

Using Fuzzy DLs to Enhance Semantic Image Analysis

S. Dasiopoulou^{1,2}, I. Kompatsiaris², and M.G. Strintzis¹

¹ Information Processing Laboratory, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Greece

² Multimedia Knowledge Laboratory, Informatics and Telematics Institute, Thessaloniki, Greece

Abstract. Research in image analysis has reached a point where detectors can be learned in a generic fashion for a significant number of conceptual entities. The obtained performance however exhibits versatile behaviour, reflecting implications over the training set selection, similarities in visual manifestations of distinct conceptual entities, and appearance variations of the conceptual entities. In this paper, we investigate the use of formal semantics in order to benefit from the logical associations between the conceptual entities, and thereby alleviate part of the challenges involved in extracting semantic descriptions. More specifically, a fuzzy DL based reasoning framework is proposed for the extraction of enhanced image descriptions based on an initial set of graded annotations, generated through generic image analysis techniques. Under the proposed reasoning framework, the initial descriptions are integrated and further enriched at a semantic level, while additionally inconsistencies emanating from conflicting descriptions are resolved. Experimentation in the domain of outdoor images has shown very promising results, demonstrating the added value in terms of accuracy and completeness of the resulting content descriptions.

1 Introduction

Digital image content is omnipresent on the Web; Google posted on August 2005, a total image size of 2,187,212,422, Yahoo estimated that its index covered 1.5 billion of images at that time, while nowadays statistics show a continuous growth in these numbers (indicatively Flickr uploads amount to an average of about 3000 images per minute). Given such numbers, the availability of machine processable semantic descriptions for this content becomes a key factor for the realisation of applications of practical interest, perpetuating the challenge of what constitutes the multimedia community holy grail, i.e. the *semantic gap* between representations that can be automatically extracted and the underlying meaning [1].

In the late 1970s and early 1980s, influenced by the Artificial Intelligence (AI) paradigm, image analysis and understanding became a problem of acquiring intelligent behaviour through computational means, resulting in the first attempts

towards knowledge-directed image analysis. A period of explosive growth in approaches conditioned by knowledge followed [2, 3]: varying knowledge representation and reasoning schemes, in accordance with the contemporary AI assets, were proposed, and knowledge attempted to address all aspects involved, ranging from perceptual characteristics of the visual manifestations to control strategies. The broad and ambitious scope targeted by the use of knowledge, resulted in representations and reasoning mechanisms that exhibited high complexity and inflexibility. As a result, research shifted to modular architectures, treating separately the individual subproblems. Machine learning (ML) approaches gained increased popularity as means to compensate for the complexity related to the explicit representation of associations between perceptual features and conceptual entities. Treating however, semantics solely as embodiments of visual manifestations, bears a number of serious limitations: not all of semantics can be expressed solely with visual means, variations pertain the possible manifestations of a single semantic entity, and distinct semantic entities share similar manifestations. Adding to the aforementioned the implications of the choices related to the training process, machine learning approaches, despite having reported satisfactory performances for given datasets, tend to scale poorly as the number of considered concepts is increased or when new content is introduced.

The advent of the Semantic Web paved a new era in knowledge sharing, reuse and interoperability, by making formal semantics explicit and accessible to heterogeneous agents and applications. The image analysis community embraced the new technologies, utilising ontologies at first in order to attach meaning to the produced annotations, and subsequently as means for assisting the very extraction of the annotations [4–8]. A characteristic shared among these approaches is that uncertainty is poorly handled. In the approaches addressing the transition from perceptual features to conceptual entities, thresholds and allowed ranges regarding the values of the considered features are used, i.e. ambiguity is treated as a separate aspect from the domain semantics [5, 4, 9, 10]. On the other hand, in the approaches that focus more on the utilisation of semantics and inference for the purpose of acquiring descriptions of higher complexity, uncertainty is not taken into consideration at all [8, 11].

Acknowledging the inherent role of uncertainty in image understanding, in this paper we propose a formal reasoning framework for the extraction of semantically coherent image descriptions, extending the earlier work presented in [12]. More specifically, we investigate the recent advances in fuzzy Description Logics (DLs) extensions [13, 14], under a generic analysis context with the purpose of overcoming part of the limitations that result from the non semantic view taken by machine learning techniques. The input of the proposed reasoning framework consists of ML obtained descriptions, along with the calculated distance values, i.e. the values indicating the similarity from the training set concepts. The graded descriptions are treated as fuzzy assertions, where each concept corresponds to a fuzzy set specified by the values learned for the parameters comprising the internal ML structure, and the corresponding distance reflects the membership of the examined image (image segment) to a given concept. The ML based extracted

descriptions may refer to different levels of granularity, and no assumptions are made regarding specific implementation issues. The possibly overlapping, contradictory or complementary, input fuzzy assertions are integrated and interpreted through the utilisation of formal semantics and reasoning.

The rest of the paper is organised as follows. Section 2 discusses the task of semantic image analysis, indicating issues and requirements related to the utilisation of formal semantics and reasoning. Section 3 introduces fuzzy DLs from an image analysis perspective, exemplifying their usability and potential. In Section 4, we present the implementation details of the proposed reasoning framework, and in Section 5, the results of the evaluation. Section 6 reports on relevant work, and finally Section 7 concludes the paper.

2 Applying Reasoning in Semantic Image Analysis

The use of formal, explicit knowledge, and subsequently inference, in semantic image analysis aims to exploit background knowledge that is characteristic of the semantic entities (objects, events) of interest within a particular context. The expected value from formally representing the semantics of a domain can be roughly summarised as “assisting the extraction of descriptions by making explicit the way semantics are perceived, ensuring thus the acquisition of interpretations that match human cognition”. In practical terms, the use of knowledge translates into the elimination of semantic inconsistencies and the acquisition of descriptions whose interpretation goes beyond visual patterns.

