

Using Multiple Domain Visual Context in Image Analysis*

Phivos Mylonas and Yannis Avrithis
National Technical University of Athens
Image, Video and Multimedia Laboratory
{fmylonas, iavr}@image.ntua.gr

Abstract

In this paper we propose an algorithm to improve the results of knowledge-assisted image analysis, based on contextual information. In order to achieve this, we utilize fuzzy algebra, fuzzy sets and relations, towards efficient manipulation of image region concepts. We provide a novel context modelling, based on the OWL language, using RDF reification. Initial image analysis results are enhanced by the utilization of domain-independent, semantic knowledge in terms of concepts and relations between them. The novelty of the presented work is the context-driven re-adjustment of the degrees of confidence of the detected concepts produced by any image analysis technique, utilizing a domain-independent ontology infrastructure to handle the knowledge, as well as multiple application domains.

1. Introduction

In content-based image search and retrieval, more and more researchers are looking beyond low-level colour, texture, and shape features in pursuit of more effective searching methods. Natural object detection in indoor or outdoor scenes, i.e. identifying key object types such as sky or ground, can facilitate content-based applications, ranging from image enhancement to coding. Contextual information in terms of specific concepts, objects and events, typically present in a beach, mountain or city scenery, could be a considerable source of useful information [6]. A significant number of misclassifications usually occur because of the similarities in colour and texture characteristics of various object types and the lack of such context information, which is a major limitation of individual object detectors. Generic algorithms for automatic object recognition and/or scene classification [7] are unfortunately not producing reliable results and restricting the problem to a specific domain does not provide a global, satisfactory solution.

The notion of visual (scene) context is introduced in [6] as an extra source of information for both object detection and scene classification. The truth is that the idea behind the use of such additional information refers to the fact that not all events are relevant in all situations and this holds also

when dealing with image analysis problems. Visual context is a difficult notion to grasp and capture and thus we restrict it herein to the notion of ontological context, defined as part of the fuzzy OWL ontology presented in subsection 2.2. Our choice is aligned with the clear research trend that exists in the literature [5] towards "fuzzification" of ontology description languages, like fuzzy DL and fuzzy OWL, as the representation and reasoning capabilities of fuzzy go clearly beyond classical.

In the following we propose our initial research progress in implementing an OWL-based context representation approach to use on top of any knowledge-assisted image analysis methodology. The ultimate goal is to be able to apply our approach on top of any given image domain; currently, we present preliminary experimental results focused on a limited set of possible domain datasets. The classification output of the specific image under consideration may be used optionally as a refinement step to the final contextualized results. Consequently, the only restriction that applies to the input image, is the fact that it should lie within the broad set of domains supported by the contextual ontology.

During the proposed contextualization process, we introduce a methodology to improve the results of image segmentation, in terms of each region's concepts, based on the introduced contextual information; in comparison to our previous research efforts [1] a novel multiple domain ontological representation for context is introduced, combining fuzzy theory and fuzzy algebra [3] with OWL [9] and RDF reification [10]. In this process, the membership degrees of the regions' concepts are re-estimated appropriately, according to a context-based membership degree readjustment algorithm.

The outline of this work is as follows: section 2 is dedicated to the multiple domain fuzzy context knowledge representation used, including some necessary notation used throughout the paper. Section 3 describes the proposed contextualization in terms of visual context algorithm steps and section 4 presents some preliminary experimental results, whereas section 5 concludes our work.

2. Fuzzy Context Knowledge Representation

The proposed context-based knowledge model can be expressed formally with the use of basic elements, such

*This research was partially supported by the European COST292 Action and the European Commission under contract FP6- 001765 aceMedia.

as concepts, relations between concepts and domains, that build an ontology structure [2]. Since relations among real-life concepts are always a matter of degree, they are best modelled using principles from fuzzy relational algebra [3], summarized in the following.

