Applying Fuzzy DLs in the Extraction of Image Semantics

Stamatia Dasiopoulou[1], Ioannis Kompatsiaris[1], and Michael G. Strintzis[1,2]

[1] Centre for Research and Technology Hellas, Informatics and Telematics Institute, Thessaloniki, Greece
[2] Information Processing Laboratory, Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Greece
{dasiop,ikom,michael}@iti.gr

Abstract. Statistical learning approaches, bounded mainly to knowledge related to perceptual manifestations of semantics, fall short to adequately utilise the meaning and logical connotations pertaining to the extracted image semantics. Instigated by the Semantic Web, ontologies have appealed to a significant share of synergistic approaches towards the combined use of statistical learning and explicit semantics. While the relevant literature tends to disregard the uncertainty involved, and treats the extracted image descriptions as coherent, two valued propositions, this paper explores reasoning under uncertainty towards a more accurate and pragmatic handling of the underlying semantics. Using fuzzy DLs, the proposed reasoning framework captures the vagueness of the extracted image descriptions and accomplishes their semantic interpretation, while resolving inconsistencies rising from contradictory descriptions. To evaluate the proposed reasoning framework, an experimental implementation using the fuzzyDL Description Logic reasoner has been carried out. Experiments in the domain of outdoor images illustrate the added value, while outlining challenges to be further addressed.

1 Introduction

Semantic image analysis has challenged researchers for decades in the quest for generalisable approaches to alleviate the so called *semantic gap* [1,2,3,4], i.e. the lack of correspondence between the descriptions that can be automatically extracted from visual content and the respective meaning a human would attach. Towards this goal, statistical learning approaches have attracted increased interest in the last couples of years, as they provide powerful and effective means to discover, capture, and manage, complex associations between perceptual features (i.e. attributes of visual manifestations such as colour and texture) and semantic concepts. Support Vector Machines (SVMs) [5] and Bayesian Networks (BNs) [6] constitute popular examples, and have been espoused in numerous approaches targeting the extraction of image semantics [7,8,9,10,11,12].

Although promising results have been reported, the attained performance varies significantly with respect to the number of concepts addressed and the

S. Spaccapietra, L. Delcambre (Eds.): Journal on Data Semantics XIV, LNCS 5880, pp. 105–132, 2009.

considered image data sets as well. The observed variability relates to core challenges in computer vision, including perceptual similarities of semantically distinct concepts and perceptual variations in the possible manifestations of a single concept [13,14], much as to limitations inherent in the assumption that semantics can be rendered in a visual fashion. As a result of the aforementioned, the learnt associations do not necessarily reflect the intended semantics, viz. the associations implicitly targeted when providing the corresponding training examples, leading often to incomplete and conflicting classifications. Indicatively, among the observations presented in [15], the weapon classifier proves to be more efficient when querying for palm trees, while when querying for instances of fire and flames, the soccer classifier provides the highest performance.

Acknowledging the value of statistical learning techniques, yet aware of their weaknesses, approaches towards the synergistic utilisation of explicit semantics have become the subject of systematic research[1]. The Semantic Web (SW) incentive influenced to a large extent the choice of the investigated representation formalisms, favouring the use of SW languages [16,17] and of the closely related Description Logics (DLs) [18,19]. As a result, a number of multimedia ontologies [20,21,22,23] have been proposed to represent perceptual features and to enable linking with domain specific ontologies, in order to formalise the transition from low-level features to semantic entities [24,25,26,27]. In addition, domain ontologies, tailored to the analysis viewpoint as well, have been proposed in order to acquire interpretations of higher abstraction through reasoning over automatically extracted descriptions [28,29,30,31,32].

However, the effects entailing from the poor utilisation of semantics in statistical learning, namely the uncertainty inherent in the extracted descriptions and the semantic inconsistencies issuing from conflicting descriptions, tend to be overlooked. Specifically, the extracted classifications are commonly treated as crisp assertions, neglecting significant information regarding the plausibility of the acquired descriptions. Furthermore, in the majority of cases, the extracted descriptions are assumed to be semantically coherent. As a result, the use of explicit conceptual models and reasoning are rendered mainly as means to acquire abstract and complex descriptions by exploiting logical associations between the extracted descriptions, such as the inference of a person instance by reasoning over instances of face and body in a certain configuration. Evidently though, both aforementioned assumptions are rather weak and hardly correspond to the pragmatics of the problem at hand.

Aiming to enhance the utilisation of semantics and alleviate part of the aforementioned effects in the accuracy and completeness of descriptions that are extracted by means of machine learning approaches, we present in this paper a reasoning framework that utilises fuzzy DLs semantics in order to interpret the

[1] Indicatively, besides individual research activities, this pursuit has been the principal objective in a substantial number of European projects including aceMedia (http://www.acemedia.org/aceMedia), K-Space (http://kspace.qmul.net), BOEMIE (http://www.boemie.org/), X-Media (http://www.x-media-project.org/), MESH (http://www.mesh-ip.eu/), SALERO (http://www.salero.eu/), etc.

output of the classifiers into a semantically consistent interpretation. The use of DLs allows us to formally capture the semantics underlying the concepts of interest, while the fuzzy extensions provide the means to model the vagueness encompassed in the extracted classifications. Furthermore, extending on earlier investigations [33], the presented framework supports the explicit representation of the constituent image regions, allowing, as explained in the sequel, the more effective utilisation of the underlying semantics. The contribution of the proposed reasoning framework can be summarised in the following.

- The uncertainty of the descriptions made available through the application of learning based approaches is formally handled and taken into consideration in the interpretation of the descriptions' semantics.
- The inconsistencies resulting from conflicting descriptions, due to the aforementioned limitations in the learning of associations between perceptual features and corresponding semantics, are identified and resolved.
- Besides formally grounding the acquisition of the most plausible interpretations in the presence of multiple possible interpretations, the proposed fuzzy DLs based reasoning framework supports the identification of image regions where concepts, missed in the initial descriptions, may be present.

The rest of the paper is organised as follows. Sections 2 and 3 outline the reasons that motivated our investigation and the particular issues involved in the application of formal reasoning in semantic image analysis. Section 4 presents the proposed reasoning framework architecture and its constituent reasoning tasks, while Section 5 elaborates the implementation details. Section 6 presents the evaluation of the proposed framework and the experiences drawn. Relevant initiatives are presented in Section 7, while Section 8 summarises the paper and outlines future research directions.

2 Motivation

Statistical concept classifiers exhibit highly variable performance, yet support generic learning for a substantial number of concepts [15,34,35,36]. As demonstrated in a recent study [37], satisfactory retrieval can be achieved, even when the detection accuracy is low, provided that sufficiently many concepts are used, as long as these concepts can be related to one another in some reasonable way. In addition, the conducted experiments reveal that when there exist semantic associations between the addressed concepts, then these concepts can serve as an intermediate layer to enhance the reliability of the extracted semantic image descriptions. The conducted experiments consider semantic video descriptions, yet the results can be easily generalised for the case of image descriptions, since the examined concept classifiers address notions detected per video frames, i.e. without the use of temporal information.

The observations drawn by the aforementioned study regarding the potential of incorporating semantics, do not outline a new direction; approaches following the knowledge-directed paradigm have been reported since the early 70s, while

they boomed in the 80s and the early 90s in accordance with the respective advances in the field of Artificial Intelligence (AI) [38,39,40]. Yet, [37] stresses the greater potential that the recent advances in statistical concept classifiers conduce regarding the utilisation of explicit knowledge and reasoning as the means to alleviate the limitations related to the discriminative capacity of perceptual features with respect to the intended semantics.

As already described though, the limitations related to the rather poor utilisation of semantics are intertwined with the uncertainty involved in extracting semantic descriptions from images. As such, the fundamental question of what constitutes the semantics of this uncertainty, emerges. The answer lies in the viewpoint adopted in learning regarding the stipulation of semantics in accordance to perceptual features. Approaches where concepts are detected on the grounds of perceptual similarity, imply a prototypical set of feature values that constitute a visual/perceptual definition of the concept. As the presence of a concept is determined based on the similarity of those values, concepts can be considered as fuzzy sets, where the similarity (distance) function serves the role of the membership function. Contrariwise, learning approaches that utilise concepts' co-occurrence and correlation, implement a probabilistic interpretation of the features to concepts transition. Support Vector Machines (SVMs) constitute a popular example of the former category, while Bayesian Nets and Hidden Markov models [41] fall in the latter.

Apparently, both types of uncertainty pertain to the extraction of image semantics, much more since they address complementary aspects. A classification indicating that a specific image region constitutes an instance of sea with a probability of 0.7, refers to the presence or not of sea; no information is provided about how blueish this sea region might be. A classifier though that assess an image region to belong to the sea concept with a degree of 0.7, quantifies the similarity of this region with what has been learned as the perceptual definition of sea. For further details on the different semantics of the two uncertainty types, the reader is referred to [42]. The investigation of a reasoning framework that considers both types of uncertainty is undoubtedly of particular interest. In this work though, we focus on the fuzzy perspective of the extracted semantic descriptions, since we consider it an significant starting point for the appropriate handling of semantic classification results, and a useful insight into the complementary role of probabilistic reasoning.

