IMPROVING IMAGE ANALYSIS USING A CONTEXTUAL APPROACH

Phivos Mylonas, Thanos Athanasiadis and Yannis Avrithis
School of Electrical and Computer Engineering
National Technical University of Athens
9, Iroon Polytechniou Str., 157 73 Zographou, Athens, Greece

ABSTRACT

Generic algorithms for automatic object recognition and/or scene classification are unfortunately not producing reliable and robust results. A common approach to cope with this still unresolved issue is to restrict the problem at hand to a specific domain. In this paper we propose an algorithm to improve the results of image analysis, based on the contextual information we have, which relates the detected concepts to any given domain. Initial results produced by the image analysis module are domain-specific semantic concepts and are being re-adjusted appropriately by the suggested algorithm, in the means of fine-tuning the degrees of confidence of each detected concept. The novelty of the presented work is twofold: i) the knowledge-assisted image analysis algorithm, that utilizes an ontology infrastructure to handle the knowledge and MPEG-7 visual descriptors for the region labeling and ii) the context-driven re-adjustment of the degrees of confidence of the detected labels.

1. INTRODUCTION

It is common knowledge that the lack of machine generated but human understandable, high level indexing mechanisms, that produce content description in a conceptual level, degrades the importance of digital multimedia content itself. State-of-the-art image analysis systems [5] are limiting themselves by resorting mostly to visual descriptions at a very low level, such as dominant color. The MPEG-7 standard [7] provides functionalities for management of multimedia content and metadata, but it lacks on the extraction of semantic description and annotation.

We use the term knowledge assisted analysis when image analysis algorithms and ontological representation of both general and domain specific knowledge are tightly coupled and there is a constant interaction between them. Ontologies [4] express key entities and relationships of multimedia content in a formal machine-processable representation and can help to bridge the semantic gap [9, 11] 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 called KAA [2], 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, KAA's output is in many cases not sufficient for the understanding of multimedia content. In the approach followed herein, we introduce a methodology for improving the results of KAA, based on contextual information obtained from application-specific domain ontologies. The main effort of this work is spent on readjusting KAA labeling information derived from the application of several classification steps on the considered scenes. A context-based labeling update algorithm is also introduced; this algorithm describes the process of readjusting the labeling information obtained from the classification step of a specific image scene, utilizing higher level contextual knowledge available. The overall methodology forms the basis on top of which ontologies can be exploited within image analysis.

2. KNOWLEDGE ASSISTED ANALYSIS

In the process of performing efficient image analysis, we developed a test-bed application called KAA, whose architecture and functionality is described briefly in this section. For KAA'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 [8] and three domain ontologies that model the content layer of multimedia with respect to specific real-world domains, i.e. sports like tennis and holidays at the beach or the mountains.

KAA includes methods that automatically segment images into areas corresponding to salient semantic objects (e.g. persons, sea, sailing boats, etc.) and provide a flexible infrastructure for further analysis as for instance object recognition, metadata generation and indexing. In this work we focus mainly on the recognition functionality of KAA, which is done by means of semantic labeling of the detected objects. A more precise description of the KAA general architecture scheme is given in Figure 1. The core of the architecture is defined by the region adjacency graph. This graph structure holds the region-based representation of the image during the analysis process. During image analysis, a set of atom-regions is generated by an initial segmentation. Each vertex of the graph corresponds to an atom-region and holds the Dominant Color and Region Shape MPEG-7 visual descriptors extracted for this specific region.

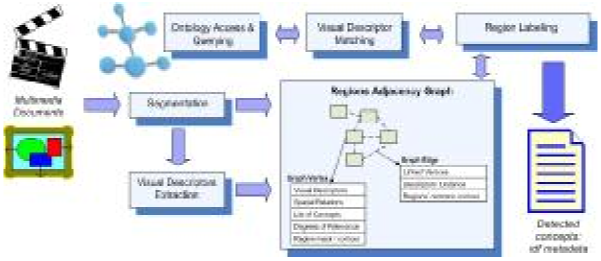

Figure 1 KAA-architecture

The next step for the analysis is to compute a matching distance value between each one of these atom-regions and each one of the prototype instances of all concepts in the domain ontology. This matching distance is evaluated by means of low-level visual descriptors. In order to combine Dominant Color and Region Shape in a unique matching distance, we use a neural network approach [10] that provides us with 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.