2.1. Background and notation

Given a universe U , a crisp set S of concepts on U is described by a membership function $\mu_S : U \rightarrow \{0, 1\}$. The crisp set S is defined as $S = \{s_i\}$, $i = 1, \dots, N$, $s_i \in S$, whereas a fuzzy set F on S is described by a membership function $\mu_F : S \rightarrow [0, 1]$. Given the set of all fuzzy sets on S , \mathcal{F}_S , then $F \in \mathcal{F}_S$. We define the crisp set of (crisp) relations between concepts as $\{r_i\}$, where $r_i : S \times S \rightarrow \{0, 1\}$, $i = 1, \dots, M$ and $M \leq \frac{N!}{(N-2)!}$ and in a similar way, we let R be the crisp set of fuzzy relations between concepts: $R = \{R_i\}$, where $R_i : S \times S \rightarrow [0, 1]$, $i = 1, \dots, M$. Let Z be the crisp set of all subdomains, that at the same time are considered to be concepts as well, then $Z \subseteq S$ and $Z = \{z_i\}$, $i = 1, \dots, L$, $L \leq N$, $z_i \in S$. Finally, the current subdomain (dis-)similarity is denoted by $p_i \in [0, 1]$, defined a priori by domain experts and characterizing the sub-domain. Index i denotes the particular subdomain and we let $|Z|$ identify the number of distinct subdomains z_i .

As the definition of a knowledge model based on fuzziness is proved to be the most suitable for the modeling of real-life information, a fuzzy ontology is constructed [1], containing subdomains, concepts and fuzzy relations together with degrees of membership for any given concept-to-concept relation. The latter can be formalized as: $\mathcal{O} = \{S, R, Z\}$, where \mathcal{O} denotes a fuzzy ontology, S is the crisp set of concepts described by the ontology, R is the crisp set of all fuzzy relations amongst these concepts and Z is the crisp set of subdomains available in \mathcal{O} .

Let also C be the fuzzy set of contexts on S and \mathcal{C} be the set of all contexts on \mathcal{O} , $\mathcal{C} \subset \mathcal{F}_S$. We consider one fuzzy set of contexts C , associated to each subdomain $z_i \in Z$ of the ontology, or in other words we define the notion of context for one subdomain $z_i \in Z$ as the fuzzy set $C_{z_i} = \sum_j c_j/w_j = \{c_1/w_1, c_2/w_2, \dots, c_n/w_n\}$, $n \leq N$ of all concepts $c_j \in S$ that are *directly* related to the subdomain z_i itself, together with the degrees of membership w_j of their relations. The membership degree w_j describes the membership function $\mu_F(c_j)$, i.e. $w_j = \mu_F(c_j)$, or for the sake of simplicity $w_j = F(c_j)$. The term *directly* is used with respect to the distance measure utilized (e.g. concepts located less than two nodes from the subdomain concept).

In order to define, extract and use each context C on top of an arbitrary set of concepts S , we rely on the semantics of their fuzzy semantic relations. Based on the relations R_i we construct relation I as a combination of relations, which further enhances the definition of fuzziness between concepts in the ontology \mathcal{O} : $I = \bigcup_i R_i^{q_i}$, $q_i \in \{-1, 0, 1\}$,

$i = 1 \dots M$. The q_i value is determined according to the semantics¹ of each relation R_i , as the latter are defined in the MPEG-7 standard [8], such as *specialization*, *part*, *property*, etc. As the construction of relation I implies, care must be taken in avoiding possible cycles, that may be present due to the cross-subdomain links between any two concepts. Cycles are avoided by not taking into consideration concepts that at the same time are subdomains as well, thus calculating the fuzzy set of contexts for all concepts other than the subdomains ones. Relation I is of great importance, as it allows us to define, extract and use contexts C out of any given set of concepts S .

2.2. Knowledge model and representation

The proposed contextual knowledge representation provides the means to exploit contextual information within an image analysis framework. In the past, we were motivated by RDF's flexible approach to representing data and modeled contextual knowledge with respect to the RDF language. In this paper we propose the enhanced use of OWL instead of RDF, since the former describes better the problem's requirements and is also expected to allow performing of useful reasoning tasks on the existing knowledge. On top of that we use OWL reification to represent the fuzzy degrees of membership.