The aforementioned incentives, in combination with the limited support for handling uncertainty and inconsistency provided by the state of the art approaches in the utilisation of explicit semantics, designated the selection of fuzzy DLs as the investigated knowledge representation. The logic grounded semantics ensure conceptual transparency and well-defined reasoning mechanisms, while maintaining a strong connection to the Semantic Web community. In addition, the fuzzy extensions allow to formally capture the imprecision in the from of vagueness that pertains to learning approaches based on perceptual similarity, such as SVMs. In combination with the particular expressivity requirements described in the following Sections, the aforementioned have been the main reasons

for preferring fuzzy DLs over some other logic based formalism, such as fuzzy first order logic, or fuzzy rules.

3 Fuzzy DLs in Semantic Image Analysis: Specifications and Requirements

Fuzzy DLs extend the model theoretic semantics of classical DLs [18] to fuzzy sets [43,44] and account for a significant share of the literature studying the representation of imprecise information [45,46,47,48,49,50]. Standardisation initiatives, such as the W3C Uncertainty Reasoning for the World Wide Web Incubator Group, which recently released the final report on reasoning under uncertainty in the Semantic Web[2], outline further the significance of handling imprecise knowledge in real world applications.

The semantics of a fuzzy DL language are given by a fuzzy interpretation $I = (\Delta^I, .^I)$, where Δ^I is an non-empty set of objects comprising the domain of interpretation, and $.^I$ a fuzzy interpretation function, which assigns each individual a to an element $a^I \in \Delta^I$, each concept name A to a membership function $A^I : \Delta^I \rightarrow [0, 1]$, and each role name R to a membership function $R^I : \Delta^I \times \Delta^I \rightarrow [0, 1]$ [47,48]. Table 1, illustrates the standard interpretation of typical DL constructors.

Table 1. Fuzzy interpretation of DL constructors following Zadeh semantics [49]

$$
\begin{aligned}
&\top^I = 1\\
&\bot^I = 0\\
&(\neg \mathrm{C})^I = 1\text{-}\mathrm{C}^I(d)\\
&(\mathrm{C} \sqcap \mathrm{D})^I = \min\{\mathrm{C}^I(d), \mathrm{D}^I(d)\}\\
&(\mathrm{C} \sqcup \mathrm{D})^I = \max\{\mathrm{C}^I(d), \mathrm{D}^I(d)\}\\
&(\forall\ \mathrm{R.C})^I = \inf_{d' \in \Delta} max\{1 - R^I(d, d'), C^I(d')\}\\
&(\exists\ \mathrm{R.C})^I = \sup_{d' \in \Delta} min\{R^I(d, d'), C^I(d')\}
\end{aligned}
$$

A fuzzy knowledge base consists of a TBox defined by a finite set of fuzzy concept inclusion and equality axioms, and an ABox defined respectively as a finite set of *fuzzy assertions*. A fuzzy assertion [47] is of the form $a : C \bowtie n$ and $(a, b) : R \bowtie n$, where $\bowtie$ stands for $\geq$, $>$, $\leq$, and $<$. 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. The standard reasoning services (e.g. instance checking, satisfiability, subsumption etc.) are adapted analogously. For example, concept satisfiability with respect to C requires the existence of an interpretation under which there will be an individual belonging to C with a degree $n \in (0, 1]$.

Using fuzzy DLs as the knowledge representation language for the semantic interpretation of descriptions provided by statistical concept classifiers, renders the

[2] http://www.w3.org/2005/Incubator/urw3/XGR-urw3-20080331/

Fig. 1. Example outdoor image and segmentation mask

$(im : Rockyside) \geq 0.50$ $(im : Countryside\_buildings) \geq 0.47$
$(im : Roadside) \geq 0.48$ $(im : Forest) \geq 0.65$
$(im : Seaside) \geq 0.46$

$(r0 : Building) \geq 0.68$ $(r0 : Trunk) \geq 0.54$
$(r1 : Sky) \geq 0.70$ $(r1 : Person) \geq 0.59$
$(r2 : Building) \geq 0.66$ $(r2 : Trunk) \geq 0.58$
$(r3 : Vegetation) \geq 0.56$ $(r3 : Rock) \geq 0.51$
$(r4 : Building) \geq 0.66$ $(r4 : Spectators) \geq 0.54$
$(r5 : Trunk) \geq 0.55$ $(r5 : Building) \geq 0.53$
$(r6 : Building) \geq 0.61$ $(r6 : Board) \geq 0.51$
$(r7 : Building) \geq 0.60$ $(r7 : Board) \geq 0.52$
$(r8 : Tree) \geq 0.56$ $(r8 : Grass) \geq 0.55$

Fig. 2. Scene and object level classifications results for the example image of Fig. 1 using SVM-based concept classifiers

available classifications into fuzzy assertions and the available domain knowledge into corresponding terminological axioms. Figures 1 and 2, illustrate an example outdoors image, its segmentation mask and the extracted classifications in the form of fuzzy DLs assertions.

As illustrated, an image may be asserted to belong to multiple scene level concepts, not necessarily semantically related; similarly, a region may belong to multiple object level concepts. This is not unusual and accounts for two equally common situations met in the extraction of semantic image descriptions. First, the use of multiple classifiers for a single concept in order to benefit from multiple sources of information, and second, classification errors that result in false positive responses for semantically contradictory concepts. For readability, we consider at most two instances per region.

Assuming the TBox of Table 2 and going through the respective assertions, one notices that there exist semantic discrepancies with respect to the extracted scene level descriptions, since according to axioms $1-6$ only one of them can be true, as well as between the scene level descriptions and the object level ones.

Table 2. Example TBox extract for the domain of outdoor images

axiom 1: Forest $\sqsubseteq$ Landscape $\sqcap \neg$ (Countryside_buildings $\sqcup$ Roadside)
axiom 2: Roadside $\sqsubseteq$ Landscape $\sqcap \neg$ (Forest $\sqcup$ Countryside_buildings)
axiom 3: Countryside_buildings $\sqsubseteq$ Landscape $\sqcap \neg$ (Forest $\sqcup$ Roadside)
axiom 4: Landscape $\sqsubseteq$ Outdoors $\sqcap \neg$ (Rockyside $\sqcup$ Seaside)
axiom 5: Seaside $\sqsubseteq$ Outdoors $\sqcap \neg$ (Landscape $\sqcup$ Rockyside)
axiom 6: Countryside_buildings $\sqsubseteq \exists$ contains.Building $\sqcup \exists$ contains.Grass
axiom 7: Countryside_buildings $\sqcap \exists$contains.(Spectators $\sqcup$ Board $\sqcap$ Rock) $\sqsubseteq \bot$

Contradictions may be straightforward, such as in the case of (im:Roadside) $\geq$ 0.52 and (im:Forest) ≥ 0.65, or implicit such as in the case of (im:Forest) ≥ 0.65 and (r0:Building) ≥ 0.68, where through inference the latter assertion entails that (im:Countryside_buildings) ≥ 0.68. Furthermore, the identification of inconsistencies depends on the scene level concept used as a reference. Assuming for example that the Forest scene description is valid, all region assertions referring to the Building, Spectators and Board concepts entail inconsistency. Assuming though that the Countryside_buildings scene description is valid, inconsistencies are raised by regions assertions referring to the concepts Spectators, Board and Rock instead.

Consequently, in order to reach a coherent interpretation, the possible alternative scene interpretations need to be identified and subsequently assessed with respect to their plausibility. This means that for all possibly satisfiable scene concepts, that is for all scene concepts for which a model exists when conflicts, the corresponding degrees of membership need to be computed in order to provide a measure for their plausibility. Due to the logical relations between the object and scene level concepts, the degree to which an image belongs to a scene concept does not necessarily equal the degree provided by the respective scene concept classifier. Hence, in our current example, the satisfiable, and thereby plausible, scene descriptions are (im:Countryside_buildings), (im:Forest), and (im:Rockyside); the corresponding minimum degrees are 0.68, 0.65 and 0.51.

Once the most plausible scene description is determined, the next step is to ensure that the object level descriptions are not introducing semantic conflicts. As in the case of scene level descriptions, the identification, tracking and resolving of such inconsistencies is intertwined with the semantics as defined in the TBox axioms. In the considered example, the identification and resolving of inconsistencies is rather straightforward, since all inconsistent assertions refer to atomic concepts (i.e. Spectators, Rock, and Board). In the presence of an axiom such as $Person \sqcap Bench \sqsubseteq Spectators$ though, the inconsistency could be resolved in multiple ways, namely by removing all Person instances, all Bench instances, or all instances of both classes. Selecting among the different alternatives needs to take into account cost criteria encompassing the assertions' degrees of confidence in order to retain the available plausibility information.

The final step towards a more complete image description is to compensate for missing assertions and enrich the existing descriptions by means of entailment. Missing assertions refer to scene or object level descriptions that are entailed by the computed scene interpretation, yet failed to be detected by the applied concept classifiers. As elaborated in subsection 4.3, for the case of object level descriptions, the proposed framework allows not only to recover the missing assertions, but also to acquire suggestions regarding which of the input region instances could be a possible match for the missing object level descriptions. Enrichment on the other hand covers those cases where scene and object level descriptions of higher abstraction can be inferred from the available ones. In the running example, where Countryside_buildings is designated as the most plausible scene description, the presence of at least one region belonging to the Building concept and one region belonging to the Grass concept is entailed, each with a degree ≥ 0.68. As a result, the degrees of the region assertions referring to the concept Building are updated, and so is the assertion concerning region $r8$, which now becomes most plausible from the initially extracted one referring to the concept Tree. Furthermore, due to axioms 3 and 4, the image is also asserted as an instance of Landscape and Outdoors.

