Multimedia Reasoning with f-$\mathcal{SHIN}$

N. Simou, Th. Athanasiadis, V. Tzouvaras, S. Kollias
Image Video and Multimedia Systems Laboratory,
School of Electrical and Computer Engineering,
National Technical University of Athens,
Iroon Polytexneiou 9, 15773 Zografou, Greece
{nsimou,thanos,tzouvaras,stefanos}@image.ntua.gr

Abstract

Effective management and exploitation of multimedia documents requires extraction of the underlying semantics. Multimedia analysis algorithms can produce fairly rich but imprecise information about a multimedia document. In this paper, a multimedia reasoning architecture is presented using the fuzzy extension of expressive $\mathcal{SHIN}$, f-$\mathcal{SHIN}$. First a segmentation algorithm generates a set of over-segmented regions and a classification process is employed to assign those regions with semantic labels. A semantic-based refinement of the segmentation is follows and this information initializes the ABox of a fuzzy-knowledge that is used for multimedia reasoning. The proposed approach was tested on outdoor domain and shows promising results.

1 Introduction

Moving from low-level perceptual features to high-level semantic descriptions that match human cognition is the final frontier in computer vision, and consequently to any multimedia application targeting efficient and effective access and manipulation of the available content. The early efforts targeting this so called *semantic gap* formed what is known as content-based (analysis and) retrieval approaches, where focus is on extracting the most representative numerical descriptions and defining similarity metrics that emulate the human notion of similarity. The limitations of such numerical-based methodologies however [9], led to the investigation of ways to enhance their performance.

To overcome the above limitations, more sophisticated semantic information extraction approaches are required to enable more efficient manipulation and retrieval of visual media. Although existing multimedia standards, like MPEG-7 [8], provide important functionalities such as manipulation and transmission of objects and metadata, their initial extraction, and that most importantly at a semantic level, is out of the scope of the standards and is left to the content developer.

In the last decade a substantial amount of work has been carried out in the context of Ontologies and Description Logics (DLs). DLs are logical reconstruction of the so called frame-based knowledge representation languages, with the aim of providing a simple well-establishing declarative semantics to capture the meaning of the most popular features of structured representation of knowledge. A main point is that DLs are considered to be attractive in multimedia applications as they form a good compromise between reasoning power and computational complexity. Experience in using DLs in applications has shown that in many cases we would like to extend the representational and reasoning capabilities of them. In particular, the use of DLs in the context of multimedia, points out the necessity of extending and using DLs with capabilities which allow the treatment of the inherent imprecision in multimedia object representation, retrieval and detection [11]. In fact, classical DLs are insufficient for describing multimedia situations since retrieval matching and detection are not usually situation of *true* or *false*.

In this paper, we present a hybrid system composed of a Knowledge Assisted Analysis (KAA) module and a fuzzy reasoning engine. The KAA module includes a semantic segmentation approach based on the RSST algorithm [7]. The fuzzy reasoning engine has been constructed on the basis of DL f-$\mathcal{SHIN}$ [10] which is the fuzzy extension of expressive $\mathcal{SHIN}$ [5]. By reasoning in this context, we refer to the automatic derivation of high-level semantic annotations from low-level multimedia data (raw and/or preprocessed to acquire audiovisual or conceptual descriptions of varying abstraction levels) through the utilization of the provided (general, domain, structural, etc.) knowledge.

The rest of the paper is organized as follows. The next section presents the architecture of the proposed hybrid approach. In section 3, the KAA module is presented and the S-RSST algorithm is outlined. In section 4, the semantics of

the f-$\mathcal{SHIN}$ in the context of multimedia reasoning is presented. Finally, in the last section, the experimental results are demonstrated.

2 Reasoning Architecture

In this section we describe the overall architecture for the proposed multimedia reasoning (see Figure 1). An image or a video frame is initially processed by the knowledge -assisted analysis module (KAA), which outputs a segmentation mask together with region-associated labels and degrees of confidence. KAA module processes the input image in three basic steps, namely: i) through an SVM classifier [4], ii) a semantic-based segmentation algorithm called Semantic-RSST [2] and iii) a region merging step. In summary, the approach for knowledge- assisted analysis is as follows: A segmentation algorithm [1] partitions the input image in a number of regions that may have symbolic interpretation. During the classification process, low level visual features are extracted from those regions feeding a Support Vector Machine (SVM) classifier. In this phase, initial classification takes place by assigning a set of semantic label and confidence value pairs to each segment. Classification results are refined by the application of a novel semantic-based segmentation algorithm, which targets to refine the initial labels and the segmentation mask. Finally, neighbor regions that share some common semantic labels and meet certain criteria are merged to form a more meaningful segmentation of the image.