Under these considerations, using formal languages to represent mappings between feature values and domain entities, such as in [4], where colour, texture and shape values are mapped to tumour types, or [10], where colour and shape values are mapped to natural objects, may be significant for purposes of sharing and reusing knowledge, but it does not leave much opportunities for utilising reasoning in terms of intelligence through computational means. It is not simply a matter of the limited datatype support provided by ontology languages such as RDFS³ and OWL⁴, but because of the non logical nature of the problem at hand, i.e. the estimation of distance between a given data structure that constitutes a feature model and the measurable feature values. As the image and knowledge representation communities get more familiar to each other, the role and potential of formal knowledge inference, in semantic image analysis is revised and reassessed. In the following we discuss common issues that arise in image analysis and how they relate to the application of reasoning.

2.1 Issues and Requirements

Still image analysis, depending on the application context, may refer to the extraction of descriptions at segment level, image level, or both. In the first

³ <http://www.w3.org/TR/rdf-schema/>

⁴ <http://www.w3.org/TR/owl-features/>

case, the extracted descriptions refer to notions corresponding to objects and are accompanied by localisation information. In the second case, the descriptions apart from global scene characteristics may as well indicate the presence of objects, however without providing any information about their location. In either case, low-level features are correspondingly extracted and processed, so that the associations that underly features and conceptual entities can be captured. Independently from the learning approach followed, the extracted associations reflect correspondences between visual manifestations and conceptual notions. These correspondences are neither sound, due to the similarities in appearance between distinct semantic entities, neither complete, due to the inability to capture all possible manifestations a semantic entity may have. Furthermore, as they rely solely on visual characteristics, their potential addresses strictly meaning that can be related to visual information. The aforementioned result in a number of implications which provide the context for the identification of requirements with respect to the utilisation of formal semantics and reasoning.

A first requirement refers to the ability to handle contradictory descriptions at the same level of granularity, e.g. at image (scene) or segment level. In the case of scene level descriptions, it is not adequate to choose the description with the highest plausibility, where plausibility may refer to the probability of the description being true or to the degree to which this description is true. The various scene level descriptions are usually logically related to each other, the simplest case being that of addressing descriptions of increasing detail. Consequently, the plausibility of each extracted description is intertwined to the plausibility of descriptions referring to the semantically related notions. Similar observations hold for the case of segment level descriptions, when the extracted descriptions address only the presence of concepts. When localisation information is also available, image processing issues arise as well, since apart from semantic coherency, the extracted descriptions need to be checked at the partitioning plane (or planes, as different implementations may adopt independent segmentations).

A second requirement refers to the ability to handle and resolve semantic conflicts between descriptions of different granularity. As in the previous case, the conceptual notions at object and scene levels are highly likely to be logically related. The visual information that the learning approaches rely on, cannot ensure that such relations are reflected in the learned associations. Such limitation applies to both directions. The plausibility of object level concepts may impact on the existence and plausibility of scene level concepts, e.g. the presence of mountain indicates high plausibility for a mountainous scene, and conversely. Similarly, the presence of a given scene level concept may affect the plausibility of a concept at object level. For example, a scene description of beach affects (negatively) an object level description of snow.

A third requirement involves the ability to support descriptions, at scene or segment level, whose semantics lie beyond the potential of visual characteristics, i.e. semantics that originate in logical associations between other concepts. Examples include cases of part-whole relations, temporal sequences of simpler concepts, the combined or exclusive presence of given concepts, etc. Contrary

to the aforementioned, this requirement impacts mostly on the selection of the knowledge representation scheme, as it embodies the main goal of knowledge representation and reasoning, i.e. the automatic extraction of implicit knowledge based on the explicitly available one. As in every knowledge-based application, the construction of the knowledge depends on the application context. For example, the concept of swimmer is more likely to be described as a face surrounded by sea, than a person in the sea.

Underlying the aforementioned requirements is the need to provide the means for handling uncertainty. This forms a crucial requirement, due to the inherent in image analysis ambiguity, where visual information plays the role of cues rather than evidences, allowing the detection of a concept only with an estimated plausibility. It is important to note that the semantics of the ambiguity are not unique. When analysis follows a similarity based approach, matching extracted feature values to prototypical ones, the resulting plausibility represents the distance from the prototypical values, i.e. the degree to which the examined image (segment) belongs to a given concept. On the other hand, when a Bayesian approach has been followed, the calculated values represent a different kind of ambiguity, i.e. the probability that the examined image (segment) is an instance of the examined concept. Naturally, under an analysis framework both kinds of ambiguity may be present.

In the following we present a brief introduction to the basics of fuzzy DLs and discuss how the aforementioned requirements relate to core DL inference services. Additionally, we indicate the functionalities, that once available would allow the extraction of image semantics under a fuzzy DL based reasoning framework. We note that in the following, uncertainty refers strictly to cases of machine learning based analysis, i.e. to fuzzy logic semantics. A combined fuzzy-probabilistic reasoning framework, although very interesting and promising towards a more complete confrontation of the image understanding challenge, is currently beyond the scope of this paper.