We note the significance of the existential ($\exists$) and the disjunction ($\sqcup$) constructors for the aforementioned tasks. The existential quantification allows to handle cases where the initial descriptions are incomplete, due to segmentation fault or to erroneous classification, while the union constructor allows to represent and reason over the alternative scene interpretations in order to assess their satisfiability. Considering rule formalisms instead, we would lack the possibility to express existential quantification or use disjunction in the head of rules to so as to state the entailment of multiple possible alternatives. Using fuzzy first order logic, the latter would not pose a problem, yet we would be unable to infer the existence of regions corresponding to concepts failed to be detected by the classifiers, as described in detail in the following Section.

Finally, it is important to stress that the TBox aims to capture generic knowledge reflecting the logical associations issuing from the semantics of the concepts at hand, rather than data set specific conceptualisations, as the latter would risk false implications when invoked on classifications over data sets with differing attributes. For example, an axiom such as $\exists$contains.Tree $\sqcap$ $\exists$contains.Trunk $\sqsubseteq$ *Forest* may be representative for forest scenes for a given data set, yet in the general case it could lead to biased inferences, as Tree and Trunk instances can be as well found in many other scene descriptions. Restricting the included axioms to strict domain semantics modelling, the domain TBox can ensure that the extracted descriptions are in compliance with the semantics of the concepts they refer to.

4 Fuzzy DLs-Based Reasoning Framework for Semantic Image Analysis

Figure 3 depicts the proposed reasoning framework for managing the tasks outlined in the previous Section. As shown, the semantic interpretation of the

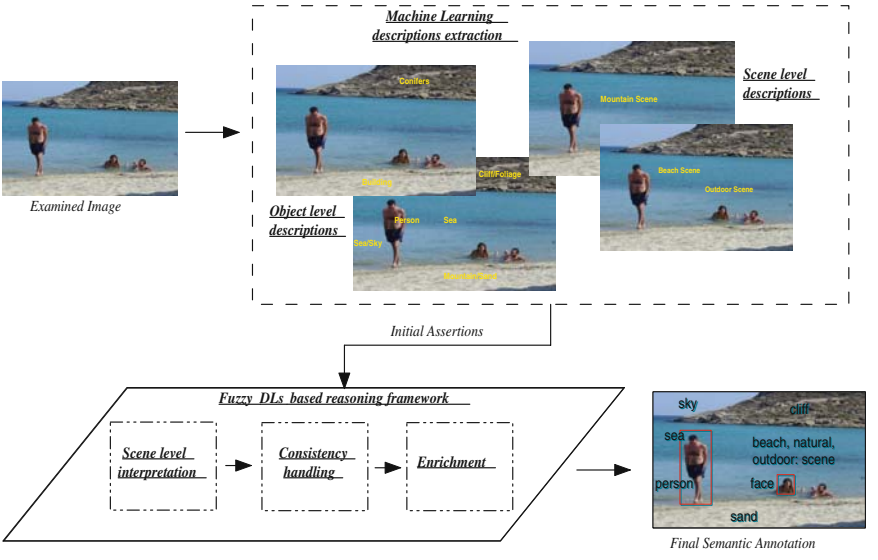

Fig. 3. Proposed fuzzy DLs based reasoning framework

descriptions that are acquired through statistical learning is performed in three steps. First, the most plausible description at scene level is determined. Next, the inconsistencies in the initial descriptions are resolved with respect to the previously computed scene level interpretation. As, more than one plausible interpretations may exist, during this step the different alternatives are ranked with respect to plausibility criteria. The set of assertions with the highest ranking is finally passed to the last step, where by means of logical entailment, assertions pertaining to complex or missing descriptions are made explicit.

Extending the conceptual framework of [33], in the current approach, we exploit localisation information for descriptions at object level. This is accomplished by explicitly representing the region which is associated with the extracted object concept descriptions, using the model of Table 3. According to it, an image and its constituent regions are associated through the role contains, while disjoint axioms make explicit the discrimination between scene and object level concepts, as well as between an image and its regions. As illustrated, there is no restriction on the number of descriptions at scene (respectively object) level that may be assigned to an image (region).

Towards a more conceptually accurate model, the first axiom would need to be revised as Image $\sqsubseteq \exists$depicts.SceneConcept, so as to capture the fact that the Image concept includes those objects in the domain of interpretation that are associated with a SceneConcept instance through a depicts property. Similarly for the third axiom which would become Region $\sqsubseteq \exists$depicts.ObjectConcept. As in the examined context though, both models would ensue the same inferences, we preferred the model of Table 3 to avoid unnecessary complexity.

Table 3. Annotation model for image and its constituent parts

$$Image \sqsubseteq SceneConcept$$
$$Region \sqsubseteq ObjectConcept$$
$$Image \equiv \exists contains.Region$$
$$Image \sqcap Region \sqsubseteq \bot$$
$$SceneConcept \sqcap ObjectConcept \sqsubseteq \bot$$

The explicit representation of the image regions has a twofold effect. On one hand, it allows to generate final descriptions of higher informative value, as object concept assertions are no longer associated only to the image (supporting loose annotation) but also to specific regions. On the other hand, in the case of inconsistent classifications, instead of simply removing the conflicting assertions and ending up with regions with no description, we are now able to compute suggestions of consistent object descriptions. In the following, the details of the three reasoning tasks are given.

4.1 Scene Level Interpretation

The alternative scene interpretations constitute the possible models regarding the interpretation of an image, and as a consequence for each SceneConcept$_i$ it suffices to find one model such that SceneConcept$_i \neq \emptyset$, instead of requiring SceneConcept$_i \neq \emptyset$ for all models. Towards this end, and due to the logical associations between concepts referring at the object level and concepts referring at the scene level, all assertions need to be taken into account in order to check the satisfiability of the alternative scene descriptions. As a consequence, all disjointness axioms in which scene level concepts are participating need to be removed before checking for satisfiability, as otherwise possible inconsistencies would reasoning and would prevent the effective utilisation of all information carried in the available extracted descriptions.

The scene interpretation procedure, summarised in Table 4, consists in the following steps. First, all disjoint axioms are removed and the TBox is revised with respect to the currently examined concept $SceneConcept_i$ so as to avoid conflicts when an image is inferred to belong both to $SceneConcept_i$ and its complement $\neg SceneConcept_i$. To accomplish this, the disjointness axioms are revised so as to entail an instance of $notSceneConcept_i$ instead of triggering an inconsistency. Considering the example TBox of 2, and for $SceneConcept_i$ equal to Rockyside, the presence of a region (r_i:Building $\geq d_i$), with $d_i \geq 0.5$ due to axiom 5 would entail (im:$\neg$ Rockyside ≥ 0.5), rendering the available classifications inconsistent. Revising axiom 5 as $Land- scape \sqsubseteq notRockyside$ though, the ABox remains consistent, and the assertion (im:notRockyside $\geq d_i$) is obtained instead.

Next, the satisfiability of each scene level description is checked considering all initial assertions besides the told scene level assertions that refer to a scene concept other than the currently examined one. Each scene concept $SceneConcept_i$, for which $notSceneConcept_i$ is not satisfiable constitutes a possible interpretation.

Table 4. Scene level interpretation

Scene Level Interpretation Algorithm
Input: scene level concepts hierarchy H_{SC}, input assertions A **Output**: glb for all satisfiable scene level concepts

```
1:  for all hierarchy levels L_i starting from the root
2:    for all scene level concepts SC_j ∈ L_i
3:      if ∃ satisfiable subsumee of SC_j or i==0{
4:        revise disjoint axioms adding nonSC_j
5:        check ¬ SC_j satisfiabity
6:          if ¬ SC_j not satisfiable{
7:            remove assertions inconsistent to SC_j
8:            update A and compute glb(SC_j)
9:          }
10:     }
11:  rank scene level concepts wrt glb
```

Thereby, compared with [33] the checks required to determine the most plausible scene descriptions, are reduced. To further improve the efficiency, the checking of the scene concepts satisfiability utilises the subsumption relations between the scene concepts. Thus, if a concept $SceneConcept_i$ is computed to be unsatisfiable, we skip the checks of all concepts subsuming it. For all satisfiable scene level concepts, the respective greater lower bound (glb) values are computed, by following the inconsistency handling methodology described in the sequel. Subsequently, the glb values are ranked and the scene concept with the highest one is selected as the most plausible scene description.

4.2 Inconsistency Handling

Having computed the scene level concepts that constitute possible interpretations, the next step is to obtain for each of them the most optimistic interpretation, in order to assess the most plausible one. Towards this end, for each satisfiable scene concept, the inconsistencies with respect to the input object concept assertions need to be identified and resolved. Following a similar procedure to the one described above, the TBox is revised so that the disjointness axioms involving the examined scene concept and object level ones, instead of causing an inconsistency, entail an instance of a correspondingly introduced $nonObjectConcept_i$. As more than one object level concepts $ObjectConcept_i$ may give rise to inconsistencies, a conjunctive expression is formed including the respective $nonObjectConcept_i$ concepts and the generic $nonObjectConcept$ concept is defined as it subsumer. To resolve the inconsistencies, we employ the procedure described in the following, until no instance of $nonObjectConcept$ with glb greater or equal than 0 exists.