The output of KAA, i.e. the region-associated semantic labels and degrees of confidence, is used by FiRE[1] as the fuzzy assertion component (ABox) of the knowledge base. FiRE is a fuzzy reasoning engine based on the expressive DL f-$\mathcal{SHIN}$ [10]. It is a tableaux implementation for fuzzy logics, i.e. logics where truth values are taken from the interval [0,1], allowing handling of imprecise information and extraction of implicit knowledge. The segments of the image are represented as DL-individuals participating in the domain concepts to a given degree. The terminology (TBox) is defined by using the domain concepts to declare more general and complex concepts.

Using such a representation, implicit knowledge about segments can be extracted. This inferred knowledge either assigns them to higher concepts or corrects possibly mistaken labels (Section 5) assigned by KAA. According to this information, a merging process takes place which merges the updated segments producing a new segmentation mask. For example in Figure 1 regions 1, 2, and 3 of the segmented image could be segmented to new upper concept $ComplexBuilding$. The resulting metadata are again

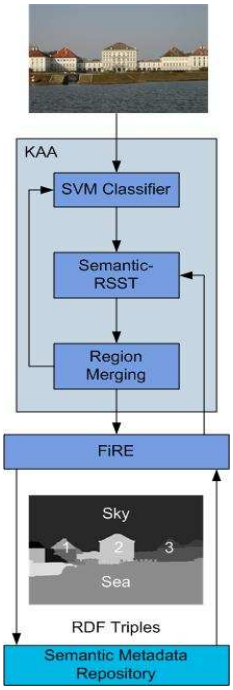

Figure 1. The multimedia reasoning architecture

used as the fuzzy assertion component by FiRE until termination criteria are met and a sufficient number of segments (semantically meaningful) is reached for the particular image.

Finally, FiRE communicates with a semantic repository, such as Sesame[2], permitting storage, querying and reasoning with RDF and RDF Schema. The resulting metadata are translated to RDF and stored to a Sesame repository by FiRE. By that way, the extracted implicit knowledge can be available to other modules through the sesame repository.

3 Knowledge-Assisted Analysis

Knowledge-assisted analysis, in the context of this work, deals with the very important and difficult task to create the link between multimedia content and concepts stored in a knowledge-base, e.g. a concept that has been detected in a specific (spatiotemporal) location of an image (video). Due to the difficulty of this task, results are prone to many errors both in detecting semantic objects and in recognizing them among many possible others.

[1] FiRE can be found at http://www.image.ece.ntua.gr/∼nsimou together with installation instructions and examples.

[2] http://www.openrdf.com

3.1 Initial Classification

In this paper we use Knowledge-Assisted Analysis (KAA) module to form the missing link and create the primary semantic metadata. KAA consists mainly of three processes: An initial (over)segmentation, descriptor extraction and region classification. Initially, a segmentation algorithm, based on low-level features such as color and texture [1], is applied in order to divide the given image into regions and corresponding low-level descriptions are computed for every resulting region. The later are employed so as to form a compound low-level descriptor vector for every image region, based on a simple concatenation mechanism. The computed feature vector is employed for generating initial set of region's candidate semantic labels [4]. This process results to an initial fuzzy labeling of the regions with concepts from the knowledge base, i.e. for region a we have the fuzzy set (following the sun notation [6]) $L_a = \sum_k c_k/w_k$, where k is the cardinality of the (crisp) set of all concepts $C = \{c_k\}$ in the knowledge base and $w_k = \mu_a(c_k)$ is the degree of membership of element c_k in the fuzzy set L_a.

