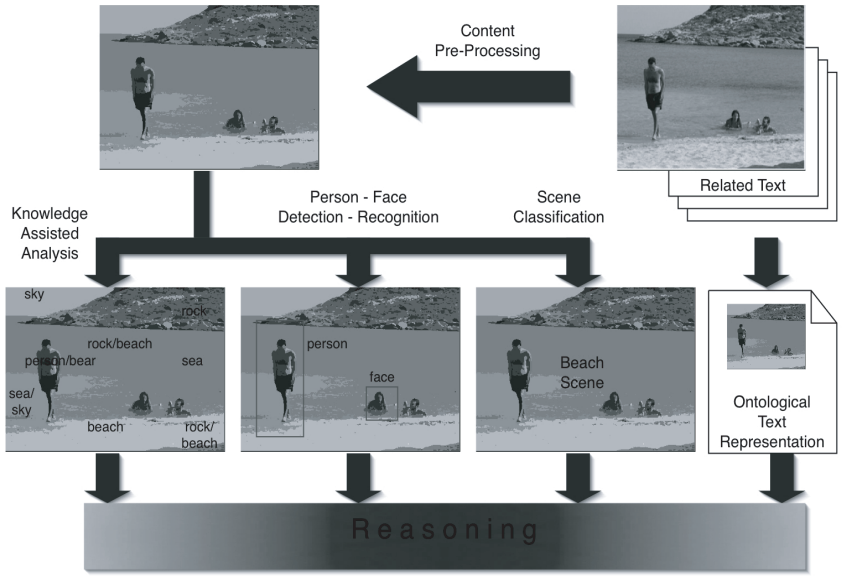
Within aceMedia, the automatically generated metadata can be exploited by the personal-

Figure 1. Overview of the aceMedia system



Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 2. aceMedia overall multimedia analysis and understanding architecture



ization module, which creates a model of user preferences and profiles enabling personalized search and presentation of content. The user model is dynamically updated by learning on user behavior as users interact with their content. Furthermore, semantic multimedia annotation in aceMedia is exploited in user-centered applications such as intelligent search and retrieval. aceMedia tools under development include user query interpretation, hybrid visual-semantic search, and retrieval, and improved relevance feedback. In the remainder of this article, the focus will be on the Knowledge Assisted Analysis module (KAA) of aceMedia and the supporting technologies.

Knowledge Assisted Analysis Within aceMedia

The overall architecture of the proposed knowledge-assisted analysis framework is illustrated in Figure 3. First segmentation is applied, and subsequently low-level descriptors and spatial relations are extracted for the generated image segments. Once the low-level descriptors are available, an initial set of hypotheses is generated for each image segment based on the distance between the segment's extracted descriptors and the domain concepts prototypical descriptors that are included in the knowledge base. Thereby, a set of plausible annotations (i.e., domain concepts) with corresponding degrees of confidence is produced for each segment. These graded hypotheses are then passed to the context analysis module that refines them utilizing the ad-hoc contextual knowledge, as will be described in more detail in the sequel. The refined hypotheses sets along with segment spatial relations are then passed to the genetic algorithm, which based on the provided domain concept definitions decides on the optimal semantic interpretation.

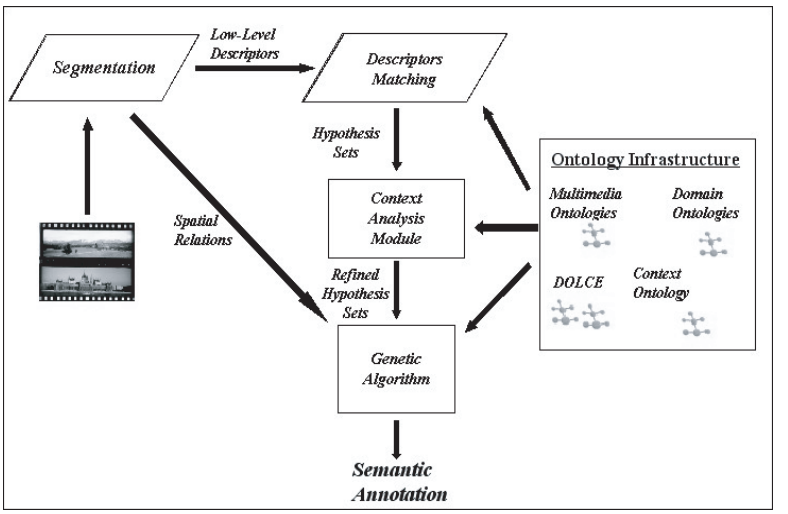
LOW-LEVEL VISUAL INFORMATION PROCESSING

Segmentation, Feature Extraction and Initial Hypotheses Generation

In order to implement the initial hypotheses generation procedure, the examined image has to be segmented into regions and suitable

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 3. Knowledge-assisted analysis framework architecture



low-level descriptions have to be extracted for every resulting segment. In the current implementation, an extension of the Recursive Shortest Spanning Tree (RSST) algorithm has been used for segmenting the image (Adamek, O'Connor, & Murphy, 2005). Considering low-level descriptions, specific descriptors of the MPEG-7 standard have been selected, namely the *Homogeneous Texture*, *Region Shape,* and *Dominant Colour* descriptors. Their extraction for each of the generated image regions is performed according to the guidelines provided by the MPEG-7 eXperimentation Model (XM) (MPEG-7 Visual Experimentation Model (XM), 2001).

In order to produce the hypotheses sets, appropriate measures need to be defined for qualitatively assessing visual similarity between the examined image segments and the defined domain concept prototypes. As MPEG-7 does not provide a standardized method for combining different descriptors distances or for estimating a single distance based on more than one descriptor, a weighted sum approach was followed, resulting in the calculation of a single scalar distance D for each hypothesis. Thereby, for each segment a similarity degree DOC is produced against each of the defined domain concepts, as follows:

$$DOC = \frac{1}{e^{mD}}$$

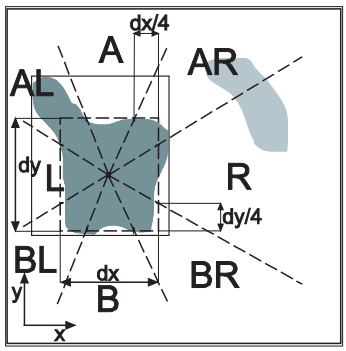
where the slope parameter m is experimentally set. The pairs of domain concept and corresponding degree of confidence that result for each segment comprise its hypotheses set.

Fuzzy Spatial Relations Extraction

Exploiting domain-specific spatial knowledge in image analysis tasks is a common practice among the object recognition community. It is generally observed that objects tend to be present in a scene within a particular spatial context and thus spatial information can substantially assist in discriminating between objects exhibiting similar visual characteristics. Among the most commonly adopted spatial relations, directional ones have received particular attention. In the present analysis framework, eight fuzzy directional relations are supported, namely *Above* (*A*), *Right* (*R*), *Below* (*B*), *Left*

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 4. Reduced MBR spatial relations definition



(*L*), *Below-Right* (*BR*), *Below-Left* (*BL*), *Above-Right* (*AR*), and *Above-Left* (*AL*).