Table 5. Expansion rules for computing the alternative sets of consistent assertions

$\sqcap$-rule	if $(a : C_1 \sqcap C_2) \in L(\mathrm{x})$ then $L(\mathrm{y})$=$L(\mathrm{x})\backslash \{(a : C_1)\}$ and $L(\mathrm{z})$=$L(\mathrm{x})\backslash\{(a : C_2)\}$ and $L(\mathrm{w})$=$L(\mathrm{x})\backslash\{(a : C_1 \sqcap C_2)\}$
$\sqcup$-rule	if $(a : C_1 \sqcup C_2) \in L(\mathrm{x})$ then $L(\mathrm{x})$=$L(\mathrm{x})\backslash\{(a : C_1 \sqcup C_2)\}$
where $C_i \longrightarrow$ A | C $\sqcap$ D | $\exists$R.D	

First, we address inconsistencies incurred directly by told descriptions. This translates into checking whether there exist asserted individuals belonging to $ObjectConcept_i$ concepts such that $ObjectConcept_i \sqsubseteq nonObjectConcept_i$. The handling of such assertions is rather straightforward and consists in their removal. Addressing asserted individuals first, prunes the search space during the subsequent tracking of the inferences that lead to an inconsistency. Next, we consider assertions referring to complex concepts, i.e. concepts for which the left hand side of the axioms in which they participate is an expression rather than an atomic concept. Contrary to the previous case, we now need to analyse the involved axioms in order to track the asserted descriptions that cause the inconsistency. Furthermore, these axioms determine which of the descriptions should be removed so as to reach a consistent interpretation. To accomplish this, we build on relevant works for resolving unsatisfiable DL ontologies [51,52], and employ a reversed tableaux expansion procedure, summarised in Table 5.

The main difference with respect to the relevant literature is that in our application framework, we consider solely the removal of assertions, rather than the removal or weakening of terminological axioms. The expansion procedure starts having as root node the $(im : nonObjectConcept \geq d_i)$ assertion, where d_i the computed degree, and continues until no expansion rule can be applied. As illustrated, in the case of inconsistencies caused by axioms involving the conjunction of concepts, there are multiple ways to resolve the inconsistency and reach a consistent interpretation. Specifically, there as many alternative interpretations as the sum of combinations $C(N, k)$, where N the number of conjuncts and $k = 1, .., N$. In order to choose among them, we rank the set of solutions according to the number of assertions that need to be removed and the average value of the corresponding degrees. Again, corresponding $nonObjectConcept_i$ definitions are added as in the case of scene concepts to avoid ending up with inconsistent ABoxes.

4.3 Enrichment

The final step considers the enrichment of the descriptions by means of typical fuzzy DLs entailment. Specifically, once the scene level interpretation is determined and all inconsistencies have been resolved, we end up with a semantically

consistent subset of the input assertions. To render the inferred descriptions explicit, corresponding queries are formulated and the responses are added to the final image interpretation. Inferred descriptions may refer either to concepts not addressed by the available classifiers (the Landscape concept constitutes such a concept for the example considered in Figures 1,2), or to concepts for whom the corresponding classifiers failed to produced a positive response.

Extending the framework of [33], the explicit representation of the constituent image regions, allows to model object level descriptions as instances referring to specific regions of the image, rather than to the entire image though the indirect representation of regions in the form of $(im : \exists contains.ObjectConcept_i)$ assertions. Such modelling allows for additional benefits besides the enhancement of loose image descriptions[3]. Specifically, once the inconsistency handling task is completed, there might be regions for which all initial assertions have been removed. Exploiting the visual coherency between the initial assertions associated to these regions and the assertions identified as missing, we can infer possible suggestions regarding object level concepts that such regions may depict. Furthermore, based again on the visual coherency of the concepts addressed by the classifiers, additional suggestions for missing assertions can be inferred with respect to regions that already have an object level concept assigned to them.

In order to capture and model the visual coherency of the considered concepts, we utilise the confusion matrixes acquired during the training phase of the classifiers and extract axioms of the form $ObjectConcept_i \sqsubseteq ObjectConcept_j \sqcup ObjectConcept_{j+1} \sqcup \ldots$, where the concepts $ObjectConcept_{j+n}$ represent object concepts that tend to be misclassified as instances of $ObjectConcept_i$ under a given scene description. The main reason for adopting such an approach, is that regions depicting visually similar object concepts often happen to be falsely segmented as one. As illustrated in the evaluation Section 6, the purpose of such suggestions is to facilitate the interaction with a subsequent step of analysis, including possibly re-segmentation and the re-application of specific classifiers on selected regions.

5 Implementation

In the previous Section, we described the individual tasks comprising the proposed fuzzy DLs-based reasoning framework for the enhancement of semantic image interpretation. In the following, we examine the proposed reasoning framework from an implementation perspective. Since, each task utilises corresponding standard fuzzy DLs reasoning services in order to accomplish its goals, central role in the proposed reasoning framework holds the reasoning engine that realises these core fuzzy DLs inference services.

The choice regarding which specific implementation should be employed under the proposed framework was based on the existing available fuzzy DL reasoning

[3] Loose (weak) annotation refers to object level descriptions that are associated to the entire image rather than the specific image regions.

engines and the requirements posed with respect to expressivity power and interaction capabilities. The sequence of works by Straccia [47,48,53] and Stoilos et al. [49,50,54] distill the advancements accomplished with respect to the formal definition of fuzzy extensions semantics and of corresponding reasoning algorithms. Complementary to the theoretic foundations, respective reasoning engine implementations have been developed, namely the *Fuzzy Reasoning Engine*[4] (FiRE) and the *fuzzyDL*[5].

FiRE [55] supports querying an f-*SHIN* knowledge base for satisfiability, consistency, subsumption, and entailment, under Zadeh semantics; general concept inclusions, roles and datatype support are among the planned future extensions. fuzzyDL [56] supports satisfiability, consistency, subsumption, and entailment for the language fuzzy *SHIF*, further extended by concrete fuzzy concepts, i.e. concept defined through an explicit fuzzy membership function, concept modifiers that allow to change the membership function of a fuzzy concept, and functional datatypes attributes. The reasoner accepts three types of semantics for the interpretation of conjunction, disjunction, complement and implication, namely Zadeh semantics, Lukasiewicz, and crisp. Although both available reasoners support very high expressivity and provide support for the standard reasoning services of satisfiability, instance checking, disjointness and subsumption, the factor that differentiates them is the handing of general concept inclusions. As illustrated in Sections 2 and 3, handling general concept inclusions is crucial as it allows to model the existence of specific regions, thus specific object level concept instances, which in turn imply corresponding scene concept instances. Otherwise, the object level instances would be reduced to scene concept instances, and subsequently the region individuals would become tautological to the respective image individual. Given the above considerations, we selected the fuzzyDL reasoner.

Fig. 4 shows an abstract view of the proposed reasoning framework architecture regarding the interaction between the proposed reasoning framework and the fuzzyDL reasoning engine. As illustrated, and already described in the detailed presentation of the procedure comprising each of the three reasoning tasks, the fuzzyDL engine provides the standard inference services required to support the semantic interpretation of an image. The proposed reasoning framework coordinates the required inference services by designating the each time considered TBox and ABox, performing appropriate translations to avoid inconsistencies and formulate respective queries so as to determine the conditions for the subsequent processing steps. Hence, it serves as an external mechanism that modularises and harmonises the interpretation into distinct subproblems on which the fuzzyDL can be invoked.

In addition, the proposed framework provides support for subtasks addressing the handling of semantics that cannot be invoked as distinct services of fuzzyDL. One such example is the tracking and resolving of inconsistencies, where besides the transformation of the TBox so that an inconsistency entails an instance of

[4] http://www.image.ece.ntua.gr/∼nsimou
[5] http://faure.isti.cnr.it/∼straccia/software/fuzzyDL/fuzzyDL.html

