Image Analysis Using Domain Knowledge and Visual Context

Phivos Mylonas, Thanos Athanasiadis and Yannis Avrithis
National Technical University of Athens
Image, Video & Multimedia Laboratory
P.C. 157 73, Zographou Campus, Athens, Greece
Phone: (+30) 210-7724351 Fax: (+30) 210-7722492 E-mail: fmylonas@image.ntua.gr

Abstract - Tackling the problems of automatic object recognition and/or scene classification with generic algorithms is not producing efficient and reliable results in the field of image analysis. Restricting the problem to a specific domain is a common approach to cope with this, still unresolved, issue. In this paper we propose a methodology to improve the results of image analysis, based on available contextual information derived from the popular *sports* domain. Our research efforts include application of a knowledge-assisted image analysis algorithm that utilizes an ontology infrastructure to handle knowledge and MPEG-7 visual descriptors for region labeling. A novel ontological representation for context is introduced, combining fuzziness with Semantic Web characteristics, such as RDF. Initial region labeling analysis results are then being re-adjusted appropriately according to a confidence value readjustment algorithm, by means of fine-tuning the degrees of confidence of each detected region label. In this process contextual knowledge in the form of domain-specific semantic concepts and relations is utilized. Performance of the overall methodology is demonstrated through its application on a real-life still image dataset derived from the *tennis* sub-domain.

1. INTRODUCTION

Digital multimedia content is rapidly gaining in size and quality following human communication and entertainment needs. It is the description of the content at a semantic level, though, the one that provides an added value to this statement. The importance of high level indexing mechanisms, which produce multimedia content description, is crucial. In the filed of image analysis, most state-of-the-art systems are limiting themselves mostly to very low level descriptions [5], such as the dominant color. The MPEG-7 standard [10] provides functionalities for management of multimedia content and metadata, but it lacks on the extraction of semantic description and annotation. Consequently, the need for efficient conceptualization in image analysis, based on some kind of knowledge, is evident.

Knowledge-assisted analysis can be defined as a tightly coupled and constant interaction between low level image analysis algorithms and higher level knowledge representation. The type of knowledge that is gaining momentum is that of *ontologies* [12]. Ontologies express key entities and relationships of multimedia content in a formal machine-processable representation and can help to bridge the semantic gap between the automatically extracted low-level arithmetic features and the high-level human understandable semantic concepts. Within this scope, we have implemented an experimentation framework that produces semantic interpretation of images by means of region-based fuzzy labeling. Still, because the results are highly dependent on the domain an image belongs to, the output of knowledge-assisted analysis is in many cases not sufficient for the understanding of multimedia content. Herein, we introduce a methodology for improving its results, based on contextual information obtained from application-specific domain ontologies. A context-based labeling update algorithm is introduced, that describes the readjustment of labeling information utilizing available contextual knowledge in the form of ontologies.

One of the main obstacles in the field of context-sensitive image analysis post-processing relies on the fact that it is very difficult to create domain applicable solutions, without in general having a large knowledge about the particular problem being solved. Otherwise, limitation in analysis capabilities is inevitable and has led researchers into the area of visual context modeling. In a recent review, Torralba [13] outlines the history of visual context modeling. Studies by Biederman et al. [2] and Palmer [8] highlight the effect of contextual information in the processing time for object recognition. Henderson and Hollingworth [4] have reported results suggesting that the choice of image regions is governed not merely by their low-level saliency but also by scene semantics. In a similar vein, several studies support the idea that scene semantics can be available early in the chain of information processing [9], and suggest that scene recognition may not require object recognition as a first step. In this framework, the term *context* [7] can take on many meanings and there is not even one definition that is felt to be globally satisfactory. It is therefore very important to establish a working interpretation, in order to exploit it.

The structure of this paper is as follows: in section 2, we discuss the main components of our experimentation framework. In section 3 we present in detail the contextual knowledge infrastructure, introducing the notion of fuzzy relations in ontologies, along with the utilized context-based confidence value readjustment algorithm. In section 4 we provide some experimental results from a real-life still image dataset and in section 5 we present our concluding remarks.