The proposed context model can be seen as a graph, in which every node represents a concept and each edge between two nodes a contextual relation between the respective concepts. Additionally each edge has a related degree of confidence, expressing the desired fuzziness within the context model. One of the main differences to previous approaches and one of the main reasons to use OWL instead of RDF, is the fact that now a new set of "special" concepts exists, i.e. the crisp set of *subdomains* $Z \subseteq S$. These

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#" ... >
<owl:Class rdf:ID="beach"> <rdfs:subClassOf rdf:resource="#holidays" />
...
</owl:Class> <owl:Class rdf:ID="holidays"/> ...
<owl:Class rdf:ID="sand"/> <rdf:Description rdf:about="s1">
<rdf:subject rdf:resource="#sand"/>
<rdf:predicate rdf:resource="#isRelated"/>
<rdf:object rdf:resource="#beach"/>
<rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Statement"/>
<context:isRelated rdf:datatype="http://www.w3.org/2001/XMLSchema
#float">0,85</context:isRelated>
</rdf:Description> ... </rdf:RDF>

```

Fig. 1: OWL reification representation.

subdomain concepts are manually inserted into the graph model by domain experts and do function as intermediaries between the *root element* and the rest of concepts. Subdomains are not considered to be part of the set of concepts an image analysis module may have provided, but are indicative of the domains supported by the context knowledge.

Representing the above graph model in OWL is not a trivial or straightforward task, due to the complexity as well

¹*direct* (i.e. $q_i = 1$) semantic interpretation of a relation may not be meaningful, whereas its *inverse* may be (i.e. $q_i = -1$) and $q_i = 0$ eliminates a relation.

variety of its constraints and validation rules. Nevertheless, our context model is fully OWL-validated, as it is built with respect to the entire set of OWL syntactic rules. The fuzzy degree of confidence associated to each relation between any two concepts is described using the well-known from RDF reification technique, only this time implemented in OWL. Reification makes a statement about the statement, which contains the degree information and is ideal for representing fuzziness in a meaningful way, as the reified statement should not be asserted automatically. For instance, having a statement, such as "water isRelatedTo mountain" and a degree of confidence of 0,85 for this statement, does obviously not entail, that water is always related to a mountain scene/image. A representative example, using *isRelated* as the representation of combined relation *I* is illustrated in Fig. 1; concepts *beach* and *sand* are provided.

2.3. Graph model and representation

In the current approach and provided the utilization of the notion of subdomains, the *root element* plays an important role, in terms of calculating all contextual information with respect to it, however, it is not the one that characterizes the application domain. Quite on the contrary, it is the amount and type of subdomains that defines the latter, so as to be able to apply the proposed technique to multiple application domains.

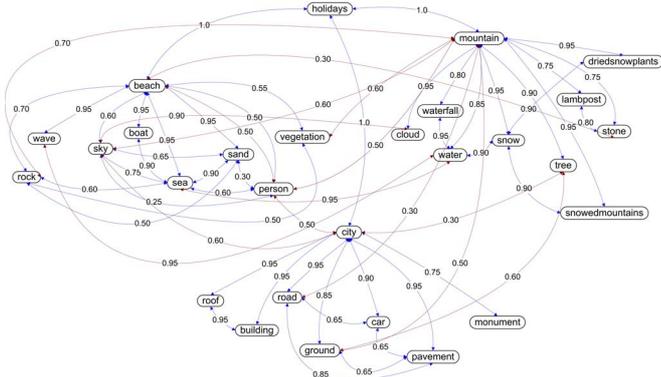


Fig. 2: Ontology graph sample - *beach*, *mountain*, *city* subdomains