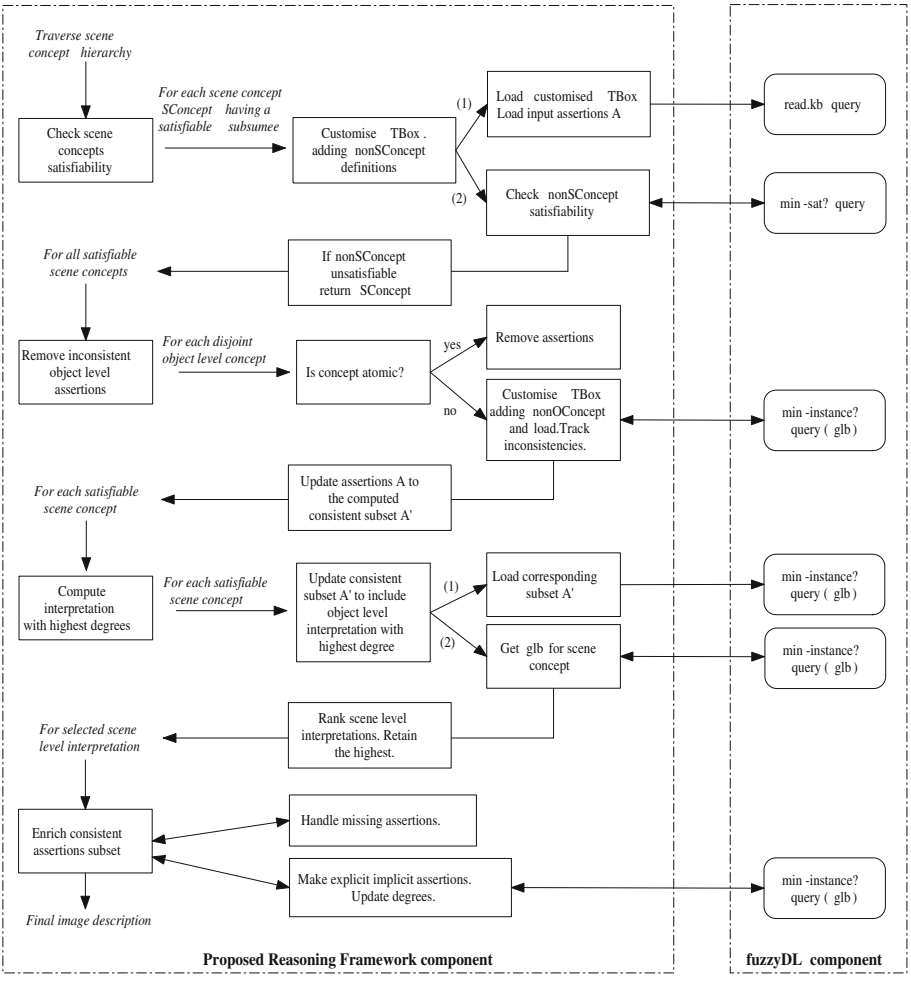

Fig. 4. Implementation architecture diagram

respectively introduced nonConcepts, the semantics of the axioms involved in the creation of an inconsistency are taken into account in order to compute the possible alternative solutions. We note that regarding the TBox revision, parts of the TBox translations, specifically the transformations during the satisfiability checks in the scene level interpretation task, have been manually performed, as the emphasis in the current experimental implementation is placed on assessing the feasibility of the proposed approach.

Another example is the computation of possible models with respect to a specific scene concept, given a TBox and the initial region assertions. In this case, first the TBox needs to be revised so that inconsistencies can be tracked and removed, and afterwards the possible models (i.e. the different configurations

regarding the association of each region to one of the disjuncts representing the available classification results) need to be explicitly provided to fuzzyDL as distinct ABoxes. The latter is essential in order to compute the corresponding glb values for the examined scene concept, as otherwise it would be impossible to entail a value different than ≥ 0 unless all disjuncts per region addressed the same concept. Going back to the example of Figure 2 for instance, it would be impossible to infer the scene concept Countryside_buildings unless there existed a region r_i for which all alternative possible classification where the same one, i.e. that of Building.

6 Evaluation

In order to assess the potential of the proposed reasoning framework for the purpose of enhancing the semantic coherency and completeness of image descriptions, we experimented in the domain of outdoor images. The sets of scene (C_{scene}) and object (C_{object}) level concepts addressed by the employed SVM based classifiers are C_{scene}={Roadside, Rockyside, Countryside_buildings, Seaside, Forest} and C_{object}={Building, Roof, Grass, Foliage, Dried-Plant, Sky, Rock, Tree, Sea, Sand, Boat, Road, Ground, Person, Trunk, Wave}.

Figure 5 illustrates example images of the addressed scene level concepts. As illustrated, Seaside images refer to coastal and beach scenes, Rockyside ones include mountainous images with little vegetation, Roadside images refer to landscape scenes depicting parts of road, Countryside_buildings represent scenes where buildings are present yet not in an urban environment, and finally Forest images correspond to natural landscapes with abundant vegetation, including trees, foliage, trunks, etc.

From an initial set of 700 outdoor images, two sets of 350 images have been assembled: one served as the training set on which the learning of concept classifiers was performed, and the second served as the test set. Ground truth for all images has been manually generated at object and scene level. The manual annotation and training of the classifiers, both comprising quite cumbersome and resource intensive activities (corresponding to an average of two to three

Fig. 5. Example images of the supported scene level classifiers

Table 6. Example extract of the outdoor image domain TBox developed for evaluation purposes

$$\begin{aligned}
&\text{Countryside_buildings} \sqsubseteq \exists \text{contains.Building} \sqcap \exists \text{contains.Grass}\\
&\text{Countryside_buildings} \sqsubseteq \text{Landscape}\\
&\text{Grass} \sqcup \text{Tree} \sqsubseteq \text{Foliage}\\
&\text{Rockyside} \sqsubseteq \exists \text{contains.Rock}\\
&\text{Roadside} \sqsubseteq \exists \text{contains.Road}\\
&\text{Roadside} \sqsubseteq \text{Landscape}\\
&\exists \text{contains.Building} \sqsubseteq \text{Countryside_buildings}\\
&\exists \text{contains.Sea} \equiv \text{Seaside}\\
&\text{Beach} \equiv \text{Seaside} \sqcap \exists \text{contains.Sand}\\
&\exists \text{contains.Sky} \sqsubseteq \text{Outdoor}\\
&\text{Trunk} \sqsubseteq \text{Tree}\\
&\text{Wave} \sqsubseteq \text{Sea}\\
&\text{Boat} \sqsubseteq \text{Sea}\\
&\text{Forest} \sqcap (\text{Roadside} \sqcup \text{Countryside_buildings}) \sqsubseteq \bot\\
&\text{Roadside} \sqcap \text{Countryside_buildings} \sqsubseteq \bot\\
&\text{Rockyside} \sqcap (\text{Seaside} \sqcup \text{Landscape}) \sqsubseteq \bot\\
&\text{Landscape} \sqsubseteq \sqcap \text{Outdoor} \bot\\
&\text{Forest} \sqcap \exists \text{ contains.(Rock} \sqcup \text{Sea} \sqcup \text{Sand} \sqcup \text{Building} \sqcup \text{Road)} \sqsubseteq \bot\\
&\text{Rockyside} \sqcap \exists \text{ contains.(Sea} \sqcup \text{Sand} \sqcup \text{Building} \sqcup \text{Road)} \sqsubseteq \bot
\end{aligned}$$

person months - for the number of concepts and images considered in the specific experiment), constitute efforts already spent for the purpose of training and assessing the performance of the employed SVM based classifiers [57]. Thus, the only extra resources required with respect to the proposed reasoning framework relate to the transformation of the already existing ground truth files and extracted descriptions to a format compliant to the one used by the proposed reasoning framework, which amounts to a negligible amount of effort. The reason for stressing this out, is to outline that the application of the proposed reasoning framework does not entail any additional resources with respect to annotation tasks.

In order to apply the proposed reasoning framework, a TBox that captures the semantics of the domain addressed by the available classifiers needs to be constructed. Table 6 illustrates an extract of the outdoor images TBox that has been developed for the carried out experimentation. It includes 25 concepts and one role. The included concepts comprise the scene and object level concepts supported by the classifiers, Landscape, Outdoor, and the generic SceneConcept and ObjectConcept concepts that are used to enforce that the discrimination between the two levels of concepts; the respective role is *contains*, which links an image to its constituent regions, and scene level concepts to object level ones.

Approximately fifty axioms, including the transformations required to avoid inconsistencies with respect to the alternative scene level interpretations, are used to capture the interrelations of the involved scene and object level concepts; this number increases further, when taking into account the additional

axioms appended during the handling of inconsistencies. We note though, that not all axioms are loaded at once to fuzzyDL, since the proposed reasoning framework coordinates, as previously explained, the axioms and assertions over which the reasoning services of fuzzyDL are invoked. As a consequence, the complexity remains too low to incur performance concerns, and similar observations have resulted when experimenting with larger TBoxes, as long as the considered ABoxes remained similar in size.

Specifically, using a virtual Linux machine, running on XP Windows, with an Intel Core quad processor, requires about three hours and twenty minutes to process the complete test set. Individual image processing times, vary from two seconds to one minute and half, depending on the given assertions and the complexity of resolving the encountered inconsistencies. Actually, the inconsistency handling process, which computes the possible consistent alternatives by tracking the definitions involved, and the satisfiability and glb queries communicated to fuzzyDL are the most time consuming tasks. The average memory required per image is 26 MBs, of which only a small fragment, namely 1/100, is consumed by the proposed framework, the rest committed by the evoked fuzzyDL services. Both observations relate to the fact that the proposed reasoning framework addresses mostly the coordination of the input and queries to be communicated to the fuzzyDL than realising itself core inference services, with the exception of tracking inconsistencies.

In order to quantify the performance of the proposed approach, we compared the accuracy and completeness of the obtained image descriptions, with the descriptions provided by means of classification, as well as with the descriptions acquired when using the reasoning framework of our previous study [33]. The last allows for a first estimate on the added value of explicitly representing the individual image regions and the alternative object descriptions associated with them. As evaluation metrics, we adopted recall, precision and and F-measure, according to the following definitions.

- Precision (p): number of correct assertions extracted/inferred per concept divided by the number of assertions that were extracted/inferred for the given concept.
- Recall (r): number of correct assertions extracted/inferred per concept divided by the number of assertions referring to that concept that are present in the ground truth image descriptions.
- F-measure: $2 * p * r/(p + r)$.