The objective of this knowledge-based analysis, is to extract high level, human comprehensible features and create automatically semantic metadata describing the multimedia content itself. For each image KAA produces an RDF file that contains a sequence of elements, one for each region/graph vertex. Each element includes a list of labels (candidate concepts) with their degree of confidence and, additionally, information about the spatial relations with other regions. One could read this RDF and use it directly as semantic annotation by associating the specific image to the number of detected concepts. That is, an image is described by the detected objects, each one of those is linked to a list of possible labels and each one along with a 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. CONTEXT-BASED LABELING UPDATE ALGORITHM

3.1. Knowledge Structure and Representation

Let us present the problem that this work attempts to address, in a more formal manner. Our algorithm readjusts in a meaningful way the initial label confidence values produced by KAA. In designing such an algorithm, contextual information residing in the ontology is utilized. In general, the notion of context is strongly related to the notion of ontologies since an ontology can be seen as an attempt for modeling real world (i.e. fuzzy) entities and context determines the intended meaning of each concept, i.e. a concept used in different context may have different meanings. Consequently, one possible way to extract and use the context is to define it in the means of fuzzy ontological relations.

Although ontologies may contain any type of relations, only taxonomic (i.e. ordering) relations and spatial relations are of our interest. As discussed in [1], the use of ordering relations is necessary for the determination of the document's context. Thus, the main challenge of this work is the meaningful exploitation of information contained in these taxonomic relations within the ontology. Fuzzy relations are suitable for representing such real life information. On the other hand, depending on the requirements of the application, the set of spatial relationships can be rich (many spatial relationships with minor differences between each other) or sparse (fewer distinct relationships). A rather complete set of semantic spatial relationships, enhanced by fuzzy degrees for greater accuracy, can be modeled as: ***above, far_above, below, far_below, beside, enclosed, enclosing*** [6].

Consequently, to tackle both types of relations we introduce a "fuzzified" definition of an ontology-based knowledge model: $O_F = \left\{ C, \left\{ r_{c_i,c_j} \right\} \right\},\ i, j = 1..n,\ i \neq j$ and $F(R_{c_i,c_j}) = r_{c_i,c_j} : C \times C \to [0,1]$, 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.

3.2. Mathematical Expressions

In the following let us agree on the mathematical notation used herein:

- $RG = \left\{ g_q \right\},\ q = 1..p,\ q \in \mathbb{N}$: the set of all regions/segments in the scene, where $p \equiv |RG|$.
- $L = \left\{ l_k \right\},\ k = 1..n,\ k \in \mathbb{N}$: the set of all possible labels associated to the scene under consideration, where $n \equiv |L|$.
- $L^{g_q} = \{ l_k \} \subseteq L$, where $k, q \in \mathbb{N}$: the set of the detected labels associated to one particular region g_q of the scene.
- $d_{g_q,l_k},\ g_q \in RG,\ l_k \in L^{g_q}$: the confidence value of each label l_k produced by KAA assigned to a particular region g_q of the scene.
- $C = \left\{ c_k \right\},\ k = 1..m,\ k \in \mathbb{N}$: the set of all possible concepts included in the ontology representation. $m \in \mathbb{N},\ m \equiv |C|$.

In this first implementation phase of our approach a "1-1" mapping between labels and concepts is assumed, i.e. $n = m$.

• r_{c_i,c_j}, $i,j \in [1,m]$: fuzzy relation degree value between any two concepts $c_i, c_j \in C$ stored in the ontology.

The proposing algorithm aims to re-adjust the belief value d_{g_q,l_k} 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. To tackle cases that more than one concept is related to multiple concepts, we introduce the term context relevance cr_{c_k} which refers to the overall relevance of concept c_k to the "root element" of the domain. Current approach aggregates each concept's values obtained i) from direct relationships of the concept with other concepts and ii) indirect relationships, calculating the maximum value of all. An example domain ontology is depicted in Figure 2:

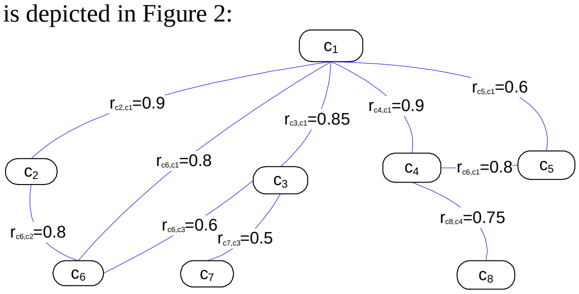

Figure 2 A fragment of a domain ontology. Concept c_1 is the "root element" of the domain, in most cases characterizing the domain (e.g. *beach*)