2. INITIAL REGION LABELING

Semantic comprehension of images involves in a great degree the tasks of image classification and object

This work is partially funded by the European Commission under contract FP6-001765 aceMedia and by the Greek Secretariat of Research and Technology (PENED Ontomedia 03 ΕΔ 475).

recognition. Restriction of images into predefined classes, namely domains, improves drastically the results of object recognition algorithms; a practice commonly used and also followed herein. We have implemented an experimentation platform called KAA (i.e. Knowledge-Assisted Analysis platform), that its main target is to identify semantic objects within an image of a given domain [1]. In this section we describe briefly the work carried out so far in KAA and all the relative foundations used. This was considered essential for the sake of scientific completeness and for the solid presentation of the results, since KAA's results serve as input for the contextual processing, or in other words, as initial region labeling.

For our platform's knowledge representation a comprehensive ontology infrastructure has been created, containing a core ontology (DOLCE [3]), two multimedia ontologies describing both the multimedia structure and the multimedia visual characteristics [11] and a domain ontology that contains information of a specific real-world domain, i.e. sports like tennis.

KAA begins with segmenting the input image into a number of regions, based on color homogeneity and shape regularity. These regions initialize the construction of an Attributed Relational Graph (ARG) which is the core of KAA's architecture. Figure 1 illustrates KAA's general architecture scheme, where the ARG is positioned in the center and interacts with several other processes.

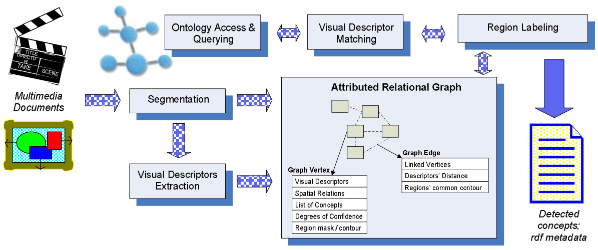


Figure 1. KAA general architecture

An ARG is a graph structure that holds the region-based representation of the image during the analysis process. An ARG is defined precisely by spatial entities represented as vertices V, each labeled with an attribute a and, binary spatial relationships represented as pairs of vertices $V \times V$ each labeled with a spatial descriptor w :

$$ARG \triangleq \langle V, a, w \rangle$$

where: V is the set of graph's vertices, $a: V \to A$, $w: V \times V \to W$ and W is the set of all spatial descriptors. **(1)**

In this particular work, A and W are complex structures defined as:

$$A: \{\text{Dominant Color,Region Shape,Homogeneous Texture}\}$$

and $W: \{\text{spatial relations}\}$

Each vertex of the graph corresponds to a region and holds the Dominant Color, Region Shape and Homogeneous Texture MPEG-7 visual descriptors extracted for this specific region. In a similar fashion, each graph's edge corresponds to the spatial adjacency of regions and is specified by one of the following relative and/or absolute spatial relations: *BelowOf, AboveOf, RightOf, LeftOf, AboveAll* and *BelowAll.*

The next step is to compute a matching distance value between each one of these regions and each one of the prototype instances of all concepts in the domain ontology. These concepts are linked with corresponding prototypical instances of the Visual Descriptor Ontology [11] descriptors (i.e. the abovementioned three MPEG-7 descriptors) and on this basis, a matching distance is computed by means of low-level visual descriptors. In order to combine Dominant Color, Region Shape and Homogeneous Texture in a unique matching distance, we use weighted average function with different weights for each concept that provides the required distance weighting. This combined distance is normalized and transformed to a degree of confidence, whereas a threshold to eliminate those labels that have a small degree is applied, keeping only those that have a strong belief of being correct. The threshold value varies for each domain allowing incorrect labels to be assigned to a region, for the benefit of retaining in all cases the correct label. Consequently, KAA serves as initial region labeling, i.e. it links a region with a possible label along with a fuzzy degree of confidence. At this point we propose an additional step that manipulates and improves the resulted list of labels taking into account accompanied contextual information.

3. VISUAL CONTEXT OPTIMIZATION

3.1 Contextual Knowledge Model and Representation

The idea behind the use of contextual information responds to the fact that not all human acts are relevant in all situations and since context is a difficult notion to grasp and capture, we restrict it to the notion of *ontological context*. The latter is defined as a "fuzzified" version of traditional ontologies. From now on we shall use the simplified term context for it. The problems to be addressed include how to represent this context, how to determine it, and how to use it to influence the results of knowledge-assisted analysis.