3.2 Semantic Segmentation

In this section we examine how a variation of a traditional segmentation technique, the Recursive Shortest Spanning Tree, also known as RSST [7], can be used to further improve the initial results. RSST is a bottom-up segmentation algorithm that begins from the pixel level and iteratively merges similar neighbor regions until certain termination criteria are satisfied. It uses internally a graph representation of image regions, like the Attributed Relation Graph (*ARG*) [3]. In the beginning, all edges of the graph are sorted according to a criterion, e.g. color dissimilarity of the two connected regions using Euclidean distance of the color components. The edge with the least weight is found and the two regions connected by that edge are merged. After each step, the merged region's attributes (e.g. region's mean color) is re-calculated. RSST will also re-calculate weights of related edges as well and resort them, so that in every step the edge with the least weight will be selected. This process goes on recursively until termination criteria are met. Such criteria may vary, but usually these are either the number of regions or a threshold on the distance.

We modify this algorithm to operate on the fuzzy sets in a similar way as if they worked on low-level features (such as color, texture, etc.). This variation follows in principles the algorithmic definition of the traditional RSST, though a few adjustments were considered necessary and were added. S-RSST aims to improve the usual oversegmentation results by incorporating region labeling in the segmentation process [2]. The modification of the traditional algorithm to S-RSST lies on the definition of the two criteria: (1) The dissimilarity criterion between two adjacent regions a and b (vertices v_a and v_b in the graph), based on which graph's edges are sorted and (2) the termination criterion.

For the calculation of the similarity between two regions, we examined two approaches. The first one is based on the definition of a metric between two fuzzy sets, those that correspond to the candidate concepts of the two regions. This dissimilarity value is computed according to the following formula and is assigned as the weight of the respective graph's edge e_{ab}:

$$w(e_{ab}) = 1 - \sup_{c_k \in C} (t - norm(\mu_a(c_k), \mu_b(c_k))) \qquad (1)$$

where a and b are two neighbor regions and $\mu_a(c_k)$ is the degree of membership of concept $c_k \in C$ in the fuzzy set L_a.

Let us now examine one iteration of the S-RSST algorithm. Firstly, the edge e_{ab} with the least weight is selected, then regions a and b are merged. Vertex v_b is removed completely from the ARG, whereas v_a is updated appropriately. This update procedure consists of the following two actions:

1. Re-evaluation of the degrees of membership of the labels fuzzy set in a weighted average (w.r.t. the regions' size) fashion.

2. Re-adjustment of the ARG edges by removing edge e_{ab} and re-evaluating the weight of the affected edges.

This procedure continues until the edge e^* with the least weight in the ARG is bigger than a threshold: $w(e^*) > T_w$. This threshold is calculated in the beginning of the algorithm, based on the histogram of all weights of the set of all edges.

4 Multimedia Reasoning with f-$\mathcal{SHIN}$

4.1 f-$\mathcal{SHIN}$ and Reasoning Services

f-$\mathcal{SHIN}$ is a fuzzy extension of DL $\mathcal{SHIN}$ [5]. As pointed out in the fuzzy DL literature [11, 10], fuzzy extensions of DLs involve only the *assertion* of individuals to concepts and the semantics of the new language. Hence, as usual we have an alphabet of distinct concept names (**C**), role names (**R**) and individual names (**I**). Then, f-$\mathcal{SHIN}$-concepts are inductively defined as follows,

1. If $C \in \mathbf{C}$, then C is a f-$\mathcal{SHIN}$-concept,

2. If C and D are concepts, R is a role and $n \in \mathbb{N}$, then $(\neg C)$, $(C \sqcup D)$, $(C \sqcap D)$, $(\forall R.C)$, $(\exists R.C)$, $(\geq nR)$ and $(\leq nR)$ are also f-$\mathcal{SHIN}$-concepts.