Table 7 gives the performance of the classifiers, of the reasoning framework presented in [33], and of the currently proposed reasoning framework, for the case of scene level concepts. Compared to the performance of the classifiers, we note that the application of the proposed reasoning framework incurs a significant improvement. Going through the respective domain axioms, it is easy to correlate the extend of enhancement to the extent of semantic relations between object level concepts with scene level, particularly axioms that entail a scene level descriptions based on object level descriptions. Compared with the respective reasoning performance of [33], the explicit representation of the individual

Table 7. Evaluation of analysis and reasoning performance for scene level concepts

	Analysis			**Reasoning [33]**			**Reasoning**		
Concept	**Recall**	**Precision**	**F-M**	**Recall**	**Precision**	**F-M**	**Recall**	**Precision**	**F-M**
Rockyside	0.68	0.70	0.69	0.68	0.79	0.74	0.65	0.77	0.72
Seaside	0.85	0.67	0.75	0.86	0.72	0.78	0.79	0.75	0.78
Beach	-	-	-	0.45	0.76	0.57	0.45	0.76	0.57
Roadside	0.68	0.69	0.69	0.72	0.70	0.70	0.72	0.63	0.67
Forest	0.75	0.63	0.69	0.74	0.68	0.71	0.76	0.68	0.72
Countryside buildings	0.30	1.0	0.46	0.60	0.86	0.71	0.60	0.86	0.71
Landscape	0.75	0.71	0.	0.87	0.85	0.85	0.87	0.85	0.85
Outdoor	-	-	-	1.0	1.0	1.0	1.0	1.0	1.0

image regions and the corresponding object level assertions appears to have a rather negligible effect, as the slight improvement observed for concepts such as Roadside and Forest is counterbalanced by the slight deterioration with respect to the Rockyside and Seaside concepts.

Table 8 compares the respective performance for descriptions at object level. As shown, besides the Boat and Grass concepts, the application of the reasoning framework of [33] improves significantly the performance compared to the sole application of the classifiers. This is a direct consequence of the fact that the considered object level concepts are characterised by rich semantics with respect to the scene level concepts that constitute their context of appearance. The behaviour observed with respect to the Boat and Grass concepts relates to the risks entailed by a false scene level interpretation, which may incur in the case of very poor classification performance, in which case the input descriptions suggest interpretations other than the actual one. Going for example through the images for which Boat assertions where falsely removed, thus incurring the observed lowering in the recall rate, we noticed that the corresponding prevailing scene level assertions were not incompliance with the actual scene semantics.

Similar considerations emerge when analysing the not so remarkable effect of reasoning in the recall of scene level concepts such as Rockyside. Going through the images depicting rocky side scenes, yet failed to be recognised as such, we noticed that in all cases the classifiers had falsely detected another scene level concept instead, despite the fact that the instantiations of the Rock concept were successfully detected in the their majority. Adding an axiom such as $\exists$contains.Rock $\sqsubseteq$ Rockyside would seem a reasonable idea for improving performance on the grounds that the available axioms did seem to overlook this knowledge. However, as in the case of Boat, such an amendment would imbalance the trade off between what constitutes domain semantics and what is mere tuning to the peculiarities of a given data set. In the discussed case, this is easy to illustrate simply considering how often it is for rocks to appear in beach scenes.

The application of reasoning however, under the model proposed in this paper that considers the individual regions, entails an even higher effect on the

Table 8. Evaluation of analysis and reasoning performance for object level concepts

	Analysis			**Reasoning**[33]			**Reasoning**		
Concept	**Recall**	**Precision**	**F-M**	**Recall**	**Precision**	**F-M**	**Recall**	**Precision**	**F-M**
Building	0.54	0.69	0.60	0.62	0.86	0.72	0.62	0.64	0.63
Roof	0.33	0.54	0.41	0.33	0.75	0.46	0.43	0.63	0.52
Grass	0.49	0.42	0.45	0.30	0.52	0.38	0.83	0.56	0.67
Vegetation	0.48	0.84	0.61	0.86	0.86	0.86	0.80	0.49	0.61
Dried-Plant	0.07	0.11	0.08	0.07	0.13	0.10	0.12	0.33	0.18
Sky	0.95	0.93	0.94	0.95	0.93	0.94	0.96	0.92	0.94
Rock	0.65	0.45	0.53	0.69	0.70	0.69	0.57	0.57	0.57
Tree	0.49	0.52	0.51	0.56	0.47	0.51	0.83	0.46	0.59
Sand	0.02	0.10	0.03	0.57	0.45	0.50	0.57	0.45	0.50
Sea	0.69	0.60	0.64	0.85	0.69	0.76	0.75	0.69	0.72
Boat	0.41	0.71	0.52	0.33	0.66	0.44	0.44	0.57	0.5
Road	0.50	0.69	0.58	0.69	0.71	0.70	0.77	0.52	0.62
Ground	0.26	0.33	0.29	0.26	0.33	0.29	0.49	0.45	0.47
Person	0.75	0.51	0.61	0.75	0.51	0.61	0.86	0.45	0.61
Trunk	0.26	0.28	0.27	0.26	0.28	0.27	0.33	0.22	0.27
Wave	0.25	0.5	0.33	0.25	0.5	0.33	0.25	0.5	0.33

completeness and accuracy of the object level descriptions. This a direct consequence of the fact that instead of leaving a region without a corresponding assertion in the case the classification results prove to be inconsistent, probable suggestions are inferred that as illustrated incur further improvement. In order to obtain the values illustrated in the Table, we considered for each region the inference-based suggestion with the highest degree. As described in Section 4 though, more than one suggestions may be inferred for a given region, independently of whether this region has been subjected to inconsistent classification, aspiring to further assist in the identification of additional descriptions. As a result, in the case of a more interactive analysis and classification module, the proposed framework has the potential for an effectively higher enhancement.

Table 9 provides a rough assessment of the potential benefit such suggestions entail, by measuring the respective recall and precision values when all suggested additional descriptions are taken into account. As expected, the concepts that exhibit the higher potential for improvement are those for which once the scene description has been identified, their perceptual similarity with already detected concepts allows them to be associated to existing region assertions. As described, these suggestions have disjunctive semantics, i.e. they do not necessitate the presence of the suggested concept but rather identify the most plausible regions at which this concepts should be sought. Figure 6 provides an estimation of the number of regions that should be searched if no such information was available, i.e. when the only knowledge relates to the region assertions that have been missed during classification, and the respective number of regions when the suggestions provided by the proposed reasoning framework are used, provided

Table 9. Evaluation of reasoning for object level concepts including the inferred suggestions

	Reasoning		
Concept	**Recall**	**Precision**	**F-M**
Building	0.66	0.83	0.74
Roof	0.35	0.69	0.46
Grass	0.60	0.75	0.67
Vegetation	0.6	0.68	0.64
Dried-Plant	0.05	0.22	0.08
Sky	0.95	0.93	0.94
Rock	0.65	0.45	0.53
Tree	0.49	0.52	0.51
Sand	0.02	0.10	0.03
Sea	0.69	0.60	0.64
Boat	0.47	0.71	0.52
Road	0.64	0.78	0.70
Ground	0.27	0.31	0.28
Person	0.75	0.51	0.61
Trunk	0.41	0.33	0.37
Wave	0.25	0.5	0.33

that they are correct; otherwise, the searching for a missing object reduces to the same situation as in the former case. As illustrated when the inferred suggestion are taken into account, the number of regions that need be examined is reduced almost by half for concepts for which semantic and perceptual information is available.

Summing up the experiences and observations drawn from the conducted evaluation, we note that the utilisation of explicit semantics has a positive impact towards the semantic interpretation of image descriptions. The use of fuzzy DLs allows to handle formally the degrees of confidence that accompany the automatically extracted and utilise them both towards the identification of the most plausible interpretation as well as for resolving inconsistencies. The preservation of the degrees information in combination with the ensured semantic coherency of the resulting image descriptions, renders the proposed framework a useful contribution for semantic retrieval tasks that address multimedia content. Furthermore, since the presented framework makes no assumption with respect to the classifiers used to provide the initial classifications, it has the potential to be employed in any image retrieval scenario involving vague. descriptions.

Indicatively, practical cases where the proposed framework could be employed include applications such as the TRECVID[6] challenge, where among the addressed tasks is the extraction of high-level visual content descriptions using statistical learning. Within such context, the proposed reasoning framework could be used to alleviate inconsistent classifications and to enhance the completeness

[6] http://www-nlpir.nist.gov/projects/trecvid/

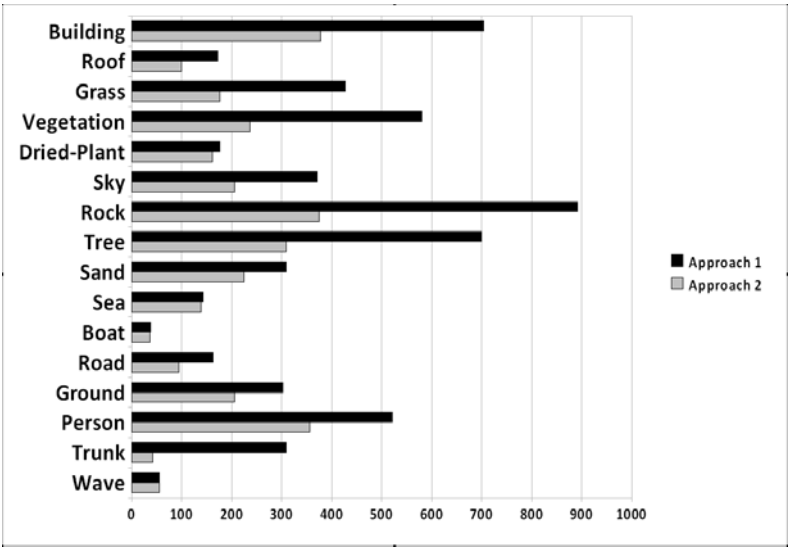

Fig. 6. Comparison of the number of regions that need to be examined for missing object level assertions, when no additional knowledge is available (approach 1) and when the inferred suggestions are taken into consideration (approach2)

of the final content descriptions. Thereby, the reliability of the descriptions is improved, while concepts that are not supported by the classifiers, but are semantically related, can be afforded, sparing the time and effort for building such classifiers. Another example application where the proposed framework could be used is the DL-Media [58] retrieval system in order to allow the ontological query service to perform over inconsistent image descriptions. In general, as exemplified in the motivating examples and the carried out evaluation, the proposed framework has the potential to enhance content descriptions, and by consequence the corresponding content management tasks, acquired by means of typical statistical learning techniques, when the underlying imprecision refers to vagueness.