Fuzzy directional relations extraction in the proposed analysis approach builds on the principles of projection- and angle-based methodologies (Skiadopoulos et al., 2005; Wang et al., 2004) and consists of the following steps. First, a *reduced box* is computed from the *ground* object's (the object used as reference, painted dark grey in Figure 4) Minimum Bounding Rectangle (MBR) so as to include the object in a more representative way. The computation of this *reduced box* is performed in terms of the MBR compactness value *c*, which is defined as the value of the fraction of the object's area to the area of the respective MBR: If the initially computed *c* is below a threshold *T*, the ground object's MBR is reduced repeatedly until the desired threshold is satisfied. Then, eight cone-shaped regions are formed on top of this reduced box as illustrated in Figure 4, each corresponding to one of the defined directional relations. The percentage of the *figure* object (the object whose relative position is to be estimated, painted light grey in Figure 4) pixels that are included in each of the cone-shaped regions determines the degree to which the corresponding directional relation is satisfied. After extensive experimentations, the value of threshold *T* was set equal to 0.85.

KNOWLEDGE INFRASTRUCTURE

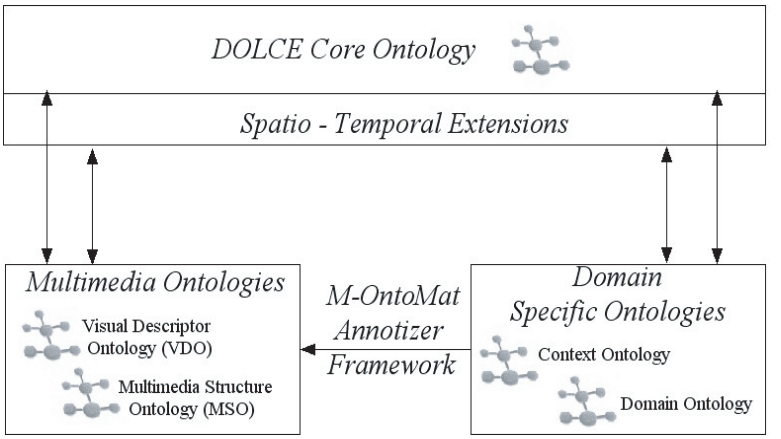
Among the possible knowledge representation formalisms, ontologies present a number of advantages (Staab & Studer, 2004). They provide a formal framework for supporting explicit, machine-processable semantics definitions, and they facilitate inference and the derivation of new knowledge based on rules and already existing knowledge. Thus, ontologies are suitable for expressing multimedia content semantics in a formal machine-processable representation that will allow automatic analysis and further processing of the extracted semantic descriptions. Following these considerations, in the aceMedia project framework, the ontology infrastructure, in a resource description framework (RDF), introduced in (Bloehdorn et al., 2005) has been used as the means for representing the knowledge components needed. As illustrated in Figure 5, it consists of a *Core Ontology* whose role is to serve as a starting point for the construction of new ontologies, a *Visual Descriptor Ontology* that contains the representations of the MPEG-7 visual descriptors, a *Multimedia Structure Ontology* that models basic multimedia entities from the MPEG-7 Multimedia Description Scheme (ISO/IEC Part:3, 2001), and *Domain Ontologies* that model the content layer of multimedia content with respect to specific real-world domains.

Core Ontology

In general, core ontologies are typically conceptualizations that contain specifications of domain-independent concepts and relations based on formal principles derived from philosophy, mathematics, linguistics, and psychology. The role of the core ontology in this overall framework is to serve as a reference point for the construction of new ontologies, to provide a reference point for comparisons among different ontological approaches, and to serve as a bridge between existing ontologies. In the presented framework, the *DOLCE* (Gangemi, Guarino, Masolo, Oltramari, & Schneider, 2002) ontology is used for this purpose. DOLCE was explicitly designed as a core ontology, is

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 5. Knowledge infrastructure



minimal in the sense that it includes only the most reusable and widely applicable upper-level categories, rigorous in terms of axiomatization, and extensively researched and documented.

Visual Descriptor Ontology

The visual descriptor ontology (VDO) (Simou, Tzouvaras, Avrithis, Stamou, & Kollias, 2005) represents the visual part of the MPEG-7 and thus, contains the representations of the set of visual descriptors used for knowledge assisted analysis. Its modelled concepts and properties describe the visual characteristics of the objects. The construction of the VDO attempted to follow the specifications of the MPEG-7 Visual Part (ISO/IEC Part:3, 2001). Because strict attachment to the MPEG-7 Visual Part became impossible, several requisite modifications were made in order to adapt the XML schema provided by MPEG-7 to an ontology and the data-type representations available in RDFS. The tree of the VDO consists of four main concepts, which are VDO: Region, VDO:Feature, VDO:VisualDescriptor, and VDO:Metaconcepts. None of these concepts is included in the XML schema defined MPEG-7, but their need was vital in order to create a correctly defined ontology. The VDO: VisualDescriptor concept contains the visual descriptors, as these are defined by MPEG-7. The VDO:Metaconcepts concept, on the other hand, contains some additional concepts that were necessary for the VDO, but they are not clearly defined in the XML schema of MPEG-7. The remaining two concepts that were defined, VDO:Region and VDO:Feature, are also not included in the MPEG-7 specification, but their definition was necessary in order to enable the linking of visual descriptors to the actual image regions. For example, consider the VDO: VisualDescriptor concept, which consists of six subconcepts, one for each category of the MPEG-7-specified visual descriptors. These are *color*, *texture*, *shape*, *motion*, *localization*, and *basic descriptors*. Each of these subconcepts includes a number of relevant descriptors. These descriptors are defined as concepts in the VDO.

Multimedia Structure Ontology

The multimedia structure ontology (MSO) models basic multimedia entities from the MPEG-7 Multimedia Description Scheme (ISO/IEC Part:5, 2001) and mutual relations like decomposition. Within MPEG-7, multimedia content is classified into five types: image, video, audio, audiovisual, and multimedia. Each of these types has its own segment subclasses.

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

MPEG-7 provides a number of tools for describing the structure of multimedia content in time and space. The Segment DS (ISO/IEC Part:5, 2001) describes a spatial or temporal fragment of multimedia content. A number of specialized subclasses are derived from the generic Segment DS. These subclasses describe the specific types of multimedia segments, such as video segments, moving regions, still regions, and mosaics, which result from spatial, temporal, and spatiotemporal segmentation of the different multimedia content types. Multimedia resources can be segmented or decomposed into sub-segments through four types of decomposition: spatial, temporal, spatiotemporal, and media source.

Domain Ontology

A domain ontology was developed for representing the knowledge components that need to be explicitly defined under the proposed approach. This contains the semantic concepts that are of interest in the examined domain (e.g., in the beach vacation domain: Sea, Sand, Person, etc.), their prototypical low-level characteristics as well as their spatial relations.