Given the sample graph illustrated in Fig. 2, all possible routes in the graph are taken into consideration forming an exhaustive approach, with respect to the facts that routes between concepts may be reciprocal and routes between any two concepts is prohibited to include a subdomain (concept). Thus, estimation of each concept's context value is initially bound to the specific subdomain the concept belongs initially and is derived from relationships of the concept with other concepts, belonging either to the same or even different subdomains, always with respect to their weighted links. Of course, a meaningful compatibility indicator or distance metric is used in the process. Depending on the nature of the domains under consideration, the best indicator could be selected among various operators found in the literature, like for instance the standard

t-norm or *t*-conorm. In general, the ideal distance metric for two concepts is one that quantifies their semantic correlation and given the three subdomains *beach*, *mountain* and *city* presented in Fig. 2, the *max* value is a meaningful measure of correlation.

3. Visual Context Analysis

Based on the previous principles and mathematical foundations, we now present the visual context algorithm. Its core functionality is the meaningful readjustment of the membership degrees of each concept associated to a region or segment of an image, obtained from any kind of image analysis module. The novelty introduced herein deals with the context value introduced in subsection 2.1, which is utilized in order to tackle cases where one concept is related to more than one concepts, lying either within its original subdomain or even in other subdomains, as well. In any case, a concept's context value refers to the overall relevance of the concept to the related subdomain and thus to the *root element* of the entire ontology. For instance, in Fig. 2 the *root element* is *holidays*, whereas *beach*, *mountain* and *city* are the considered subdomains and concept *water* is related to both *beach* and *mountain* ones. The role of subdomains is to act as a barrier when considering all possible routes in the graph during each concept's context value estimation: no route is allowed to cross through a subdomain concept, since this would incorrectly mix the semantics of each subdomain. Using the standard *t*-conorm and the algebraic product as the *t*-norm, we have (see notation in 2.1):

1. for each subdomain $z_i \in Z$ identify the set of its contexts C_{z_i} .
2. for each concept $s_k \in S$ associated to an image region with a membership degree $\mu(s_k)$, identify its existing relations to the set of contexts: $\{R_{s_k, c_j} : c_j \in C_{z_i}\}$.
3. *iff* $R_{s_k, c_j} \neq 0$, then obtain the particular contextual information in the form of a new degree of membership $\mu(s_k)$ associated to the image region, according to the recursive formula:
$$\mu^t(s_k) = \mu^{t-1}(s_k) - p_i \cdot (\mu^{t-1}(s_k) - w_j) \quad (1)$$
where t denotes the iteration used.
4. prune the set of concepts that are not directly related to the subdomain z_i the image belongs to, according to its original classification output².

The physical meaning of (1) is that p_i controls whether or not context will be applied, i.e. $p_i = 0$ denotes no context influence, whereas $p_i = 1$ indicates a strong context leverage and ignoring of initial degrees $\mu(s_k)$. Equivalently to (1), for an arbitrary iteration i , we have:

$$\mu^t(s_k) = (1 - p_i)^t \cdot \mu^0(s_k) + (1 - (1 - p_i)^t)w_j \quad (2)$$