3 Fuzzy DLs and Semantic Image Analysis

Description Logics (DLs) [15] are a family of knowledge representation formalisms characterised by logically founded formal semantics and well-defined inference services. Starting from the basic notions of atomic concepts and atomic roles, arbitrary complex concepts can be described through the application of corresponding constructors (e.g., \neg , \sqcap , \sqcup). Terminological axioms (TBox) allow to capture equivalence and subsumption semantics between concepts and relations, while real world entities are modelled through concept ($a : C$) and role ($R(a, b)$) assertions (ABox). The semantics of DLs are formally defined through an interpretation I . An interpretation consists of a non-empty set Δ^I (the domain of interpretation) and an interpretation function \cdot^I , which assigns to every atomic concept A a set $A^I \subseteq \Delta^I$ and to every atomic role R a binary relation $R^I \subseteq \Delta^I \times \Delta^I$. The interpretation of complex concepts follows inductively [15].

Table 1. Example of TBox.

$$\begin{aligned}
\text{Natural} &\equiv \text{Outdoor} \sqcup \neg \text{ManMade} \\
&\quad \exists \text{ contains.Sky} \sqsubseteq \text{Outdoor} \\
\text{Beach} &\equiv \exists \text{ contains.Sea} \sqcap \exists \text{ contains.Sand} \\
&\quad \text{Beach} \sqsubseteq \text{Natural} \\
\text{Cityscape} &\equiv \exists \text{ contains.}(\text{Road} \sqcup \text{Car}) \sqcup \exists \text{ contains.Building} \\
&\quad \text{Cityscape} \sqsubseteq \text{ManMade}
\end{aligned}$$

In addition to the means for representing knowledge about concepts and assertions, DLs come with a powerful set of inference services that make explicit the knowledge implicit in the TBox and ABox. *Satisfiability*, *subsumption*, *equivalence* and *disjointness* constitute the main TBox inferences. Satisfiability allows to check for concepts that correspond to the empty set, subsumption and equivalence check whether a concept is more specific or respectively identical to another, while disjointness refers to concepts whose conjunction is the empty set. Regarding the ABox, the main inferences are *consistency*, which checks whether there exists a model that satisfies the given knowledge base, and *entailment*, which checks whether an assertion ensues for a given knowledge base.

Assuming a TBox that describes a specific domain, one can build an ABox from the analysis extracted descriptions and benefit from the inferences provided to detect inconsistencies and obtain more complete descriptions. Let us assume the TBox of Table 1, and the analysis extracted assertions $(\text{image1} : \text{Cityscape})$, $(\text{region1} : \text{Sea})$, $(\text{region2} : \text{Sand})$, $(\text{image1}, \text{region1}) : \text{contains}$, and $(\text{image1}, \text{region1}) : \text{contains}$. The following assertions entail: $(\text{image1} : \text{ManMade})$, $(\text{image1} : \text{Beach})$, and $(\text{image1} : \text{Natural})$. Furthermore an inconsistency is detected, caused by the disjointness axiom relating the concepts Natural and ManMade. However, the aforementioned apply only in the case of crisp assertions, which is not the common case in image analysis. Applying thresholds in the initial descriptions and transforming them to binary, does not overcome the problem; instead additional issues are introduced. Let us assume, the following set of initial descriptions, $(\text{region1} : \text{Sea}) \geq 0.8$, $(\text{region1} : \text{Sand}) \geq 0.9$, $(\text{image1} : \text{Cityscape}) \geq 0.9$, and that all role assertions involving the role *contains* have a degree of ≥ 1.0 . Transforming them directly to crisp, would result in the crisp assertions $(\text{region1} : \text{Sea})$, $(\text{region1} : \text{Sand})$, and $(\text{image1} : \text{Cityscape})$, which would cause an inconsistency with no clear clues about which assertions are more prevailing, as the degree information has been omitted.

In the case of a fuzzy DL language, the ABox consists of a finite set of *fuzzy assertions* of the form $a : C \bowtie n$ and $(a, b) : R \bowtie n$, where \bowtie stands for \geq , $>$, \leq , and $<$ ⁵. The semantics are provided by a fuzzy interpretation, which in

⁵ Intuitively a fuzzy assertion of the form $a : C \geq n$ means that the membership degree of the individual a to the concept C is at least equal to n

accordance to the crisp DLs case, is a pair $I = (\Delta^I, \cdot^I)$ where Δ^I is a non-empty set of objects called the domain of interpretation, and \cdot^I is a *fuzzy*, this time, interpretation function which maps: an individual a to an element $a^I \in \Delta^I$, i.e., as in the crisp case, a concept name A to a membership function $A^I : \Delta^I \rightarrow [0, 1]$, and a role name R to a membership function $R^I : \Delta^I \times \Delta^I \rightarrow [0, 1]$.

Regarding fuzzy DLs extensions, two main efforts exist currently that address formally both the semantics and the corresponding reasoning algorithms. In [14, 16], the DL language *SHIN* has been extended according to fuzzy set theory leading to the so called *f-SHIN*. The fuzzy extensions address the assertion of individuals and the extension of the language semantics. In [13], a fuzzy extension of *SHOIN(D)* is presented, which constitutes a continuation of earlier works of the authors on extending *ALC*, *SHIF*, and *SHIF(D)* to fuzzy versions [17, 18]. In addition to extending the *SHOIN(D)* semantics to *f-SHOIN(D)*, the authors present a set of interesting features: concrete domains as fuzzy sets, fuzzy modifiers such as *very* and *slightly*, and fuzziness in entailment and subsumption relations.

Additionally to the theoretic foundations for the fuzzy extensions, respective reasoning algorithms have been presented and implemented, namely the *Fuzzy Reasoning Engine* (FiRE)⁶ and the *fuzzyDL*⁷.