As opposed to concepts themselves that are manually defined by domain experts, prototypical visual descriptor instances for each of the concepts of interest, which are required for the initial hypotheses generation during the matching process described in the subsection "Segmentation, Feature Extraction and Initial Hypotheses Generation," and spatial relations are extracted using a training set of images. More specifically, to populate the domain knowledge with prototypical visual descriptor instances, sample images of a training set are processed with the M-Ontomat-Annotizer tool, that allows linking domain concepts with low-level visual descriptor values (Saathoff, 2006). The values of spatial relations for the concepts of the given domain are estimated according to the following ontology population procedure:

Let $S = \{s_i, i = 1,...I\}$ denote the set of regions produced for an image by segmentation, $C = \{c_p, p = 1,...P\}$ denote the set of concepts defined in the employed domain ontology and:

$$\Pi = \{\rho_k, k = 1,...K\} = \{A, AL, AR, B, BL, BR, L, R\}$$

denote the set of supported spatial relations. Then, the degree to which s_i satisfies relation ρ_k with respect to s_j can be denoted as $I_{\rho_k}(s_i,s_j)$, where the values of function I_{ρ_k} are estimated according to the procedure of the subsection "Fuzzy Spatial Relations Extraction" and belong to [0, 1]. To populate the ontology, this function needs to be evaluated over a set of segmented images with ground truth annotations that serves as a training set. More specifically, the mean values, $I_{\rho_k mean}$, of I_{ρ_k} are estimated, for every k over all region pairs of segments assigned to objects (c_p,c_q), $p \neq q$. The calculated values are stored in the ontology. These constitute the constraints input to the spatial optimization problem which is solved by the genetic algorithm, as will be described in the subsection "Spatial Optimization."

Context Ontology

A "Fuzzified" Context Model

As found in the literature, the term *context* has many interpretations, as well as definitions (Mylonas & Avrithis, 2005), none of which is globally applicable. It is therefore very important to establish a working interpretation, in order to benefit from and contribute to multimedia analysis. The problems to be addressed include how to represent context, how to determine it, and how to use it to optimize the results of knowledge-assisted analysis. Results of the latter are highly dependent on the domain an image belongs to and thus in many cases are not sufficient for the understanding of multimedia content. The lack of contextual information (Mylonas et al., 2005) in the process is a major limitation toward a better analysis performance and together with similarities in numerous low-level characteristics of various object types results in a significant number of

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

misclassifications. We introduce a method for further improving the results of the proposed knowledge-based approach, based on a contextual ontology.

In general, it is possible to formally describe an ontology as the entire set of concepts and semantic relations between concepts within a given universe:

$$O = \{C,\{R_{c_i,c_j}\}\},\ i,j = 1,...n,\ R_{c_i,c_j}: C \times C \rightarrow \{0,1\},\ i,j = 1,...n$$

where O forms an ontology, C is the set of all possible concepts it describes and R_{c_i,c_j} the semantic relation amongst two concepts c_i,c_j. Any type of relation may be included in an ontology, however, for the problem at hand a "fuzzified," ad-hoc context ontology is introduced in order to express all relationships between participating concepts. In order for this ontology to be highly descriptive, it must contain a representative number of distinct and even diverse relations among concepts, so as to scatter information among them and meaningfully describe context. In this work we utilize a set of relations whose semantics are defined in MPEG-7 (Benitez, Zhong, Chang, & Smith, 2001), namely: PartOf (*P*), SpecializationOf (*Sp*), PropertyOf (*Pr*), inContextOf (*Ct*), Location (*Loc*), InstrumentOf (*Ins*), and PatientOf (*Pat*).

However, when modelling real-life information governed by uncertainty and fuzziness, only *fuzzy* relations can handle such issues. In fact, the previously encountered relations can be modelled as fuzzy relations. Thus, in order to extract and use the desired ontological context, we define it by means of fuzzy ontological relations:

$$O_F = \{C,\{r_{c_i,c_j}\}\},\ i,j = 1,...n$$

where O_F forms a domain-specific "fuzzified" ontology, C is the set of all possible concepts it describes, $r_{c_i,c_j} = F(R_{c_i,c_j}): C \times C \rightarrow [0,1]$, $R_{c_i,c_j}: C \times C \rightarrow \{0,1\}$, $i,j = 1,...n$ denotes a fuzzy ontological relation amongst two concepts c_i,c_j and R_{c_i,c_j} is a crisp semantic relation amongst the two concepts. We shall use this "fuzzified" definition of the knowledge model throughout this article.

Contextual Knowledge Representation and Ontological Relations

The proposed contextual ontology model is able to represent any type of fuzzy relation between concepts $F(R_{c_i,c_j}) = r_{c_i,c_j}$. All relations between concepts are contained within an RDF-based representation, forming the overall contextual knowledge. Describing the accompanying degree of confidence is carried out using reification (W3C, RDF Reification, 2004) (i.e., by making a statement about the statement, which contains the degree information). Reification was used in order to achieve the desired expressiveness, whereas representing fuzziness with reified statements is an acceptable way, since the reified statement should not be asserted automatically. For instance, having a statement, such as *Car inContextOf MotorsportScene* and a degree of confidence of 0.85 for this statement, does obviously not entail, that a car is always in the context of a motorsports scene.

More specifically, let us select one fuzzy relation, namely the *partOf* relation *P*, which is a fuzzy taxonomic relation on the set of concepts. $P(a,b) > 0$ means that *b* is a part of *a*. For example *a* could be a *boat* and *b* could be a *sail*. An example of its formal representation is presented in Figure 6.

The proposed model can be seen as a graph in which every node represents a concept and each edge between two nodes a contextual relation between the respective concepts. Additionally each edge has a corresponding degree of confidence that represents fuzziness existing within the context model. Non-existing edges are implying non-existing relations (i.e., relations with zero confidence values are omitted).