One possible way to describe ontologies is formalized as:

$$O = \left\{C, \{R_{c_i,c_j}\}\right\}, i,j = 1 \ldots n, R_{c_i,c_j} : C \times C \to \{0,1\} \quad \textbf{(2)}$$

where O is an ontology, C the set of concepts it describes and R_{c_i,c_j} the semantic relation amongst two concepts $c_i, c_j \in C$. Although ontologies may contain any type of relations, only ordering relations are of our interest. In order to extract and use the desired ontological context, we define it in the means of fuzzy ontological relations. Thus, the main challenge of this work is the meaningful exploitation of information contained in these taxonomic relations within the ontology framework. To tackle this type of relations we propose a "fuzzified" definition of the ontology:

$$O_F = \left\{C, \{r_{c_i,c_j}\}\right\},\ i,j = 1..n$$
$$F(R_{c_i,c_j}) = r_{c_i,c_j} : C \times C \to [0,1] \quad \textbf{(3)}$$

where O_F forms a domain-specific "fuzzified" ontology, C is the set of all possible concepts it describes and

$F\left(R_{c_i,c_j}\right) = r_{c_i,c_j}$ denotes a fuzzy relation amongst two concepts c_i, c_j. This "fuzzified" definition of the knowledge model is the most suitable for the modeling of real life information, which without doubt is governed by uncertainty and fuzzy relations.

The proposed contextual knowledge representation is easy to implement and provides all means to exploit contextual information within the image analysis framework we have set. We decided to model our contextual knowledge following a standardized language as RDF, introducing at the same time novel Semantic Web characteristics, such as reification [14] to obtain enhanced functionality. Finally, this concludes to only one possible way of representing context, that demands creation of new domain ontologies, as context demands a much richer set of relations and much more complex modeling primitives, than typical classification hierarchies provide.

Our context model is basically 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, which represents the fuzziness within the context model. Representing the graph model in RDF is a straight forward task; RDF is based upon a graph model, so that graph representation of contextual knowledge can naturally be represented in RDF. The second issue needing our attention is how to represent the degree of confidence associated with each relation. Describing the additional degree of confidence can be carried out using "manual" reification, i.e. making a statement about the statement, which contains the degree information. Moreover, having a statement, such as "*Car inContextOf Motorsports*" and a degree of confidence of 0.8 for this statement, does obviously not entail, that a car is always in the context of a motorsports scene. Consequently, representing fuzziness with reification seems to be a natural way, as the reified statement should not be asserted automatically.

This knowledge model is able to represent any type of real-life fuzzy relation between concepts $F\left(R_{c_i,c_j}\right) = r_{c_i,c_j}$. The "fuzzified" ontology is composed by a number of relations that cover every possible concept. In the current approach, all relations between concepts are represented in the ontology, each one of them forming a part of the overall contextual knowledge. We focus on a representative example, i.e. the *partOf* relation, *P*. The *partOf* relation, *P*, is a fuzzy partial ordering on the set of concepts. $P(x,y)>0$ means that *y* is part of *x*. For example *x* could be *tennis* and *y* could be a tennis *ball*. Its representation is presented in Figure 2; we provide two concepts and supposing an RDF namespace &dom following sample part of RDF representation is produced:

```
<rdf:Description rdf:about="#s1">
    <rdf:subject rdf:resource="&dom;tennis"/>
    <rdf:predicate rdf:resource="&dom;partOf"/>
    <rdf:object>rdf:resource="&dom;ball"</rdf:object>
    <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns
      #Statement"/>
    <context:partOf rdf:datatype="http://www.w3.org/2001/XMLSchema
      #float">0.85</context:partOf>
</rdf:Description>
```

Figure 2. RDF representation of the *partOf* relation

Other fuzzy semantic relations considered in the framework of this paper are: *SpecializationOf* (*Sp*), *inContextOf* (*Ct*), *InstrumentOf* (*Ins*), *PatientOf* (*Pat*), *LocationOf* (*Loc*) and *PropertyOf* (*Pr*). Each of these relations is represented by a specific layer of the overall "fuzzified" ontology built according to the guidelines already presented, in order to express all relationships between participating concepts belonging to the specific tennis sub-domain selected.