Continuing the previous example, let us assume that the following fuzzy assertions result from analysis: $(image1 : Cityscape) \geq 0.4$, $(image1 : Outdoor) \geq 0.82$, $(region1 : Sea) \geq 0.8$, $(region2 : Sand) \geq 0.65$, $(region3 : Sky) \geq 0.9$, and $(image1, region_i) : contains \geq 1.0$, for $i = 1, 2, 3$. Querying for the greatest lower bound for the individual *image1* with respect to the concepts *Outdoor* and *Beach* we retrieve 0.9 and 0.65 respectively. Note that *Beach* is defined as the conjunction of two existential restrictions involving the fillers *Sea* and *Sand*. Since the assertions referring to the role *contains* have a degree ≥ 1.0 , the degrees of the two existential definitions depends only on the degrees of their respective fillers. Under Zadeh's semantics, the T-norm equals the minimum of the involved degrees, i.e. $\min\{0.65, 0.8\} = 0.65$. The \exists *contains.Sky* \sqsubseteq *Outdoor* axiom, gives greatest lower bound for $(image1 : Outdoor)$ equal to 0.9, updating the explicitly given value of 0.82. As the detection of an outdoor image does not give any further information about a more specific scene description or of the objects that may be depicted, the degrees of the other assertions should not be affected. On the contrary the presence of any of the three concepts, means that it is an outdoor concept, with a plausibility greater or equal to the corresponding degree of the respective concept. The subsumption axioms between the concepts *Beach*, *Natural* and *Cityscape* capture this knowledge, ensuring an appropriate behaviour.

Furthermore, let us examine in a greater detail the implication of the fuzzy conjunction semantics appearing in axiom *Beach* $\equiv \exists$ *contains.Sea* $\sqcap \exists$ *contains.Sand*. Under a case where for one of the existential restriction fillers there can be no assertion, no entailment can be made with respect to the *Beach* concept. Assum-

⁶ <http://www.image.ece.ntua.gr/~nsimou>

⁷ <http://faure.isti.cnr.it/~straccia/software/fuzzyDL/fuzzyDL.html>

ing now an assertion of the Beach concept, it entails the presence of assertions for the Sea and Sand concepts, and with degrees greater or equal to that of the Beach assertion. One can observe that again the fuzzy DLs semantics reflect the desired behaviour. Note that in the fuzzy case, disjointness has different semantics than in the crisp case, i.e. two concepts that are disjoint raise no inconsistency as long as the T-norm (conjunction) of their degrees is not equal or greater than 0.5. Generally, in the examined image analysis context, a degree less than 0.5 indicates a rather poor match in terms of visual similarity, resulting in practically ignoring the corresponding assertions. We note however, that in case the corresponding concepts classifiers are implemented in a binary fashion, i.e. low values indicate the non presence of a concept, then such fuzzy assertions play an important role, as they entail assertions of high degrees for the negated concepts.

We observe that fuzzy DLs, through the expressivity and the inference services that they proffer, constitute a very promising technology for supporting the extraction of image semantics. The available functionalities allow one to formally describe the different kinds of logical associations that define the semantics of the concepts and roles of the respective domain of interest, and ensure the entailment of intended descriptions, the semantic based update of the degrees, and the detection of inconsistencies, in case the initial description are violating the semantic model of the domain. However, in order to utilise fuzzy DLs in semantic image analysis, there are still two main issues: the definition of a framework under which the requirements described in section 2.1 can be realised based on fuzzy DLs inference services, and the handling of inconsistencies, so that a final set of consistent descriptions can be obtained. Handling inconsistencies in DLs knowledge bases usually refers to approaches targeting revision of the terminological axioms [19, 20]. In the examined case however, the inconsistencies result from the limitations in associating semantics with visual features. Thus it is the ABox that needs to be appropriately managed. The methodology that we followed, is described in the following section, where the implementation of the proposed fuzzy DLs based reasoning framework is detailed.

4 A Fuzzy DLs-based Reasoning Framework for Semantic Image Analysis

In the previous Section we highlighted how fuzzy DLs relate to the tasks involved in supporting semantic image analysis. In this section, we present the details of the proposed fuzzy DLs based reasoning framework, which utilising the core fuzzy DLs inference services, accomplishes the requirements highlighted in Section 2.1. Summarising, the latter address three key issues: i) consistency checking and handling of assertions of the same granularity (i.e. scene and object level), ii) consistency checking and handling between assertions of different granularity, and iii) enrichment of the descriptions by means of logical entailment. In the current study, localisation information of object level descriptions is not taken into account, reducing the first task to consistency checking at scene level. Using the

fuzzyDL system for the core fuzzy DLs inferences, the proposed reasoning framework realises the extraction of image semantics as follows. First, the descriptions that apply to an image at scene level are determined through reasoning, utilising the terminological box semantics. In the sequel, based on the previously inferred scene level descriptions, inconsistencies in the initial set of scene and object level assertions are tracked and resolved, leading to a semantically meaningful description for the image. The last step refers to the enhancement of the description by making explicit assertions that result by logical entailment. The details of each step are presented follow.

Selection of scene descriptions. Due to the logical associations between concepts that refer to objects and concepts that refer to scene level notions, all the available analysis produced assertions need to be taken into account at this step. First, the hierarchy of the scene level concepts is computed, based on the respective TBox. Starting from the more specialised scene concepts, i.e. concepts that are not subsumed by other scene concepts, and moving upwards the hierarchy, the assertions of the corresponding concepts are processed with respect to fuzzy semantics. In each level of the hierarchy, the assertion, explicit or inferred, with the greatest degree prevails the assertions of disjoint concepts. The latter are stored in a list as they indicate sources of inconsistency to be addressed in the following step. In case of assertions referring to concepts that are not disjoint, all of them are preserved. Moving to the next level of the scene concepts hierarchy, the procedure is repeated, checking additionally whether the prevailing concept of the current level is subsumed by the concept(s) selected at the previous level. If a subsumption relation holds, the degree at the current level is updated accordingly so that the degree of the subsumer is greatest or equal to that of the subsumee.