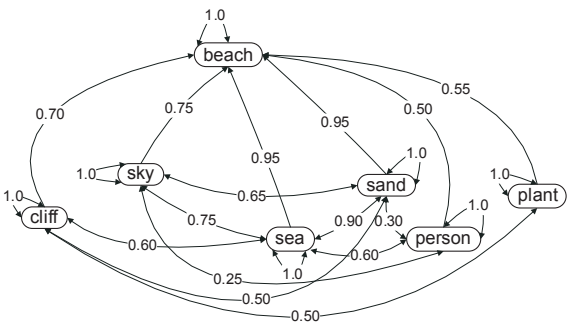
Finally, another important point to consider is the fact that each concept has a different probability to appear in the scene. A flat context model (i.e., relating concepts only to the

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 6. Reified RDF/XML representation of fuzzy part of relation

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:context="&dom;"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
    <rdf:Description rdf:about="#partOf">
        <rdfs:domain>
            <rdf:Description rdf:about="http://www.w3.org/1999/02/22-rdf-syntax-ns#Statement"/>
        </rdfs:domain>
        <rdfs:range>
            <rdf:Description rdf:about="http://www.w3.org/2001/XMLSchema#float"/>
        </rdfs:range>
    </rdf:Description>
    <rdf:Description rdf:about="#relation1">
        <rdf:subject rdf:resource="&dom;sail"/>
        <rdf:predicate rdf:resource="&dom;partOf"/>
        <rdf:object> rdf:resource="&dom;boat"</rdf:object>
        <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Statement"/>
        <context:partOf
rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.85</context:partOf>
    </rdf:Description>
</rdf:RDF>
```

Figure 7. locationOf ontology fragment from the beach holidays domain



respective scene type) would not be sufficient in this case. We model a more detailed graph where ideally concepts are all related to each other, implying that the graph relations used are in fact transitive. As observed in Figure 7, every concept participating in the contextualized ontology has at least one link to the root element. Additional degrees of confidence exist between any possible connections of nodes in the graph, whereas the root beach element could be related either directly or indirectly with any other concept. This results to the notion of *context relevance*, described in greater detail in the following section of this work.

CONTEXT AND SPATIAL OPTIMIZATION

Context Optimization

Once contextual knowledge structure is defined and corresponding representation is implemented, a context-based confidence value readjustment (CCVR) algorithm is introduced to aid in the field of multimedia

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

analysis. Our contextualization approach acts as a post-processing step on top of the initial set of hypotheses that re-estimates the initial labels degree of confidence for each image segment. In the process, it utilizes contextual information residing in the constructed context ontology and passes the optimized results to the genetic algorithm. We exploit context, constructed by a semantically meaningful combination of the previously selected fuzzy relations. More specifically, each *label* is related to a specific *concept* c_k of the application domain ontology and stored together with its relationship degrees to any other related concept. To tackle cases that more than one concept is related to multiple concepts, we introduce the term *context relevance* $cr_{dm}(c_k)$ which refers to the overall relevance of concept c_k to the *root element* of the domain *dm*. An exhaustive approach is followed considering all possible routes in the graph, with respect to the fact that all routes between concepts are reciprocal.

Estimation of each concept's context relevance is derived from two sources:

1. *Direct relationships* of the concept with other concepts.
2. *Indirect relationships*, utilizing a suitable distance metric operator.

Let us present a simplified but illustrative example (Figure 8) derived from the beach holidays contextualized ontology part, presented in Figure 7, assuming that the only available concepts were c_{beach}, c_{sea}, c_{sand}, and c_{person}. Let concept c_{sea} be related to concepts c_{beach}, c_{sky}, and c_{sand} directly with: $r_{c_{sea},c_{beach}} = 0.95$, $r_{c_{sea},c_{sky}} = 0.75$ and $r_{c_{sea},c_{sand}} = 0.90$, while concept c_{sky} is related to concept c_{beach} with $r_{c_{sky},c_{beach}} = 0.75$ and to concept c_{sand} with $r_{c_{sky},c_{sand}} = 0.65$ and concept c_{sand} is additionally directly related to concept c_{beach} with $r_{c_{sand},c_{beach}} = 0.95$. Given the semantic perspective on the correlation between any two concepts, we select the *max* operator as the appropriate distance metric operator. Then, we calculate the value for $cr_{beach}(c_{sky})$ as follows:

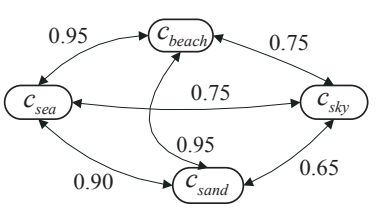
$$
\begin{aligned}
&cr_{beach}(c_{sky}) \\
&= \max\{r_{c_{sky},c_{beach}},\ r_{c_{sky},c_{sea}} \times r_{c_{sea},c_{beach}},\ r_{c_{sky},c_{sand}} \times r_{c_{sand},c_{beach}},\ r_{c_{sky},c_{sand}} \times r_{c_{sand},c_{sea}} \times r_{c_{sea},c_{beach}},\ r_{c_{sky},c_{sea}} \times r_{c_{sea},c_{sand}},\ r_{c_{sand},c_{beach}}\} \\
&= \max\{0.75,\ 0.7125,\ 0.6175,\ 0.55575,\ 0.64125\} \\
&= 0.75
\end{aligned}
$$

In this case, we observe that the direct relationship between the two concepts dominates the context relevance value for concept *sky*. A similar approach is followed for every concept participating in the context ontology.

After estimating each concept's context relevance value and according to the CCVR algorithm described in Mylonas et al. (2006), we identify the optimal normalization parameter for the domain and define the minimum considerable value of an initial degree of confidence. For each label accompanied by a degree of confidence higher than this value, we examine the supplied domain ontology and identify the concept in the domain that is related to it. Then for each identified concept we obtain the particular contextual information in the form of its relations to the set of any other concepts and calculate the new degree of confidence for the label associated to the region, based on the normalization parameter and the context's relevance value. In the case a concept is related to additional concepts apart from the root element of the ontology, an intermediate aggregation step is applied to calculate the concept's context relevance value, as already explained.

Key points in this approach are the identification of the intra-concepts relation-

Figure 8. Simplified context ontology graph



Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

ships between all concepts, the definition of a meaningful normalization parameter and the identification of the optimal initialization value for the initial confidence values. When re-evaluating these values, the ideal normalization parameter is always defined with respect to the particular domain of knowledge and is the one that quantifies their semantic correlation to the domain. The overall process is terminated when belief to the labelling output provided initially is not strong enough, that is, there are no more labels with an acceptable initial confidence value above the specified initialization value. The result of this contextualization step is the meaningful readjustment of the initial degrees of confidence accompanying each image segment, increasing the efficiency and robustness of the proposed hybrid semantic image analysis methodology and providing optimized input to the genetic algorithm, as described in the next section.

Spatial Optimization

As outlined in "System Overview," after the initial set of hypotheses is generated based solely on visual features and these are refined using context, a genetic algorithm (GA) is introduced to decide on the optimal image interpretation using the fuzzy spatial relations that have been computed for every pair of image segments. The GA is employed to solve a global optimization problem, while exploiting the available domain spatial knowledge, and thus overcoming the inherent visual information ambiguity. Spatial knowledge is obtained as described in the subsection "Domain Ontology" and the resulting learnt fuzzy spatial relations serve as constraints denoting the allowed domain objects spatial topology.

Fitness Function

The proposed optimization process uses as input the context-refined hypotheses sets (as already described in the subsection "Context Optimization"), the fuzzy spatial relations extracted between the examined image segments, and the spatial-related domain knowledge as produced by the particular training process. Under the proposed approach, each chromosome represents a possible solution. Consequently, the number of the genes comprising each chromosome equals the number I of the segments s_i produced by the segmentation algorithm and each gene assigns a defined domain concept to an image segment.

An appropriate *fitness function* is introduced to provide a quantitative measure of each solution's fitness (i.e., to determine the degree to which each interpretation is plausible):

$$f(CR) = \lambda \cdot FS_{norm} + (1 - \lambda) \cdot SC_{norm}$$

where CR denotes a particular chromosome, FS_{norm} refers to the degree of low-level descriptors matching, and SC_{norm} stands for the degree of consistency with respect to the provided spatial domain knowledge. The variable λ is introduced to adjust the degree to which visual features matching and spatial relations consistency should affect the final outcome.

The value of FS_{norm} is computed as follows:

$$FS_{norm} = \frac{\sum_{i=1}^{N} I_M\left(g_{ip}\right) - I_{\min}}{I_{\max} - I_{\min}}$$

where

$$I_M(g_{ip}) \equiv DOC_{ip}$$

denotes the degree to which the visual descriptors extracted for segment s_i match the ones of concept c_p, where g_{ip} represents the particular assignment of c_p to s_i. Thus, $I_M(g_{ip})$ gives the degree of confidence, DOC_{ip} (as defined in the subsection "Segmentation, Feature Extraction and Initial Hypotheses Generation"), associated with each hypothesis. $I_{\min} = \sum_{i=1}^{N} \min_p I_M\left(g_{ip}\right)$ is the sum of the minimum degrees of confidence assigned to each region hypotheses set and $I_{\max} = \sum_{i=1}^{N} \max_p I_M\left(g_{ip}\right)$ is the sum of the maximum degrees of confidence values respectively. For the computation of SC_{norm} the approach described in the following subsection is followed.

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Spatial Constraints Verification

Estimating the degree to which the spatial constraints between two objects to segment mappings g_{ip}, g_{jq} are satisfied is a prerequisite for exploiting spatial information in the analysis procedure. In this work, this degree of satisfaction is expressed by the function $I_S(g_{ip}, g_{jq})$. $I_S(g_{ip}, g_{jq})$ is defined with the help of a normalized Euclidean distance $d(g_{ip}, g_{jq})$, which is calculated according to the following equation:

$$d\left(g_{ip}, g_{jq}\right) = \frac{\sqrt{\sum_{k=1}^{8}\left(I_{\rho_k mean}\left(c_p, c_q\right) - I_{\rho_k}\left(s_i, s_j\right)\right)^2}}{\sqrt{8}}$$

where $I_{\rho_k mean}$ is part of the knowledge infrastructure, as discussed in the section "Knowledge Infrastructure," $I_{\rho_k}(s_i, s_j)$ denotes the degree to which spatial relation ρ_k is verified for a certain pair of segments s_i, s_j of the examined image and c_p, c_q denote the domain defined concepts assigned to them respectively. Distance $d(g_{ip}, g_{jq})$ receives values in the interval [0, 1]. The function $I_S(g_{ip}, g_{jq})$ is then defined as:

$$I_S(g_{ip}, g_{jq}) = 1 - d(g_{ip}, g_{jq})$$

and takes values in the interval [0, 1] as well, where 1 denotes an allowable relation and 0 denotes an unacceptable one. Using this, the values of SC_{norm} is computed according to the following equation:

$$SC_{norm} = \frac{\sum_{l=1}^{W} I_{S_l}(g_{ij}, g_{pq})}{W}$$

where W denotes the number of the constraints that had to be examined.

Implementation Issues

To implement the previously described optimization process, a population of 200 chromosomes is employed, and it is initialized with respect to the input set of hypotheses. After the population initialization, new generations are iteratively produced until the optimal solution is reached. Each generation results from the current one through the application of the following operators.

- **Selection:** A pair of chromosomes from the current generation are selected to serve as parents for the next generation. In the proposed framework, the Tournament Selection Operator (Goldberg & Deb, 1991) with replacement, is used.
- **Crossover:** Two selected chromosomes serve as parents for the computation of two new offsprings. Uniform crossover with probability of 0.7 is used.