where $\mu^0(s_k)$ represents the initial degree of membership. Typically, the iteration stops when t equals 3 or 5.

²in order to decrease the influence of context, this step may be omitted and thus the application domain "smoothened".

4. Experimental Results

In order to test the implementation and efficacy of the presented algorithm, we carried out experiments utilizing 573 images evenly distributed to the three supported subdomains of interest, i.e. 193 *beach*, 184 *mountain* and 196 *city* images acquired from personal collections and the Internet. In the following we present a representative set of both indicative and overall quantitative results. A ground truth was manually constructed, consisting of a number of connected, non-overlapping regions associated to a unique concept.

Initially, two versions of an indicative *beach* image are presented (Fig. 3): (a) the original input image and (b) the segmented set of regions of the image. In order to obtain the segmented output, we implemented segmentation using the RSST algorithm [4] and a distance threshold for termination in a meaningful number of regions. Utilizing 100 images (merely 20% of the dataset) as a training set, we selected the values for p_i that resulted in the best overall evaluation.



(a) Input image (b) Segmented regions

Fig. 3: Indicative *beach* image example.

Table 1 summarizes the final degrees of membership before and after applying visual context for the *beach* image example. Boldface degrees indicate the ground truth information of the image regions, whereas the first column of the Table indicates the specific region ID under consideration. For the sake of space, concepts that are totally unrelated to the *beach* subdomain (as indicated in Fig. 2) are omitted.

Further, overall precision scores on a per concept basis, from the application of the proposed methodology to the entire dataset of the 573 images and for the entire holidays domain, are presented in Table 2. Each concept's row displays the scores before and after the use of context.

5. Conclusions

The methodology presented herein can be exploited towards the development of a more efficient, context-based image analysis environment. Its core contribution has been the implementation of a novel, multi-domain visual context interpretation utilizing a fuzzy, OWL-based, ontological representation of knowledge, as well as a visual context algorithm. Early research results were presented, indicating a significant aid (i.e. 3%-15% per concept - 9,38% overall) of visual context to the image analysis chain and it is the authors' belief that better results can be achieved with an optimized contextualization approach.

Table 1: Final degrees of membership before and after applying visual context for the beach image example of Fig. 3

ID	Concepts											
	rock		wave		sky		boat		sea		sand	
	before	after	before	after	before	after	before	after	before	after	before	after
1	0,26	0,09	0,63	0,54	0,70	0,46	0,34	0,11	0,71	0,79	0,47	0,14
2	0,63	0,42	0,13	0,07	0,67	0,24	0,70	0,28	0,17	0,03	0,77	0,46
3	0,73	0,48	0,82	0,41	0,69	0,25	0,68	0,27	0,37	0,07	0,98	0,59
4	0,33	0,11	0,70	0,59	0,67	0,45	0,31	0,10	0,77	0,85	0,27	0,08
5	0,69	0,59	0,24	0,12	0,11	0,10	0,48	0,32	0,18	0,16	0,85	0,94
6	0,67	0,58	0,25	0,13	0,13	0,12	0,37	0,24	0,29	0,26	0,79	0,87
7	0,62	0,41	0,41	0,21	0,31	0,10	0,69	0,27	0,17	0,04	0,78	0,47
8	0,53	0,35	0,44	0,22	0,37	0,12	0,70	0,28	0,24	0,11	0,77	0,46
9	0,49	0,32	0,48	0,24	0,41	0,14	0,94	0,38	0,09	0,07	0,27	0,16