In the opposite case, the concepts of the previous level are moved to the list of inconsistent concepts. To give an example, assume that the concepts Beach and Cityscape are at the same level of the hierarchy, and that the next level includes the concepts Natural and ManMade to which the former are related through respective subsumption axioms. Assuming the assertions $(image1 : Beach) \geq 0.8$, $(image1 : Cityscape) \geq 0.6$, and $(image1 : ManMade) \geq 0.9$ the Beach concept is preserved at the first step. Moving to the next level however, the concept ManMade prevails that of Natural, which means that the Beach referring assertion needs to be considered for inconsistency. The aforementioned process iterates till reaching the top level concepts of the hierarchy. Obviously, the disjointness axioms, need to be removed from the TBox on which fuzzy DLs reasoning is performed, and handled separately, in order to prevent halting the inference in case of contradictions.

Handling of inconsistency. The previous procedure results in the identification of the concepts at scene level that are inferred as valid and of the scene level concepts that constitute sources of inconsistency. The first step of the consistency handling tasks is to query for assertions that refer to object level concepts that are directly disjoint to the previously selected scene level concepts. In the existence of such assertions, if the referred concept is atomic, the assertions are

directly removed. If the referred concept is complex, the assertions that led to its inference are tracked and based on the semantics of the DLs constructors possible solutions are identified and stored. To give a simplified example, in the case of a complex concept whose definition consists in the conjunction of atomic concepts, the possible solutions equal the number of conjuncts. The next step, considers inconsistencies that arise due to disjointness axioms between scene level concepts, which includes the case of assertions referring to object level concepts that entail scene level assertions. Again, based on the terminological axioms the assertions that cause the inconsistency are tracked, and solutions are computed based on the involved constructors.

In order to enable the unhindered running of the fuzzy DL reasoner inference services, yet preserve all axioms in the TBox, based on the list of inconsistent concepts resulting during the first task, corresponding definitions of *non-concepts* are introduced, and respective subsumption axioms are added with respect to the original concepts. For example, assuming that Beach is included in the inconsistent concepts list, the axiom $\text{Beach} \sqsubseteq \text{Non-Beach}$ is added to the TBox, allowing thus the tracking of inconsistency without halting the reasoner. The alternative solutions calculated for resolving the inconsistencies are computed collectively, over the entire ABox, so that possible dependencies among them are taken into account. Checking and tracking all inconsistencies results in the general case in a set of possible solutions. In order to choose among the alternative solutions, we rank the set of solutions according to the number of assertions that need to be removed per solution, and the average of the corresponding degrees. The solution that involves the removal of the fewer assertions is eventually preferred (the average degrees are used to select between solutions of equal size).

Enrichment of descriptions. The enrichment of the descriptions by means of entailment is the most straightforward of the considered tasks. Once the scene level concepts are selected, and the assertions ensuing inconsistencies either directly or through complex definitions, are resolved, we end up with a semantically consistent set of assertions, that constitute the description of the image. Consequently, all that is left is to make explicit the assertions that are implicit in this final set. For this reason, appropriately queries are posed and the responses are included in the image description.

5 Experimental Results and Evaluation

In the previous sections, we described the reasons that motivated our investigation into a fuzzy DL based reasoning framework for supporting and enhancing semantics extraction from images. In order to assess the utilisation of formal semantics under the proposed reasoning framework, we carried out two experiments in the domain of outdoor images. An extract of the constructed Tbox is shown in Table 2. The test set consisted of 350 images, for which ground truth was manually generated at scene and object level according to the constructed TBox. In the evaluation, we compare the reliability of the image descriptions extracted through machine learning to that of the descriptions resulting after

Table 2. Extract of outdoor images TBox.

$$\begin{aligned}
& \text{Countryside_buildings} \sqsubseteq \exists\text{contains.Buildings} \sqcap \exists\text{contains.Foliage} \\
& \text{Foliage} \sqcup \text{Grass} \sqcup \text{Tree} \sqsubseteq \text{Foliage} \\
& \text{Rockside} \sqsubseteq \exists\text{contains.Cliff} \\
& \text{Roadside} \sqsubseteq \exists\text{contains.Road} \\
& \text{Coastal} \equiv \exists\text{contains.Sea} \\
& \text{Forest} \sqsubseteq \text{Landscape} \\
& \text{Forest} \sqsubseteq \exists\text{contains.Foliage} \\
& \text{Beach} \equiv \text{Coastal} \sqcap \exists\text{contains.Sand} \\
& \text{Beach} \sqsubseteq \text{Natural} \\
& \text{Cityscape} \sqsubseteq \text{ManMade} \\
& \exists\text{contains.Sky} \sqsubseteq \text{Outdoor} \\
& \exists\text{contains.Mountain} \sqsubseteq \text{Mountainous} \\
& \text{Forest} \sqcap (\text{Roadside} \sqcup \text{Countryside_buildings}) \sqsubseteq \perp \\
& \text{Roadside} \sqcap \text{Countryside_buildings} \sqsubseteq \perp \\
& \text{Landscape} \sqcap (\text{Mountainous} \sqcup \text{Coastal}) \sqsubseteq \perp \\
& \text{Natural} \sqcap \text{ManMade} \sqsubseteq \perp
\end{aligned}$$

the application of reasoning. We adopted the precision, recall, and F-measure metrics, where F-measure is defined as $2 * p * r / (p + r)$.