- **Mutation:** Every gene of the processed offspring chromosome is likely to be mutated with probability of 0.008. If mutation occurs for a particular gene, then its corresponding value is modified, while keeping unchanged the degree of confidence.

Parameter λ regulating the relative weights of low-level descriptor matching and spatial context consistency was set to 0.35 after experimentation. The resulting weight of SC_{norm}, points out the importance of spatial context in the optimization process.

To ensure that chromosomes with high fitness will contribute to the next generation, the overlapping populations approach was adopted. More specifically, assuming a population of m chromosomes, m_s chromosomes are selected according to the employed selection method, and by application of the crossover and mutation operators, m_s new chromosomes are produced. Upon the resulting $m + m_s$ chromosomes, the selection operator is applied once again in order to select the m chromosomes that will comprise the new generation. After experimentation, it was shown that choosing $m_s = 0.4 \cdot m$ resulted in higher performance and faster convergence. The previous iterative procedure continues until the diversity of the current generation is equal to/less than 0.001 or the number of generations exceeds 50.

EXPERIMENTAL RESULTS

In this section, we present experimental results from testing the proposed approach in the domains of beach and mountain vacation images. First, two individual domain ontologies

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

were developed to represent the domain concepts of interest and their spatial relations. For the case of the beach vacation domain, under the current implementation, six concepts, namely *Sky*, *Sea*, *Sand*, *Plant*, *Cliff*, and *Person*, have been defined. On the other hand, seven concepts, namely *Rock*, *Snow*, *Ground*, *Vegetation*, *Sky*, *Person*, and *Water*, have been defined for the case of the mountain vacation domain.

To acquire the visual descriptors prototypes and the membership values for the spatial relations, a training set of 200 images was assembled (100 for every domain), using a variety of beach/mountain vacations images, and manually annotated according to the domain ontology. Subsequently, segmentation was performed as described earlier, and the *Dominant Colour*, the *Homogeneous Texture* and the *Region Shape* descriptors, i.e., the currently supported descriptors, of the annotated segments were extracted. Approximately 10 prototype descriptor instances resulted for each of the defined domain concepts after the elimination of the redundant ones, i.e., of prototypes almost identical to each other that do not offer any additional discriminative power. Additionally, for each pair of segments the degree to which each spatial relation is satisfied was estimated and thus, following the procedure described earlier for each possible combination of the defined domain concepts, the domain ontology spatial relations were enhanced with fuzzy degrees.

After building the domain knowledge, semantic annotation of images can be performed following the proposed approach. For each of the examined images, the steps described in the

Table 1. Numerical evaluation for the beach vacation domain

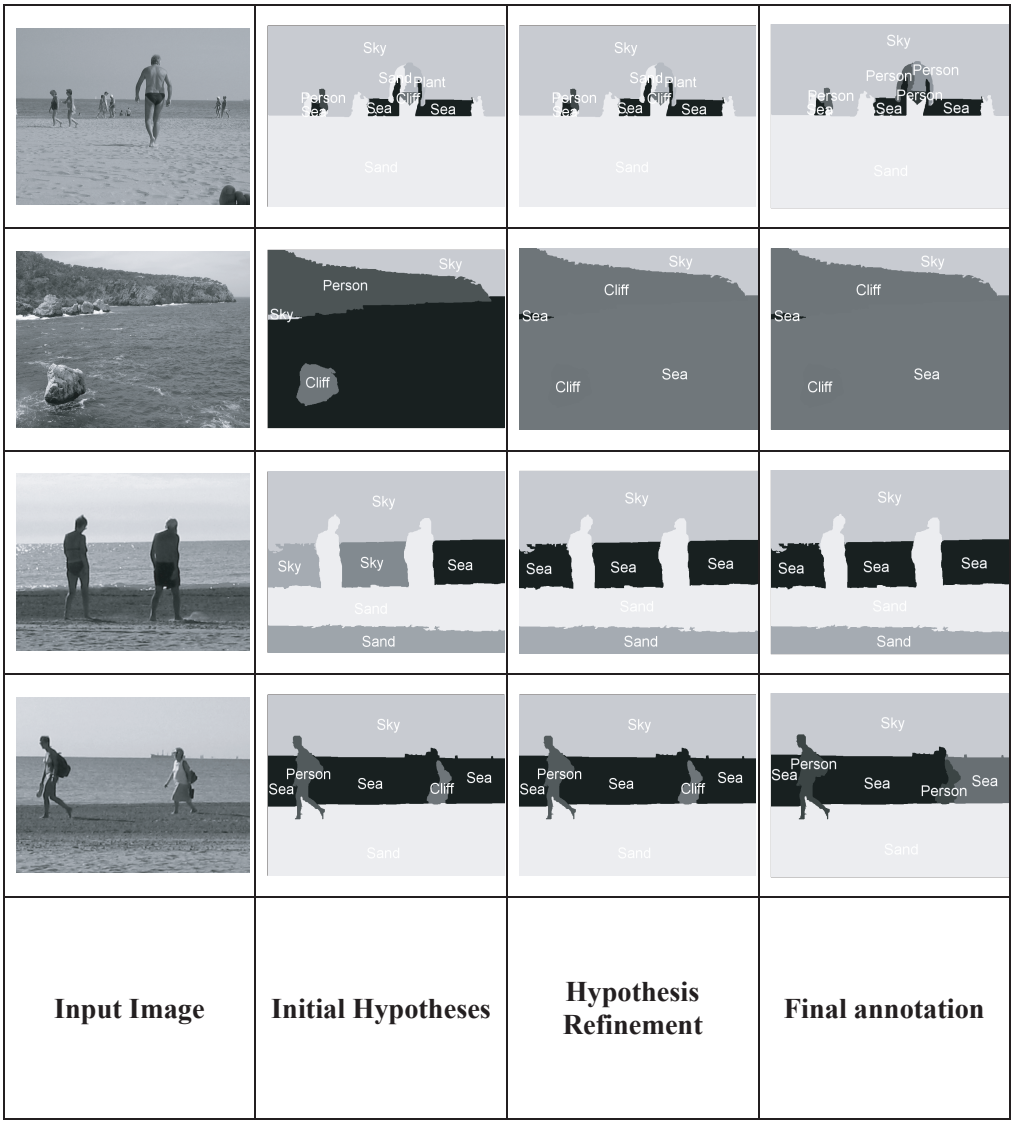
	Initial hypothesis		Hypothesis refinement		Final interpretation	
object	precision	recall	precision	recall	precision	recall
Sky	83.33%	94.74%	92.78%	94.74%	95.79%	92.86%
Sea	93.55%	87.00%	90.95%	95.50%	94.50%	90.00%
Cliff	51.92%	65.85%	59.02%	87.81%	82.93%	69.39%
Plant	17.24%	50.00%	23.53%	40.00%	60.00%	33.33%
Sand	82.69%	94.51%	89.58%	94.51%	96.70%	95.65%
Person	97.03%	71.02%	98.99%	71.02%	81.16%	99.12%
Accuracy	82.76%		87.07%		89.66%	

Table 2. Numerical evaluation for the mountain vacation domain

	Initial Hypothesis		Hypothesis Refinement		Final Interpretation	
object	precision	recall	precision	recall	precision	recall
Rock	26.67%	28.57%	40.00%	28.57%	53.33%	57.14%
Snow	75.00%	60.00%	75.00%	60.00%	60.00%	60.00%
Ground	12.50%	50.00%	14.29%	50.00%	98.20%	99.10%
Vegetation	87.00%	88.78%	85.32%	94.90%	90.00%	91.84%
Sky	93.85%	85.92%	95.31%	85.92%	95.71%	94.37%
Person	37.50%	33.33%	33.33%	22.22%	50.00%	55.56%
Water	60.00%	60.00%	60.00%	60.00%	100.00%	60.00%
Accuracy	79.02%		81.46%		86.83%	

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 9. Experimental results for the beach vacation domain

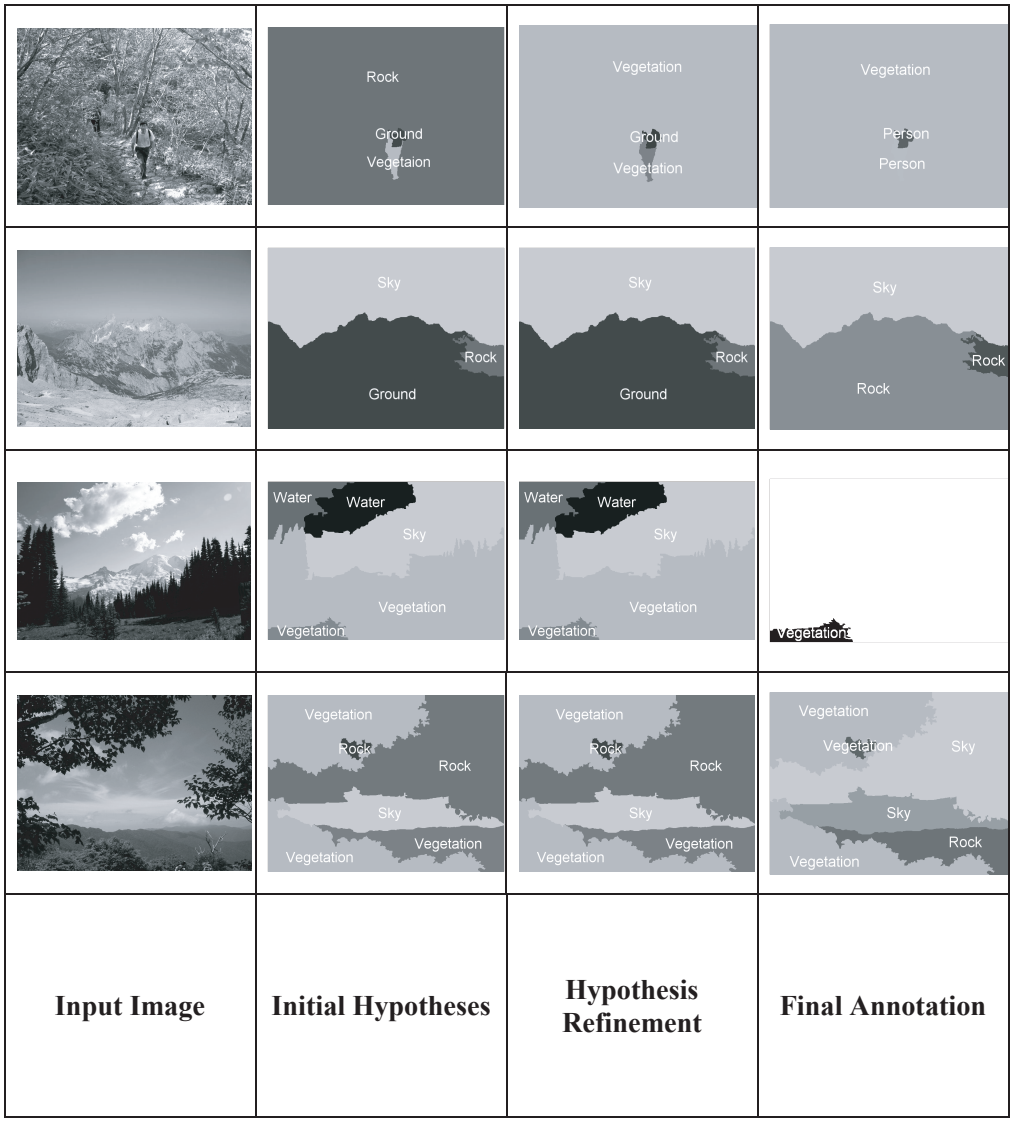


low-level visual information processing section (i.e., segmentation, descriptors extraction, and spatial relations extraction, are performed at first). Then, based on the prototype descriptor instances, initial hypotheses are generated for the examined image segments following the matching approach described in the subsection "Segmentation, Feature Extraction and Initial Hypotheses Generation," which are in turn refined through the application of the context analysis presented in "Context and Spatial Optimization." Finally, the updated graded hypotheses along with the extracted spatial relations are passed to the genetic algorithm that determines the final image interpretation.

In Tables 1-2, quantitative performance measures are given in terms of precision and recall for the two examined domains. It must

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Figure 10. Experimental results for the mountain vacation domain



be noted that for the numerical evaluation, any object present in the examined test set images that was not included in the domain ontologies was not taken into account. In Figures 9-10, indicative results are given showing the input image and the annotations resulting from the application of the genetic algorithm on the initial hypotheses and on the hypotheses refined by the context. As illustrated, the proposed system achieves satisfactory results that are further improved through the exploitation of contextual knowledge. Thereby, the use of a genetic algorithm to treat image interpretation as an optimization problem is justified, as well as the added value entailed by the introduction and utilization of context into the analysis and interpretation chain.