Letting concept c_6 be related to concepts c_1, c_2 and c_3 directly with: $r_{c_6,c_1} = 0.8$, $r_{c_6,c_2} = 0.8$ and $r_{c_6,c_3} = 0.6$, while concept c_2 is related to concept c_1 with $r_{c_2,c_1} = 0.9$ and concept c_3 is related to concept c_1 with $r_{c_3,c_1} = 0.85$, we calculate the value for cr_{c_6} as follows:

$$cr_{c_6} = \max\{r_{c_6,c_1}, r_{c_6,c_2} \cdot r_{c_2,c_1}, r_{c_6,c_3} \cdot r_{c_3,c_1}\} = \max\{0.8, 0.8 \cdot 0.9, 0.6 \cdot 0.85\} = 0.8$$

3.3. Label Confidence Re-adjustment Algorithm

The general structure of the confidence re-evaluation algorithm, adjusted for the needs of the problem at hand, is as follows:

1. Identify an optimal normalization parameter np to use within the confidence re-evaluation algorithm, according to the considered domain(s). The np is also referred to as domain similarity, or dissimilarity, measure and $np \to [0,1]$.
2. Define a threshold T for the minimum considerable value of an initial confidence value d_{g_q,l_k}, with respect to the particular classification information of the scene.
3. For each label l_k accompanied by a confidence value d_{g_q,l_k} above T examine the supplied domain ontology and identify the concept c_k in the domain that is related to l_k.
4. For each identified concept c_k in the considered domain, obtain the particular contextual information in the form of its relations to the set of any other concepts $C - \{c_k\}$: r_{c_k,c_j}.
5. Calculate the new labeling confidence value d'_{g_q,l_k} of label l_k associated to region g_q, based on np and the context's relevance value. In the case of multiple concept relations in the ontology, relating concept c_k to more than one concepts, rather than relating c_k solely to the "root element" c_1, as described already in Figure 2, an intermediate aggregation step should be applied for c_k: $cr_{c_k} = \max\{r_{c_k,c_1}, .., r_{c_k,c_m}\}$. Finally: $d^t_{g_q,l_k} = (1-np)^t \cdot d^0_{g_q,l_k} + \left(1-(1-np)^t\right) \cdot cr_{c_k}$, where t denotes the iteration parameter used and $d^0_{g_q,l_k}$ represents the original confidence value obtained by KAA. For $t = 0$, formula 1 degrades to the identity formula and the initial confidence value are propagated without being re-adjusted by the algorithm, whereas for $t = 1$ only one iteration is considered and formula 1 is transformed as: $d^1_{g_q,l_k} = d^0_{g_q,l_k} - np \times \left(d^0_{g_q,l_k} - cr_{c_k}\right)$. Typical values for t reside between 3 and 5.

Key points in this approach are the identification of the related concepts at step 3, the definition of a meaningful normalization parameter np and the identification of the optimal threshold T for the initial confidence values. When re-evaluating these values, the ideal np is always defined with respect to the particular domain of knowledge and is the one that quantifies their semantic correlation to the domain. The overall process should terminate 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_{g_q,l_k} above the specified threshold T.

4. RESULTS

We conducted experiments in the domains of beach, mountain and tennis, utilizing 95 images. Results are very promising and even in cases where detection of specific labels is rather difficult, system's performance can be initially measured by the associated degree of confidence for each label. In other words the probability of the fact that the detected label indeed describes correctly the image (or part of the image) is produced.