3.2 Context-based Confidence Value Readjustment

The proposed algorithm readjusts the initial label confidence values produced by KAA, utilizing the above presented contextual information residing in the constructed ontology. The approach followed is based on the formal methodology founded on the structures and semantics presented so far, as well as the mathematical notation summarized in [6]. In principal, we aim to re-adjust the confidence value $d_{l_k}^{g_q} \in [0,1]$ of each detected label l_k associated to a region g_q in a scene. Each label l_k is related to a specific concept c_k present in the application-domain's ontology, stored together with its relationship degrees r_{c_k,c_j} to any other related concept c_j.

To tackle cases that more than one concept is related to multiple concepts, the term *context relevance* cr_{c_k} is introduced, which refers to the overall relevance of concept c_k to the *root element* characterizing each domain. For instance the root element of tennis domain is concept c_{tennis}. All possible routes in the graph are taken into consideration forming an exhaustive approach to the domain, with respect to the fact that routes between concepts may be reciprocal. Estimation of each concept's value is derived from direct and indirect relationships of the concept with other concepts, using a meaningful *compatibility indicator* or distance metric. Depending on the nature of the domains under consideration, the best indicator could be selected among various operators found in the literature, like the *max* or the *min* operator. The ideal distance metric for two concepts is one that quantifies their semantic correlation. Given the tennis domain, the *max* value is a meaningful measure of correlation. A representative sample of the contextualized *tennis* ontology is presented in Figure 3; it represents relation *inContextOf* (*Ct*).

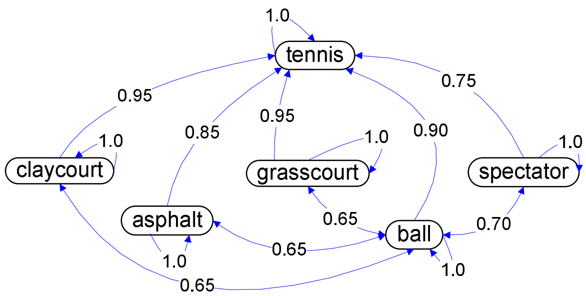


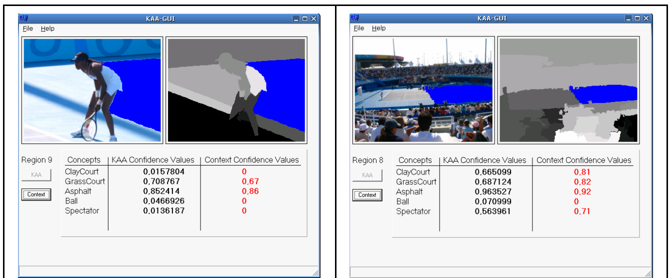
Figure 3. A fragment of the *tennis* ontology.

According to the context-based confidence value readjustment algorithm [6], we proceed by identifying as the optimal normalization parameter to use, according to the considered domains. We also define the minimum considerable value of an initial confidence value $d_{l_k}^{g_q}$, with respect to the particular classification information of the scene. For each label accompanied by a confidence value above the latter, we examine the domain ontology and identify the concept in the domain that is related to it. Then for each identified concept we obtain the particular contextual information in the form of its fuzzy relations to the set of any other concepts and calculate the new labeling confidence value $d_{l_k}^{g_q}$ of the label associated to the region, based on the normalization parameter and the context's relevance value. The overall process is terminated when belief to the labeling output provided by KAA is not strong enough, i.e. there are no more labels l_k with an acceptable initial confidence value $d_{l_k}^{g_q}$ above the specified initial confidence value. The result of this contextualization step is the meaningful readjustment of the initial label confidence values produced by KAA, optimizing the efficiency and robustness of the proposed image analysis methodology.