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

CONCLUSION

In this article, the aceMedia approach to semantic image analysis was presented. This is formulated as an optimization problem that couples ontologies with a genetic algorithm. The employed knowledge considers both high- and low-level information, represented using an ontology paradigm. The employed high-level knowledge includes the general domain knowledge in terms of concepts of interest and their spatial relations as well as contextual knowledge in form of fuzzy ontological relations, whereas low-level knowledge consists of low-level visual descriptors required for the analysis process. Following such an approach, images from different domains can be semantically annotated as long as the knowledge based is appropriately populated. The use of ontologies, due to the well-defined semantics that they provide, enables as well the application of inference services on top of the defined conceptualization that can lead to further enhanced annotations that can be inferred based on spatial reasoning.

ACKNOWLEDGMENT

The work presented in this article was partially supported by the European Commission under contract FP6-001765 aceMedia, FP6-027026 K-Space and FP6-507482 Knowledge-Web.

REFERENCES

aceMedia. (n.d.). *Integrating knowledge, semantics, and content, for user centered intelligent media services*. The aceMedia Project. Retrieved from http://www.acemedia.org

Adamek, T., O'Connor, N., & Murphy, N. (2005). Region-based segmentation of images using syntactic visual features. In *Proceedings of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Montreux, Switzerland.

Al-Khatib, W., Day, Y. F., Ghafoor, A., & Berra, P. B. (1999). Semantic modeling and knowledge representation in multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, *11*(1).

Assfalg, J., Berlini, M., Del Bimbo, A., Nunziat, W., & Pala, P. (2005). Soccer highlights detection and recognition using HMMs. *IEEE International Conference on Multimedia & Expo (ICME)* (pp. 825-828).

Athanasiadis, T., Tzouvaras, V., Petridis, K., Precioso, F., Avrithis, Y., & Kompatsiaris, I. (2005). Using a multimedia ontology infrastructure for semantic annotation of multimedia content. In *Proceedings of SemAnnot '05*, Galway, Ireland.

Benitez, A., Zhong, D., Chang, S., & Smith, J. (2001). MPEG-7 MDS content description tools and applications. In *Proceedings of International Conference on Computer Analysis of Images and Patterns (CAIP)*, Warsaw, Poland.

Bloehdorn, S., Petridis, K., Saathoff, C., Simou, N., Tzouvaras, V., Avrithis, Y., et al. (2005). Semantic annotation of images and videos for multimedia analysis. In *Proceedings of the 2nd European Semantic Web Conference, ESWC*, Heraklion, Greece.

Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V. K., & Strintzis, M. G. (2005). Knowledge-assisted semantic video object detection. *IEEE Transactions, CSVT, Special Issue on Analysis and Understanding for Video Adaptation*, *15*(10), 1210-1224.

Edmonds, B. (1999). The pragmatic roots of context. In *Proceedings of the 2nd International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT-99), LNAI 1688* (pp. 119-132). Berlin: Springer.

Gangemi, A., Guarino, N., Masolo, C. Oltramari, A., & Schneider, L. (2002). Sweetening ontologies with DOLCE in Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web. In *Proceedings of the 13th International Conference on Knowledge Acquisition, Modeling, and Management, EKAW, LNCS 2473*, Siguenza, Spain.

Goldberg, D., & Deb, K. (1991). A compara-

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

tive analysis of selection schemes used in genetic algorithms. In G. Rawlins (Ed.), *Foundations of genetic algorithms* (pp. 69-93).

Hollink, L., Little, S., & Hunter, J. (2005). Evaluating the application of semantic inferencing rules to image annotation. In *Proceedings of the 3rd International Conference on Knowledge Capture (K-CAP05)*, Banff, Canada.

Hollink, L., Nguyen, G., Schreiber, G., Wielemaker, J., Wielinga, B., & Worring, M. (2004). Adding spatial semantics to image annotations. In *Proceedings of International Workshop on Knowledge Markup and Semantic Annotation (ISWC)*.

ISO/IEC 15938-5. (2001, March). FCD information technology—Multimedia content description interface—Part 5: Multimedia description schemes. Singapore.

ISO/IEC 15938-3. (2001, March). FCD information technology—Multimedia content description interface—Part 3: Visual. Singapore.

Klir, G., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic, theory, and applications*. NJ: Prentice Hall.

Millet, C., Bloch, I., Hede, P., & Moellic, P. A. (2005). Using relative spatial relationships to improve individual region recognition. In *Proceedings of EWIMT*, London.

MPEG-7 Visual Experimentation Model (XM). (2001). Version 10.0, ISO/IEC/JTC1/SC29/WG11, Doc. N4062.

Mylonas, P., Athanasiadis, T., & Avrithis, Y. (2006). Improving image analysis using a contextual approach. In *Proceedings of International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Seoul, Korea.

Mylonas, P., & Avrithis, Y. (2005). Context modeling for multimedia analysis and use. In *Proceedings of 5th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT)*, Paris, France.

Saathoff, C., Petridis, K., Anastasopoulos, D., Timmermann, N., Kompatsiaris I., & Staab, S. (2006). M-OntoMat-Annotizer: Linking ontologies with multimedia low-level features for automatic image annotation. In *Demos and Posters of the 3rd European Semantic Web Conference (ESWC)*, Budva, Montenegro.

Simou, N., Tzouvaras, V., Avrithis, Y., Stamou, G., & Kollias, S. (2005). A visual descriptor ontology for multimedia reasoning. In *Proceedings of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, Montreux, Switzerland.

Skiadopoulos, S., Giannoukos, C., Sarkas, N., Vassiliadis, P., Sellis, T., & Koubarakis, M. (2005). 2D topological and direction relations in the world of minimum bounding circles. *IEEE Transactions on Knowledge and Data Engineering*, *17*(12), 1610-1623.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1349-1380.