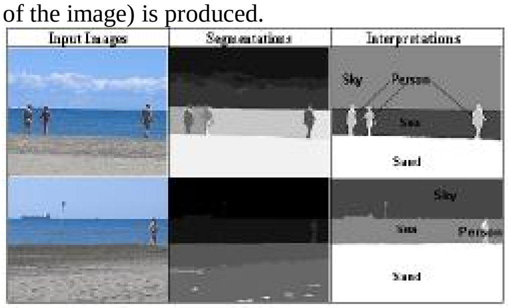

Figure 3 Holiday-results

Moreover, initial KAA results, as illustrated in Figure 3, include a segmentation mask outlining the semantic description of the scene. The different colors assigned to the generated atom-regions correspond to the object classes defined in the domain ontologies, allowing the user a visual control of the results. The proposing algorithm is then applied on the labels and in the following we present summarized results for two representatives, i.e. *sky* and *person*, derived initially from the beach and tennis domains. Initial degrees of confidence are provided, np is set to 0.15 for both domains and the acceptable threshold T used has a value of 0.20. Assuming that concepts *sky* and *person* are present in three different domains, i.e. beach, mountain and tennis, there is a relevance value for each one in every domain ontology. Thus, the re-evaluated KAA degrees of confidence are computed accordingly, whereas the optimal iteration value of the algorithm is considered to be a value of 3.

label l_k	degree d_{g_q,l_k}	domain	concept c_k	degree cr_{c_k}	$d^{t=3}_{g_q,l_k}$
sky	0.93	beach	sky	0.85	**0.899**
		mountain		0.80	**0.880**
		tennis		0.30	**0.687**
person	0.65	beach	person	0.50	**0.592**
		mountain		0.55	**0.611**
		tennis		0.80	**0.708**

Table 1. Application examples of proposed algorithm.

In the first example we consider that KAA performs well in all three cases, and suggests a 93% confidence on the detected region for *sky*. However in the different domain ontologies, different contextual relations exist for *sky* and thus initial KAA degrees are influenced in a different manner. In all three domains, contextual relationships introduce smaller -than KAA's 93%- values for the concept *sky*, resulting into lowering the initial degrees of confidence. Since the first beach domain introduces a 0.85 degree of relevance to sky, degradation of confidence value is considered to be small, i.e. only 0.031, resulting to a re-evaluated value of 0.899 instead of 0.93. Moreover, in the mountain domain, we encounter *sky* with a degree of relevance of 0.8, thus overall degree of confidence is lowered by 0.05 to 0.88. Third domain tennis results into a 0.687. In the second example, the classification label suggests a 0.65 confidence on a *person* for a detected region in the three scenes. Information obtained from the ontologies introduces a set of three contextual relations, varying from 0.50 to 0.80. Thus, initial confidence values are readjusted to a set of three new values, each one appropriate for the particular domain, as illustrated in the last column of Table 1.

5. CONCLUSIONS

Evaluation of the proposing context-based labeling update algorithm, based on real-life data was fulfilled, as well as evaluation and improvement of the feasibility and performance of KAA. An outline is presented for exploiting the contextual knowledge in order to re-adjust the region labeling procedure and improve its performance.

Acknowledgment. This research was partially supported by the European Commission under contract FP6-001765 aceMedia.

REFERENCES

1. Akrivas G. et al., Semantic Association of Multimedia Document Descriptions through Fuzzy Relational Algebra and Fuzzy Reasoning, IEEE Trans. on Systems, Man, and Cybernetics, March 2004.
2. Athanasiadis T. et al., Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content, Proc. of SemAnnot '05, November 2005
3. Gangemi A. et al., Sweetening Ontologies with DOLCE, Proc. of the 13th EKAW, 2002.
4. Hunter J. et al., Realizing the hydrogen economy through semantic web technologies, IEEE Intelligent Systems Journal, 2004
5. Lin C.-Y. et al, "VideoAL: A novel end-to-end MPEG-7 semantic video automatic labelling system," in Proc. IEEE Int. Conf Image Processing, Barcelona, Spain, Sep. 2003.
6. Mylonas P. et al., Context modelling for multimedia analysis and use, Proc. of 5th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT '05), July 2005.
7. Sikora T., The MPEG-7 Visual standard for content description - an overview, IEEE Trans. on Circuits and Systems for Video Technology, June 2001
8. Simou N. et al., A Visual Descriptor Ontology for Multimedia Reasoning, Proc. of WIAMIS '05, April 2005
9. Smeulders A.W.M. et al., Content-based image retrieval at the end of the early years, IEEE Trans. on Pattern Analysis and Machine Intelligence, December 2000
10. Spyrou E. et al., Fusing MPEG-7 visual descriptors for image classification, Proc. of ICANN 05, September 2005
11. Zhao R. et al, Narrowing the Semantic Gap – Improved Text-Based Web Document Retrieval Using Visual Features, IEEE Transactions on Multimedia, 2002