- **Experiment I.** In the first experiment, we employed two classifiers at scene level and two classifiers at segment level. The scene classifiers use colour and texture features following a support vector machine [21], and a randomised clustering trees approach [22] respectively. The segment level classifiers use respectively a distance based feature matching approach based on prototypical values [10], and a clustering trees approach [23]. The sets of analysis supported scene and object level concepts are respectively Outdoor, Indoor, Natural, Man-Made, Landscape, Beach, Mountainous, Beach and Building, Grass, Foliage, Cliff, Tree, Sea, Sand, Conifers, Boat, Road, Ground, Sky, Trunk, Person. The TBox includes also the concepts Coastal and Cityscape, which are to be entailed through reasoning.

In Tables 3 and 4, the evaluation metrics are given for the scene and segment level concepts respectively. For the case of scene level concepts, we treated the outcome of the analysis based on the semantics of the concepts that the classifiers supported. For example, in case the Landscape concept was detected we assumed that the concepts Natural and Outdoor were also detected, although this was not necessarily the case, i.e. the corresponding detectors had no provided a positive outcome. This accounts partially for the low impact of reasoning in the case of scene level descriptions, since the descriptions that were to be inferred through the subsumption axioms that apply to between the scene level concepts, were made explicit. In the case of the Beach and Coastal concepts, where no classifier for Coastal is included, the result of the application of reasoning becomes apparent. A second reason for the relative low affect of reasoning relates to the semantics of the scene concepts themselves. Observing the corresponding TBox

Table 3. Evaluation of analysis and reasoning performance for scene level concepts - Experiment I.

Concept	Analysis			Reasoning		
	Recall	Precision	F-M	Recall	Precision	F-M
<i>Indoor</i>	-	-	-	1.00	0.75	0.85
<i>Outdoor</i>	0.99	0.99	0.99	0.99	0.99	0.99
<i>Natural</i>	0.97	0.96	0.97	0.98	0.96	0.97
<i>ManMade</i>	0.18	0.40	0.25	0.18	0.40	0.25
<i>Cityscape</i>	0.18	0.40	0.25	0.18	0.40	0.25
<i>Landscape</i>	0.75	0.63	0.68	0.76	0.68	0.71
<i>Mountainous</i>	0.64	0.28	0.39	0.48	0.30	0.37
<i>Coastal</i>	-	-	-	0.86	0.49	0.63
<i>Beach</i>	0.89	0.26	0.40	0.90	0.31	0.47

(Table 2), one notices that the scene level concepts are in their majority either atomic concepts or concepts appearing in the left-hand of subsumption axioms, i.e. concepts that cannot be logically entailed through the existence of others.

This is not the case with respect to the object level concepts, where the impact of the reasoning is higher. As illustrated in Table 4, there are cases for which precision is increased, which correspond to concepts involved in disjointness axioms, cases where both recall and precision are improved, which correspond to complex concepts definitions or concepts that appear on the right hand side of subsumption axioms (e.g. Sea, Mountain), and cases where the performance is invariable, which involve concepts not participating in any axiom (e.g. Person) or concepts participating solely in the left hand side of subsumption axioms (e.g. Trunk). Evaluating collectively the performance of analysis amounts to 0.68, 0.49, and 0.57, for recall, precision, and F-measure. The respective values for reasoning are 0.68, 0.63, and 0.65. If we take into consideration only concepts whose semantics are affected by logical associations, the corresponding values become 0.70, 0.64, and 0.67.

Experiment II. In the second experiment, we considered a method based on the combined use of global and local information for the detection of both scene and object level descriptions. Colour, texture and shape descriptors are used, and learning is implemented using support vector machines. The sets of analysis supported scene and object level concepts are respectively Country-side_Buildings, Roadside, Rockyside, Beach and Building, Roof, Grass, Foliage, Dried_Plant, Sky, Cliff, Tree, Sea, Sand, Boat, Road, Ground, Person, Trunk, Wave. The concepts Outdoor, Natural, Coastal and Mountainous are to be supported solely through reasoning.

In the Tables 5 and 6, the evaluation metrics are presented for the case of scene and object level concepts respectively. A first observation is the improvement in terms of the descriptions completeness, i.e. the subsumption axioms between the scene level concepts allow to enhance the descriptions supported by analysis and acquire descriptions of more generic concepts such as Natural,

Table 4. Evaluation of analysis and reasoning performance for object level concepts - Experiment I.

Concept	Analysis			Reasoning		
	Recall	Precision	F-M	Recall	Precision	F-M
<i>Building</i>	0.35	0.17	0.22	0.09	0.83	0.17
<i>Grass</i>	0.06	0.40	0.10	0.01	0.94	0.05
<i>Foliage</i>	0.99	0.70	0.82	0.90	0.80	0.85
<i>Cliff</i>	0.98	0.21	0.35	0.54	0.42	0.47
<i>Tree</i>	0.22	0.65	0.33	0.18	0.58	0.27
<i>Sand</i>	0.49	0.37	0.42	0.92	0.41	0.56
<i>Sea</i>	0.72	0.46	0.56	0.88	0.49	0.63
<i>Conifers</i>	1.00	0.01	0.02	0.50	0.02	0.03
<i>Mountain</i>	0.14	0.01	0.01	0.43	0.04	0.06
<i>Boat</i>	0.10	0.40	0.16	0.10	0.50	0.17
<i>Road</i>	0.15	0.50	0.23	0.02	0.25	0.03
<i>Ground</i>	0.06	0.57	0.19	0.11	0.57	0.19
<i>Sky</i>	<i>0.93</i>	<i>0.87</i>	<i>0.89</i>	<i>0.93</i>	<i>0.87</i>	<i>0.89</i>
<i>Trunk</i>	<i>0.38</i>	<i>0.65</i>	<i>0.48</i>	<i>0.38</i>	<i>0.65</i>	<i>0.48</i>
<i>Person</i>	<i>0.49</i>	<i>0.54</i>	<i>0.52</i>	<i>0.49</i>	<i>0.54</i>	<i>0.52</i>