Staab, S., & Studer, R. (2004). *Handbook on ontologies, international handbooks on information systems*. Heidelberg: Springer-Verlag.

Wang, Y., Makedon, F., Ford, J., Shen, L., & Golding, D. (2004). *Generating fuzzy semantic metadata describing spatial relations from images using the R-Histogram*. Tucson, AZ: JCDL.

W3C. (2004). *RDF reification*. Retrieved from http://www.w3.org/TR/rdf-schema/

Zhang, L., Lin, F., & Zhang, B. (2001). Support vector machine learning for image retrieval. In *Proceedings of the International Conference on Image Processing* (Vol. 2, pp. 721-724).

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Zhao, J., Shimazu, Y., Ohta, K., Hayasaka, R., & Matsushita, Y. (1996). An outstandingness oriented image segmentation and its applications. In *Proceedings of the International Symposium on Signal Processing and its Applications*.

END NOTE

[1] http://www.acemedia.org

Georgios Th. Papadopoulos was born in Thessaloniki, Greece in 1982. He received a Diploma degree in electrical and computer engineering from Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece in 2005. Currently he is pursuing his PhD at AUTH and is a postgraduate research fellow with the Informatics and Telematics Institute (ITI)/Centre for Research and Technology Hellas (CERTH), Thessaloniki, Greece. His research interests include still image segmentation, knowledge-assisted multimedia analysis, content-based and semantic multimedia indexing and retrieval, information extraction from multimedia, multimodal analysis and adaptive learning techniques. He is a member of the Technical Chamber of Greece.

Phivos Mylonas, MSc (computer science), is currently a researcher at the Image, Video and Multimedia Laboratory, School of Electrical and Computer Engineering, Department of Computer Science of the National Technical University of Athens, Greece. He obtained his Diploma in electrical and computer engineering from the National Technical University of Athens (NTUA) in 2001, his Master of Science in advanced information systems from the National & Kapodestrian University of Athens (UoA) in 2003, and is currently pursuing his PhD at the former university. His research interests lie in the areas of content-based information retrieval, visual context representation and analysis, knowledge-assisted multimedia analysis, issues related to personalization and user profiling and fuzzy ontological knowledge. He has published five international journals and book chapters, he is the author of 16 papers in international conferences and workshops, and a reviewer for Multimedia Tools and Applications *and* IEEE Transactions on Circuits and Systems for Video Technology *journals. He is an IEEE member since 1999, an ACM member since 2001, a member of the Technical Chamber of Greece since 2001, and a member of the Hellenic Association of Mechanical & Electrical Engineers since 2002.*

Dr. Vasileios Mezaris received the Diploma degree and PhD in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2001 and 2005, respectively. He is a postdoctoral research fellow with the Informatics and Telematics Institute/Centre for Research and Technology Hellas, Thessaloniki, Greece. Prior to this, he was a postgraduate research fellow with the Informatics and Telematics Institute/Centre for Research and Technology Hellas, Thessaloniki, Greece, and a teaching assistant with the Electrical and Computer Engineering Department of the Aristotle University of Thessaloniki, Greece. His research interests include still image segmentation, video segmentation and object tracking, multimedia standards, knowledge-assisted multimedia analysis, knowledge extraction from multimedia, content-based and semantic indexing and retrieval. He is a member of the IEEE and the Technical Chamber of Greece.

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Dr. Yannis Avrithis was born in Athens, Greece in 1970. He received the Diploma degree in electrical and computer engineering (ECE) from the National Technical University of Athens (NTUA) in 1993, an MSc in communications and signal processing (with distinction) from the Department if Electrical and Electronic Engineering of Imperial College of Science, Technology and Medicine, University of London, in 1994, and a PhD in ECE from NTUA in 2001. He is currently a senior researcher at the Image, Video and Multimedia Systems Laboratory of the ECE School of NTUA, conducting research in the area of semantic image and video analysis, and coordinating R&D activities in national and European projects. His research interests include spatiotemporal image/video segmentation and interpretation, knowledge-assisted multimedia analysis, content-based and semantic indexing and retrieval, video summarization, automatic and semi-automatic multimedia annotation, personalization, and multimedia databases. He has been involved in 13 European and nine national R&D projects, and has published 23 articles in international journals, books and standards, and 50 in conferences and workshops in areas relating to his research interests. He has contributed to the organization 13 international conferences and workshops, and is a reviewer for 15 conferences and 13 scientific journals. He is an IEEE member, and a member of ACM, EURASIP and the Technical Chamber of Greece.

Dr. Yiannis Kompatsiaris received the Diploma degree in electrical engineering and a PhD in 3-D model based image sequence coding from Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece in 1996 and 2001, respectively. He is a senior researcher with the Informatics and Telematics Institute, Thessaloniki and currently he is leading the Multimedia Knowledge Group. His research interests include multimedia content processing, multimodal techniques, multimedia and the Semantic Web, multimedia ontologies, knowledge-based, context-aware inference for semantic multimedia analysis, semantic metadata representation, semantic adaptation, personalization and retrieval and MPEG-7 standards. He is the coauthor of five book chapters, 14 papers in refereed journals, and more than 50 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences. He is a member of IEEE and of the IEE VIE TAP.

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.