10	0,23	0,09	0,65	0,55	0,73	0,48	0,39	0,13	0,79	0,87	0,18	0,05
11	0,40	0,24	0,25	0,13	0,22	0,05	0,78	0,34	0,44	0,33	0,92	0,55
12	0,37	0,22	0,36	0,18	0,47	0,11	0,70	0,30	0,67	0,51	0,77	0,46
13	0,53	0,32	0,42	0,21	0,39	0,09	0,68	0,29	0,35	0,26	0,98	0,59
14	0,73	0,62	0,30	0,15	0,27	0,18	0,31	0,20	0,25	0,23	0,69	0,75
15	0,69	0,58	0,34	0,29	0,11	0,07	0,81	0,87	0,18	0,10	0,85	0,73
16	0,67	0,57	0,35	0,30	0,13	0,09	0,67	0,73	0,07	0,04	0,77	0,66
17	0,65	0,55	0,45	0,38	0,21	0,14	0,69	0,74	0,27	0,17	0,61	0,52

ID	Concepts											
	person		vegetation		cloud		waterfall		water		stone	
	before	after	before	after	before	after	before	after	before	after	before	after
1	0,21	0,05	0,15	0,02	0,70	0,56	0,36	0,18	0,71	0,68	0,19	0,06
2	0,77	0,92	0,76	0,27	0,37	0,30	0,30	0,15	0,27	0,14	0,76	0,42
3	0,73	0,87	0,82	0,29	0,42	0,34	0,18	0,09	0,36	0,18	0,62	0,34
4	0,14	0,04	0,24	0,04	0,67	0,54	0,50	0,25	0,70	0,63	0,12	0,04
5	0,59	0,53	0,74	0,67	0,31	0,22	0,48	0,24	0,18	0,14	0,74	0,67
6	0,67	0,61	0,75	0,67	0,33	0,23	0,22	0,11	0,14	0,11	0,75	0,67
7	0,95	0,96	0,81	0,28	0,31	0,25	0,29	0,15	0,67	0,22	0,81	0,45
8	0,83	0,87	0,64	0,22	0,67	0,54	0,33	0,17	0,19	0,10	0,45	0,25
9	0,89	0,89	0,88	0,31	0,51	0,41	0,44	0,22	0,09	0,07	0,61	0,34
10	0,13	0,03	0,23	0,03	0,63	0,50	0,43	0,22	0,59	0,56	0,05	0,01
11	0,80	0,88	0,65	0,23	0,72	0,58	0,28	0,14	0,44	0,33	0,65	0,39
12	0,77	0,92	0,76	0,27	0,17	0,14	0,30	0,15	0,26	0,20	0,76	0,46
13	0,73	0,87	0,82	0,29	0,25	0,20	0,48	0,24	0,13	0,10	0,82	0,49
14	0,73	0,62	0,70	0,63	0,67	0,47	0,71	0,35	0,26	0,19	0,70	0,63
15	0,69	0,23	0,74	0,35	0,51	0,41	0,18	0,09	0,38	0,28	0,74	0,56
16	0,67	0,22	0,75	0,36	0,43	0,34	0,24	0,12	0,47	0,34	0,75	0,56
17	0,95	0,31	0,81	0,39	0,36	0,29	0,30	0,15	0,50	0,37	0,81	0,61

Table 2: Overall precision scores per concept on 573 images

Concepts	bef. aft.		Conc.	bef. aft.		Conc.	bef. aft.	
building	0,78	0,87	wave	0,54	0,60	sky	0,75	0,86
lambpost	0,49	0,51	sea	0,74	0,82	sand	0,68	0,74
vegetation	0,78	0,83	person	0,73	0,82	cloud	0,66	0,68
waterfall	0,32	0,33	water	0,37	0,40	snow	0,75	0,79
snowedmountains	0,46	0,49	stone	0,64	0,70	boat	0,67	0,75
driedsnowplants	0,52	0,54	tree	0,56	0,58	roof	0,74	0,83
monument	0,57	0,65	ground	0,69	0,76	car	0,84	0,89
pavement	0,70	0,74	road	0,76	0,82	rock	0,67	0,73
Overall	0,64	0,70						

References

- [1] Th. Athanasiadis, Ph. Mylonas, Y. Avrithis, S. Kollias, *Semantic Image Segmentation and Object Labeling*, IEEE TCSVT SICVIR, to appear, 2007.
- [2] T. R. Gruber, *A Translation Approach to Portable Ontology Specification*, Knowledge Acquisition 5: 199-220, 1993.
- [3] S. Miyamoto, *Fuzzy Sets in Information Retrieval and Cluster Analysis*, Kluwer Academic Publishers, 1990.
- [4] O. J. Morris, M. J. Lee, A. G. Constantinides, *Graph theory for image analysis: An approach based on the shortest spanning tree*, IEEE Proceedings, Vol. 133, pp. 1461-152, 1986.
- [5] E. Sanchez, *Fuzzy Logic and the Semantic Web*, Elsevier Science Inc., New York, NY, USA, 2006.
- [6] A. Torralba, *Contextual influences on saliency*, Neurobiology of attention, Academic Press Inc, London, 2005.
- [7] A. Torralba, *Contextual priming for object detection*, Int. J. Comp. Vis., vol. 53, pp. 169-191, 2003.
- [8] ISO/IEC FDIS 15938-5, ISO/IEC JTC 1/SC 29 M 4242, *IT MCDI Part 5: MDS*, pp. 442-448, October 2001.
- [9] W3C, *OWL*, <http://www.w3.org/2004/OWL/>
- [10] W3C, *RDF Reification*, http://www.w3.org/TR/rdf-schema/#ch_reificationvocab