and Mountainous. Secondly, we observe that the application of reasoning improves in general the precision of the extracted descriptions. This is a direct outcome of the fact that there exists strong semantic association between the scene and segment level concepts semantics, i.e. there is a significant number of axioms (subsumption and disjointness ones) between them. This explains also the stronger, compared to the first experiment, affect of reasoning in the case of object level descriptions also (Table 6). The aggregated evaluation of scene and object level concepts amounts to 0.37, 0.65, and 0.47 for recall, precision, and F-measure. The respective values for reasoning are 0.77, 0.81, and 0.79. Taking into consideration only concepts which participate to axioms, the corresponding values become 0.29, 0.61, and 0.39 for the case of analysis, and 0.79, 0.82, and 0.81 for the case of reasoning, constituting hence a significant improvement.

6 Relevant Work

In the majority of relevant literature, only crisp DLs approaches have been investigated: in [6], crisp DLs are proposed for inferring descriptions whose semantics lie in logical aggregation, in [8], DLs have been extended with a rule-based approach to realise abductive inference over crisp analysis assertions, while in [11], DLs and rules have been utilised for video annotation using crisp semantics. Fuzzy DLs have been proposed in [24] for the purpose for semantic multimedia retrieval; the fuzzy annotations however are assumed to be available. Fuzzy DLs have been proposed recently in [25] and [26] for enhancing machine learning based extracted image annotations and document classification respectively;

Table 5. Evaluation of analysis and reasoning performance for scene level concepts - Experiment II.

Concept	Analysis			Reasoning		
	Recall	Precision	F-M	Recall	Precision	F-M
<i>Countryside_buildings</i>	0.30	1.0	0.46	0.60	0.86	0.71
<i>Rockyside</i>	0.68	0.70	0.69	0.68	0.79	0.74
<i>Roadside</i>	0.68	0.69	0.69	0.68	0.72	0.70
<i>Forest</i>	0.75	0.63	0.69	0.74	0.68	0.71
<i>Coastal</i>	0.85	0.67	0.75	0.86	0.72	0.78
<i>Outdoor</i>	-	-	-	0.99	1.00	0.99
<i>Natural</i>	-	-	-	0.97	1.00	0.98
<i>Mountainous</i>	-	-	-	0.67	0.80	0.74
<i>Beach</i>	-	-	-	0.45	0.76	0.57

however, neither approach addresses the problem of resolving semantic inconsistencies in the initially extracted descriptions.

7 Conclusions

In this paper, we presented a fuzzy DLs based reasoning framework with the aim to enhance the extraction of image semantics through the utilisation of formal semantics. The application of fuzzy DLs semantics allows us to formally address the uncertainty confronted in descriptions extracted through machine learning analysis. Furthermore, through the utilisation of the semantics characterising the available domain knowledge, the proposed reasoning framework addresses and resolves inconsistencies among the initial descriptions. Thereby, and free of assumptions regarding the preceding analysis, it provides the means to integrate descriptions acquired through typical image analysis into a semantically consistent, semantically enhanced annotation. The experiments, though not conclusive, have shown very promising results, indicating that the impact of reasoning is proportional to the level of semantic associations underlying the domain concepts. Future directions include the investigation of extending the framework in order to handle spatial relations semantics, and the combination with probabilistic knowledge as complementary means to handle the uncertainty in semantic image analysis.

8 Acknowledgements

This work was partially supported by the European Commission under contracts FP6-001765 aceMedia and FP6-507482 KnowledgeWeb.

Table 6. Evaluation of analysis and reasoning performance for object level concepts - Experiment II.

Concept	Analysis			Reasoning		
	Recall	Precision	F-M	Recall	Precision	F-M
<i>Building</i>	0.54	0.69	0.60	0.62	0.86	0.72
<i>Roof</i>	0.33	0.54	0.41	0.33	0.75	0.46
<i>Grass</i>	0.49	0.42	0.45	0.30	0.52	0.38
<i>Foliage</i>	0.48	0.84	0.61	0.86	0.86	0.86
<i>Dried-Plant</i>	0.07	0.11	0.08	0.07	0.13	0.10
<i>Sky</i>	0.95	0.93	0.94	0.95	0.93	0.94
<i>Cliff</i>	0.65	0.45	0.53	0.69	0.70	0.69
<i>Tree</i>	0.49	0.52	0.51	0.56	0.47	0.51
<i>Sand</i>	0.02	0.10	0.03	0.57	0.45	0.50
<i>Sea</i>	0.69	0.60	0.64	0.85	0.69	0.76
<i>Boat</i>	0.41	0.71	0.52	0.33	0.66	0.44
<i>Road</i>	0.50	0.69	0.58	0.69	0.71	0.70
<i>Ground</i>	<i>0.26</i>	<i>0.33</i>	<i>0.29</i>	<i>0.26</i>	<i>0.33</i>	<i>0.29</i>
<i>Person</i>	<i>0.75</i>	<i>0.51</i>	<i>0.61</i>	<i>0.75</i>	<i>0.51</i>	<i>0.61</i>
<i>Trunk</i>	<i>0.26</i>	<i>0.28</i>	<i>0.27</i>	<i>0.26</i>	<i>0.28</i>	<i>0.27</i>
<i>Wave</i>	<i>0.25</i>	<i>0.5</i>	<i>0.33</i>	<i>0.25</i>	<i>0.5</i>	<i>0.33</i>