7 Related Work

The majority of relevant literature considers the investigation of crisp DLs-based approaches. In the series of works presented in [28,59,29], crisp DLs are proposed for inferring descriptions modelled as logical aggregates. A probabilistic approach is described in the more recent one as a possible solution to the handling of the ambiguity introduced in visual analysis [29]. Although the proposed approach outlines an interesting direction, it lacks the technical details and evaluation that would establish the potential contribution; furthermore, considering probabilistic information, it addresses a different kind of uncertainty than what is presented in this paper.

In [31], DLs have been extended with a rule-based approach to realise abductive inference over crisp analysis assertions. Alternative consistent interpretations are computed by means of abduction and ranked using as criteria the number of new individuals that need to be introduced[7] and the number of assertions that need to be left out in order to reach a consistent interpretation. Examining the combined use of such an abductive reasoning framework with the proposed one fuzzy, could be interesting for investigating the effect in the ranking of alternative interpretations.

In [60] DLs are used to realise the interpretation of feature values pertaining to colour, texture and background knowledge to semantic objects. To this end a pseudo fuzzy algorithm is presented to reason over the calculated feature values with respect to the prototypical values constituting the definition of semantic objects. Additionally, topological knowledge is utilised to exclude inconsistent associations of semantic objects to given image segments. More specifically, in addition to the axioms representing the domain topology, axioms are introduced to capture topologically inconsistent relations. During a post processing step, individuals participating in the latter type of axioms are iteratively removed. Compared to the approach to inconsistency handling presented in this paper, [60] does not address the semantics of expressive constructors, while neither the implementation details of this postprocessing step nor evaluation results are given.

In [30], DLs and rules have been utilised for video annotation using crisp semantics. Additionally, there is no mentioning to what happens in the case of inconsistency. In [61], a DLs based approach to medical image annotation is presented under the assumption of crisp, consistent analysis extracted descriptions. In [62] a reasoning approach adhering to fuzzy logic principles was investigated for the purpose of integrating image descriptions extracted by means of visual analysis and textual analysis, regarding user entered descriptions, while in a more recent study presented in [33], a fuzzy DLs based reasoning framework has been proposed for the enhancement of initial descriptions acquired through statistical classifiers. As aforementioned, the presented approach extends on the last two investigations, building upon the acquired experiences.

Fuzzy DLs have been proposed in [58] for the purpose of semantic multimedia retrieval; the fuzzy annotations however are assumed to be available. Similar considerations have been investigated in older works such that of [63], where again the usefulness and significance of multimedia content descriptions that reflect the uncertainty present is pointed out. In the context of analysis, fuzzy DLs have been only recently explored in [55], where fuzzy DLs reasoning is used to infer semantic concepts based on part-of relations and to subsequently merge at image level. Possible inconsistencies in the analysis extracted description on which part-whole reasoning is employed is not addressed. Fuzzy logic semantics have also been investigated in [64] under a different application context, namely

[7] This is a direct result of treating the concepts to be inferred as aggregates of simpler ones and an interpretation as the quest of those aggregated and simple assertions that once introduced make use of the analysis extracted assertions.

for the purpose of supporting personalised information retrieval. In contrast to the approach presented in this paper, the emphasis is placed on weighted fuzzy concepts that are used to represent user preferences and contextualised preferences in order to allow the ranking of retrieved documents with respect to their relevance.

8 Conclusions and Future Work

The richness of visual information and the growth in the volume made available, render the potential for the exploitation of image content tremendous. Although the role of machine learning in the extraction of image semantics continues to grow, the reported endeavours show that the weakness to effectively incorporate semantics bears significant limitations in terms of the number of concepts that can be supported and the robustness of the attained performance. At the same time, the utilisation of explicit semantics as means to partially alleviate and enhance descriptions extracted through statistical learning presents an appealing potential, as suggested by recent studies addressing both research and industrial aspects [37,65].

Utilising fuzzy DLs semantics, the proposed reasoning framework captures the uncertainty of the extracted descriptions and accomplishes their integrated interpretation, while resolving inconsistencies rising from contradictory descriptions. In addition, by means of logical entailment, the final interpretation is further enriched; thereby, the need for training classifiers for semantically related concepts is partially alleviated, while missing descriptions due to segmentation and classification errors can be partially compensated. Experimentation has shown promising results, that although not conclusive yet, suggest that the proposed framework has the potential to serve as a useful contribution.

As indicated earlier in the paper, the investigation of a reasoning framework that combines fuzzy and probabilistic reasoning constitutes a challenging direction for future work. The motivation issues from the fact that the two types of uncertainty serve complementary purposes, hence suggesting a strong potential for achieving mutual benefit. However, more immediate directions for future investigations constitute on one hand on the extension of the presented reasoning framework so as to handle spatial knowledge, as well as the formalisation of the proposed reasoning tasks based on the drawn experiences with respect to the aspects that render the typical DL services inappropriate for direct exploitation in the problem of semantic image interpretation. Finally, towards more conclusive observations, we plan to extend our experimentation to larger, public data sets.

Acknowledgements

This work was partially supported by the European Commission under contracts FP6-001765 aceMedia, FP6-507482 KnowledgeWeb and FP7-215453 WeKnowIt.

References

1. Smeulders, A.W.M., Worring, M., 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), 1349–1380 (2000)
2. Chang, S.F.: The holy grail of content-based. IEEE MultiMedia 9(2), 6–10 (2002)
3. Naphade, M., Huang, T.: Extracting semantics from audio-visual content: the final frontier in multimedia retrieval. IEEE Transactions on Neural Networks 13(4), 793–810 (2002)
4. Hanjalic, A., Lienhart, R., Ma, W., Smith, J.: The holy grail of multimedia information retrieval: So close or yet so far away. IEEE Proceedings, Special Issue on Multimedia Information Retrieval 96(4), 541–547 (2008)
5. Burges, C.: A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2(2), 121–167 (1998)
6. Heckerman, D.: A tutorial on learning with bayesian networks. Learning in Graphical Models, 301–354 (1998)
7. Chapelle, O., Haffner, P., Vapnik, V.N.: Support vector machines for histogram-based image classification 10(5), 1055–1064 (1999)
8. Naphade, M., Huang, T.: A probabilistic framework for semantic video indexing, filtering, and retrieval. IEEE Transactions on Multimedia 3(1), 141–151 (2001)
9. Assfalg, J., Bertini, M., Colombo, C., Bimbo, A.D.: Semantic annotation of sports videos. IEEE MultiMedia 9(2), 52–60 (2002)
10. Christmas, W.J., Jaser, E., Messer, K., Kittler, J.: A multimedia system architecture for automatic annotation of sports videos. In: ICVS, pp. 513–522 (2003)
11. Town, C., Sinclair, D.: A self-referential perceptual inference framework for video interpretation. In: Crowley, J.L., Piater, J.H., Vincze, M., Paletta, L. (eds.) ICVS 2003. LNCS, vol. 2626, pp. 54–67. Springer, Heidelberg (2003)
12. Snoek, C., Worring, M., van Gemert, J., Geusebroek, J., Smeulders, A.: The challenge problem for automated detection of 101 semantic concepts in multimedia. In: Proc. 14th ACM International Conference on Multimedia, Santa Barbara, CA, USA, October 23-27, pp. 421–430 (2006)
13. Rao, A., Jain, R.: Knowledge representation and control in computer vision systems. IEEE Expert, 64–79 (1988)
14. Crevier, D., Lepage, R.: Knowledge-based image understanding systems: A survey. Computer Vision and Image Understanding 67, 161–185 (1997)
15. Snoek, C., Huurnink, B., Hollink, L., Rijke, M., Schreiber, G., Worring, M.: Adding semantics to detectors for video retrieval. IEEE Transactions on Multimedia 9(5), 975–986 (2007)
16. Horrocks, I., Patel-Schneider, P., van Harmelen, F.: From shiq and rdf to owl: the making of a web ontology language. J. Web Sem. 1(1), 7–26 (2003)
17. Horrocks, I., Patel-Schneider, P., Bechhofer, S., Tsarkov, D.: Owl rules: A proposal and prototype implementation. J. Web Semantics 3(1), 23–40 (2005)
18. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F.: The description logic handbook: Theory, implementation, and applications. In: Description Logic Handbook. Cambridge University Press, Cambridge (2003)
19. Baader, F., Horrocks, I., Sattler, U.: Description logics as ontology languages for the semantic web. In: Mechanizing Mathematical Reasoning, pp. 228–248 (2005)
20. Hunter, J.: Adding Multimedia to the Semantic Web: Building an MPEG-7 Ontology. In: Proc. The First Semantic Web Working Symposium (SWWS), California, USA (July 2001)