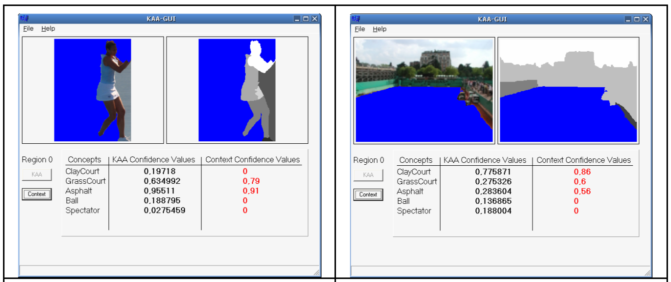
4. EXPERIMENTAL RESULTS

We conducted experiments in the tennis domain utilizing 262 images in total. In order to demonstrate our methodology, we have integrated the described techniques in one simple application developed in Java, utilizing a user-friendly graphical interface. This application illustrates both steps of knowledge-assisted analysis and contextual optimization. In the following we present a set of four application examples (Figure 4).

In the first tennis example (Figure 4a), the selected court region is successfully identified, although initial knowledge-assisted analysis results are confused by its colour. However, context successfully determines *asphalt* and corrects the misleading behavior by reducing the confidence value of *grasscourt*. In Figure 4b KAA results identify the court region correctly, however detected confidence values are confusing. Context gives a clear priority to *asphalt*, but given the distance and diversity of the shot (i.e. a mixture of spectators and terrain), maintains all four possibilities. In Figure 4c, although context slightly decreases the confidence value for *asphalt* in favor of *grasscourt*, because of the color of the court, it also extinguishes the incorrect *ball* option. Continuing, in Figure 4d we observe successful detection of dominant *claycourt* and simultaneous extinction of misleading *ball* and *spectator* labels.



(a) (b)



(c) (d)

Figure 4. Experimental results for the *tennis* domain

5. CONCLUSIONS

The methodologies presented in this paper can be exploited towards the development of more intelligent and self-confident image analysis environments. In order to verify their efficiency when faced with real-life data, we have implemented a research application and tested them thoroughly. The core contribution of the overall approach has been the provision of a novel RDF-based representation for visual context information, utilized in the process of image analysis to optimize its initial results. This is accomplished by basing our efforts on a novel ontological knowledge model, utilizing contextual information and RDF representations of fuzzy ordering relations, thus forming an interesting perspective to knowledge-assisted analysis.

REFERENCES

[1] Th. Athanasiadis et al., "Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content", SemAnnot '05, 2005.

[2] L. Biederman et al., "Scene perception: Detecting and judging objects undergoing relational violations", Cogn. Psychol. Vol. 14, pp. 143–177, 1982.

[3] A. Gangemi et al., "Sweetening Ontologies with DOLCE", 13th EKAW, 2002.

[4] J.M. Henderson and A. Hollingworth,. "High level scene perception", An. Rev. Psychol., vol. 50, pp. 243–271, 1999.

[5] C.-Y. Lin et al, "VideoAL: A novel end-to-end MPEG-7 semantic video automatic labeling system", ICIP, 2003.

[6] Ph. Mylonas et al., "Improving image analysis using a contextual approach", WIAMIS, 2006.

[7] Ph. Mylonas and Y. Avrithis, "Context modeling for multimedia analysis and use", CONTEXT '05, 2005.

[8] S.E. Palmer, "The effects of contextual scenes on the identification of objects", Memory and Cogn., vol. 3, pp. 519–526, 1975.

[9] P.G. Schyns A. Oliva, "From blobs to boundary edges: evidence for time and spatial scale dependent scene recognition", Psychol. Science, vol. 5, pp. 195–200, 1994.

[10] T. Sikora, "The MPEG-7 Visual standard for content description - an overview", IEEE Trans. on Circuits and Systems for Video Technology, 2001.

[11] N. Simou et al., "A Visual Descriptor Ontology for Multimedia Reasoning", WIAMIS, 2005.

[12] S. Staab, R. Studer, "Handbook on Ontologies", International Handbooks on Information Systems, Springer-Verlag, Heidelberg, 2004.

[13] A. Torralba, "Contextual influences on saliency", "Neurobiology of attention", 2005.

[14] W3C, RDF Reification, http://www.w3.org/TR/rdf-schema/#ch_reificationvocab.