References

1. Smeulders, A., M.Worring, .Santini, S., .Gupta, A., .Jain, R.: Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(12) (2000) 1349–1380
2. Rao, A., Jain, R.: Knowledge representation and control in computer vision systems. *IEEE Expert* (1988) 64–79
3. Draper, B., Hanson, A., Riseman, E.: Knowledge-directed vision: control, learning and integration. *Proc. of the IEEE* **84**(11) (1996) 1625–1681
4. Little, S., Hunter, J.: Rules-by-example - a novel approach to semantic indexing and querying of images. In: *International Semantic Web Conference (ISWC)*, Hiroshima, Japan. (Nov. 7-11 2004) 534–548
5. Schober, J.P., Hermes, T., Herzog, O.: Content-based image retrieval by ontology-based object recognition. In: *Proc. KI-2004 Workshop on Applications of Description Logics (ADL)*, Ulm Germany. (Sept. 24 2004)
6. Neumann, B., Moller, R.: On scene interpretation with description logics. (FBI-B-257/04) (2004)
7. Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V.K., Strintzis, M.G.: Knowledge-assisted semantic video object detection. **15**(10) (October 2005) 1210–1224
8. Espinosa, S., Kaya, A., Melzer, S., Möller, R., Wessel, M.: Multimedia interpretation as abduction. In: *Proc. International Workshop on Description Logics (DL)*, Brixen-Bressanone, Italy. (Jun. 8-10 2007)
9. Hollink, L., Little, S., Hunter, J.: Evaluating the application of semantic inferencing rules to image annotation. In: *Proc. International Conference on Knowledge Capture (K-CAP)*, Banff, Alberta, Canada. (Oct. 2-5 2005) 91–98

10. Petridis, K., Bloehdorn, S., Saathoff, C., Simou, N., Dasiopoulou, S., Tzouvaras, V., Handschuh, S., Avrithis, Y., Kompatsiaris, I., Staab, S.: Knowledge representation and semantic annotation of multimedia content. *IEE Proceedings on Vision Image and Signal Processing, Special issue on Knowledge-Based Digital Media Processing* **153** (June 2006)
11. Bagdanov, A., Bertini, M., DelBimbo, A., Serra, G., Torniai, C.: Semantic annotation and retrieval of video events using multimedia ontologies. In: *Proc. IEEE International Conference on Semantic Computing (ICSC)*, Irvine, CA, USA. (2007)
12. Dasiopoulou, S., Heinecke, J., Saathoff, C., Strintzis, M.: Multimedia reasoning with natural language support. In: *Proc. IEEE International Conference on Semantic Computing (ICSC)*, Irvine, CA, USA. (Sept. 17-19 2007)
13. Straccia, U.: A fuzzy description logic for the semantic web. In Sanchez, E., ed.: *Fuzzy Logic and the Semantic Web. Capturing Intelligence*, Elsevier (2006) 73–90
14. Stoilos, G., Stamou, G., Tzouvaras, V., Pan, J., Horrocks, I.: The fuzzy description logic f-SHIN. In: *International Workshop on Uncertainty Reasoning For the Semantic Web (URSW)*, Galway, Ireland. (Nov. 7 2005)
15. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., P. F. Patel-Schneider, e.: *The description logic handbook: Theory, implementation, and applications*. In: *Description Logic Handbook*, Cambridge University Press (2003)
16. Stoilos, G., Stamou, G., Pan, J.: Handling imprecise knowledge with fuzzy description logic. In: *Proc. International Workshop on Description Logics (DL)*, Lake District, UK. (2006)
17. Straccia, U.: Reasoning within fuzzy description logics. *J. Artif. Intell. Res. (JAIR)* **14** (2001) 137–166
18. Straccia, U.: Transforming fuzzy description logics into classical description logics. In: *Proc. European Conference on Logics in Artificial Intelligence (JELIA)*, Lisbon, Portugal. (Sept. 27-30 2004) 385–399
19. Haase, P., van Harmelen, F., Huang, Z., Stuckenschmidt, H., Sure, Y.: A framework for handling inconsistency in changing ontologies. In: *Proc. of International Semantic Web Conference (ISWC)*, Galway, Ireland. (Nov. 6-10 2005) 353–367
20. Kalyanpur, A., Parsia, B., Sirin, E., Grau, B.C.: Repairing unsatisfiable concepts in owl ontologies. In: *Proc. of European Semantic Web Conference (ESWC)*, Budva, Montenegro. (Jun. 11-14 2006) 170–184
21. LeBorgne, H., Guérin-Dugué, A., O'Connor, N.: Learning midlevel image features for natural scene and texture classification. *IEEE Trans. Circuits Syst. Video Techn.* **17**(3) (2007) 286–297
22. Moosmann, F., Triggs, B., Jurie, F.: Randomized clustering forests for building fast and discriminative visual vocabularies. In: *Neural Information Processing Systems (NIPS)*. (Nov 2006)
23. Moller, R., Neumann, B., Wessel, M.: Towards computer vision with description logics: Some recent progress. In: *Proceedings Integration of Speech and Image Understanding*, Corfu, Greece. (1999) 101–115
24. Umberto, S., Giulio, V.: DImedia: an ontology mediated multimedia information retrieval system. In: *Proc. International Workshop on Description Logics (DL)*, Brixen-Bressanone, Italy. (Jun. 8-10 2007)
25. Simou, N., Athanasiadis, T., Tzouvaras, V., Kollias, S.: Multimedia reasoning with f-shin, 2nd International Workshop on Semantic Media Adaptation and Personalization, London, UK (2007)
26. Mylonas, P., Simou, N., Tzouvaras, V., Avrithis, Y.: Towards semantic multimedia indexing by classification and reasoning on textual metadata, Knowledge Acquisition from Multimedia Content Workshop, Genova, Italy (2007)