21. Simou, N., Saathoff, C., Dasiopoulou, S., Spyrou, E., Voisine, N., Tzouvaras, V., Kompatsiaris, I., Avrithis, Y., Staab, S.: An ontology infrastructure for multimedia reasoning. In: Proc. International Workshop on Very Low Bitrate Video Coding (VLBV), Sardinia, Italy, September 15-16, pp. 51–60 (2005)
22. Arndt, R., Troncy, R., Staab, S., Hardman, L., Vacura, M.: COMM: Designing a well-founded multimedia ontology for the web. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 30–43. Springer, Heidelberg (2007)
23. Dasiopoulou, S., Tzouvaras, V., Kompatsiaris, I., Strintzis, M.G.: Capturing mpeg-7 semantics. In: Proc. International Conference on Metadata and Semantics (MTSR), Corfu, Greece, October 11-12 (2007)
24. Troncy, R.: Integrating structure and semantics into audio-visual documents. In: Fensel, D., Sycara, K., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870, pp. 566–581. Springer, Heidelberg (2003)
25. Hunter, J., Drennan, J., Little, S.: Realizing the hydrogen economy through semantic web technologies. IEEE Intelligent Systems Journal - Special Issue on eScience 19, 40–47 (2004)
26. 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, 255–262 (2006)
27. Little, S., Hunter, J.: Rules-by-example – A novel approach to semantic indexing and querying of images. In: McIlraith, S.A., Plexousakis, D., van Harmelen, F. (eds.) ISWC 2004. LNCS, vol. 3298, pp. 534–548. Springer, Heidelberg (2004)
28. Moller, R., Neumann, B., Wessel, M.: Towards computer vision with description logics: Some recent progress. In: Proc. Workshop on Integration of Speech and Image Understanding, Corfu, Greece, September 21, pp. 101–115 (1999)
29. Neumann, B., Moller, R.: On scene interpretation with description logics, FBI-B-257/04 (2004)
30. 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, pp. 713–720 (2007)
31. 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, June 8-10, pp. 323–331 (2007)
32. Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V., Strintzis, M.: Knowledge-assisted semantic video object detection. IEEE Trans. Circuits Syst. Video Techn. 15(10), 1210–1224 (2005)
33. Dasiopoulou, S., Kompatsiaris, I., Strintzis, M.: Using fuzzy dLs to enhance semantic image analysis. In: Duke, D., Hardman, L., Hauptmann, A., Paulus, D., Staab, S. (eds.) SAMT 2008. LNCS, vol. 5392, pp. 31–46. Springer, Heidelberg (2008)
34. Maron, O., Ratan, A.: Multiple-instance learning for natural scene classification. In: Proc. 15th International Conference on Machine Learning (ICML), Madison, Wisconson, USA, July 24-27, pp. 341–349 (1998)
35. Vailaya, A., Figueiredo, M., Jain, A., Zhang, H.: Image classification for content-based indexing. IEEE Transactions on Image Processing 10(1), 117–130 (2001)
36. Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D., Jordan, M.: Matching words and pictures. Journal of Machine Learning Research 3, 1107–1135 (2003)

37. Hauptmann, A., Yan, R., Lin, W.H., Christel, M., Wactlar, H.: Can high-level concepts fill the semantic gap in video retrieval? a case study with broadcast news. IEEE Transactions on Multimedia 9(5), 958–966 (2007)
38. Niemann, H., Sagerer, G., Schröder, S., Kummert, F.: Ernest: A semantic network system for pattern understanding. IEEE Trans. Pattern Anal. Mach. Intell. 12(9), 883–905 (1990)
39. Reiter, R., Mackworth, A.K.: A logical framework for depiction and image interpretation. Artif. Intell. 41(2), 125–155 (1989)
40. Russ, T., MacGregor, R., Salemi, B., Price, K., Nevatia, R.: Veil: Combining semantic knowledge with image understanding. In: ARPA Image Understanding Workshop, Palm Springs, CA, USA, February 12-17 (1996)
41. Rabiner, L., Juang, B.: An introduction to hidden markov models. IEEE ASSP Magazine, [see also IEEE Signal Processing Magazine] 3(1), 4–16 (1986)
42. Dubois, D., Prade, H.: Possibility theory, probability theory and multiple-valued logics: A clarification. Annals of Mathematics and Artificail Intelligence 32(1-4), 35–66 (2001)
43. Zadeh, L.: Fuzzy sets. Information and Control 8(32), 338–353 (1965)
44. Klir, G., Yuan, B.: Fuzzy sets and fuzzy logic: Theory and applications. Prentice-Hall, Englewood Cliffs (1995)
45. Yen, J.: Generalizing term subsumption languages to fuzzy logic. In: Proc. 12th International Joint Conference on Artificial Intelligence (IJCAI), Sydney, Australia, August 24-30, pp. 472–477 (1991)
46. Straccia, U.: A fuzzy description logic. In: Proc. International Conference on Artificial Intelligence and 10th Innovative Applications of Artificial Intelligence Conference (AAAI/IAAI), Madison, Wisconsin, July 26-30, pp. 594–599 (1998)
47. Straccia, U.: Reasoning within fuzzy description logics. J. Artif. Intell. Res. (JAIR) 14, 137–166 (2001)
48. Straccia, U.: Transforming fuzzy description logics into classical description logics. In: Alferes, J.J., Leite, J. (eds.) JELIA 2004. LNCS (LNAI), vol. 3229, pp. 385–399. Springer, Heidelberg (2004)
49. 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, November 7, pp. 67–76 (2005)
50. Stoilos, G., Stamou, G., Pan, J.: Handling imprecise knowledge with fuzzy description logic. In: Proc. International Workshop on Description Logics (DL), Lake District, UK, pp. 119–127 (2006)
51. Bell, D., Qi, G., Liu, W.: Approaches to inconsistency handling in description-logic based ontologies. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM-WS 2007, Part II. LNCS, vol. 4806, pp. 1303–1311. Springer, Heidelberg (2007)
52. Lam, J., Sleeman, D., Pan, J., Vasconcelos, W.: A fine-grained approach to resolving unsatisfiable ontologies. In: Spaccapietra, S. (ed.) Journal on Data Semantics X. LNCS, vol. 4900, pp. 62–95. Springer, Heidelberg (2008)
53. Straccia, U.: A fuzzy description logic for the semantic web. In: Sanchez, E. (ed.) Fuzzy Logic and the Semantic Web. Capturing Intelligence, pp. 73–90. Elsevier, Amsterdam (2006)
54. Stoilos, G., Stamou, G., Pan, J., Tzouvaras, V., Horrocks, I.: Reasoning with very expressive fuzzy description logics. J. Artif. Intell. Res. (JAIR) 30, 273–320 (2007)
55. Simou, N., Athanasiadis, T., Tzouvaras, V., Kollias, S.: Multimedia reasoning with f-shin. In: 2nd International Workshop on Semantic Media Adaptation and Personalization (SMAP), London, UK, pp. 413–420 (2007)

56. Bobillo, F., Straccia, U.: fuzzydl: An expressive fuzzy description logic reasoner. In: Proc. International Conference on Fuzzy Systems (FUZZ), Hong Kong, June 1-6, pp. 923–930. IEEE Computer Society, Los Alamitos (2008)
57. Papadopoulos, G.T., Mylonas, P., Mezaris, V., Avrithis, Y., Kompatsiaris, I.: Knowledge-assisted image analysis based on context and spatial optimization (2006)
58. Umberto, S., Giulio, V.: Dlmedia: an ontology mediated multimedia information retrieval system. In: Proc. International Workshop on Description Logics (DL), Brixen-Bressanone, Italy, June 8-10, pp. 467–475
59. Neumann, B., Weiss, T.: Navigating through logic-based scene models for high-level scene interpretations. In: Crowley, J.L., Piater, J.H., Vincze, M., Paletta, L. (eds.) ICVS 2003. LNCS, vol. 2626, pp. 212–222. Springer, Heidelberg (2003)
60. 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, September 24, pp. 1–10 (2004)
61. Hu, B., Dasmahapatra, S., Lewis, P., Shadbolt, N.: Ontology-based medical image annotation with description logics. In: Proc. 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), Sacramento, California, USA, November 3-5, pp. 77–83 (2002)
62. 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, September 17-19 (2007)
63. Meghini, C., Sebastiani, F., Straccia, U.: A model of multimedia information retrieval. J. ACM 48(5), 909–970 (2001)
64. Mylonas, P., Vallet, D., Castells, P., Fernandez, M., Avrithis, Y.: Personalized information retrieval based on context and ontological knowledge 23(1), 73–100 (March 2008)
65. Leger, A., Heinecke, J., Nixon, L., Shvaiko, P., Charlet, J., Hobson, P., Goasdoue, F.: Semantic web take-off in a european industry perspective. In: Garcia, R. (ed.) Semantic Web for Business: Cases and Applications, ch. 1, pp. 1–29. IGI Global (2008)