$\mathcal{T} = \{$PartOfBuildingComplex	$\equiv$	Building $\sqcap$ (($\exists$left $-$ of.Building) $\sqcup$ ($\exists$right $-$ of.Building)),
CloudedSky	$\equiv$	Cloud $\sqcap$ Sky,
Sand	$\equiv$	$\exists$below $-$ of.Sea,
Sky	$\equiv$	$\exists$above $-$ of.Sea,
Body	$\equiv$	Natural $-$ Person $\sqcap$ ($\exists$above $-$ of.Natural $-$ Person) $\sqcap$ ($\exists$below $-$ of.Natural $-$ Person)),
Leg	$\equiv$	Natural $-$ Person $\sqcap$ ($\exists$below $-$ of.Body),
Head	$\equiv$	Natural $-$ Person $\sqcap$ ($\exists$above $-$ of.Body),
PartOfHuman	$\equiv$	Body $\sqcup$ Head $\sqcup$ Leg$\}$
$\mathcal{R} = \{$Trans(above $-$ of), Trans(below $-$ of), above $-$ of $-$ adjacent $\sqsubseteq$ above $-$ of, below $-$ of$^-$ = above $-$ of,		
below $-$ of $-$ adjacent $\sqsubseteq$ below $-$ of, below $-$ of $-$ adjacent$^-$ = above $-$ of $-$ adjacent, left $-$ of$^-$ = right $-$ of$\}$		
Trans(right $-$ of), right $-$ of $-$ adjacent $\sqsubseteq$ right $-$ of, left $-$ of $-$ adjacent $\sqsubseteq$ left $-$ of, Trans(left $-$ of),		
left $-$ of $-$ adjacent$^-$ = right $-$ of $-$ adjacent, left $-$ of $-$ adjacent$^-$ = right $-$ of $-$ adjacent,		

Table 1. Knowledge Base ($TBox$)

Moreover, if R is a role then R^- is also a role, namely the inverse of R. Furthermore, DL concept axioms are of the form $C \equiv D$ or $C \sqsubseteq D$, where C, D are concepts, saying that C is equivalent or a sub-concept of D, respectively. A set of such axioms defines a $TBox$ ($\mathcal{T}$). Additionally, we can have role axioms of the form Trans(R) saying that R is transitive or $R \sqsubseteq S$ saying that R is a sub-role of S. A set of role axioms defines an $RBox$ ($\mathcal{R}$)

The semantics of fuzzy DLs are provided by a *fuzzy interpretation* [11, 10]. A fuzzy interpretation is a pair $\mathcal{I} = \langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ where the domain $\Delta^{\mathcal{I}}$ is a non-empty set of objects and $\cdot^{\mathcal{I}}$ is a fuzzy interpretation function, which maps an individual name a to elements of $\mathtt{a}^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ and a concept name A (role name R) to a membership function $\mathsf{A}^{\mathcal{I}} : \Delta^{\mathcal{I}} \rightarrow [0, 1]$

Hence a fuzzy knowledge base Σ is a triple $\langle \mathcal{T}, \mathcal{R}, \mathcal{A} \rangle$, where $\mathcal{T}$ is a fuzzy $TBox$, $\mathcal{R}$ is a fuzzy $RBox$ and $\mathcal{A}$ is a fuzzy $ABox$. $TBox$ and $RBox$ introduce the terminology i.e the vocabulary of the application domain while $ABox$ contains the assertions about named individuals in terms of this vocabulary.

The main reasoning services provided by crisp reasoners are *entailment* and *subsumption*. These services are also available by FiRE together with greatest lower bound queries which take the advantage of the fuzzy element. Fuzzy entailment queries ask whether an individual participates in a concept in a specific degree. Subsumption queries on the other hand ask whether a concept is sub-concept of another concept, e.g. $Head \sqsubseteq PartOfHuman$. Finally, since a fuzzy $ABox$ $\mathcal{A}$ might contain many positive assertions for the same individual (pair of individuals), without forming a contradiction, it is in our interest to compute what is the best lower and upper truth-value bounds of a fuzzy assertion. The concept of *greatest lower bound* of a fuzzy assertion w.r.t. Σ was defined in [11]. Greatest lower bound ask for the degree of participation of an individual in a concept.

4.2 The Fuzzy knowledge base

The extraction of implicit knowledge from explicit one requires an expressive terminology, which defines higher concepts. In our case, the holiday domain has been used and -using the concepts and the spatial relationships that the knowledge-assisted module evaluates for an input image- the following terminology has been defined.

The spatial relations which are extracted for a region, provide information about it's location relatively to the neighboring regions. These relations represent the roles in our terminology and form the following set:

$Roles = \{above-of, above-of-adjacent, below-of, below-of-adjacent, left-of, left-of-adjacent, right-of, right-of-adjacent\}$.

Similarly, the concepts that the knowledge -assisted analysis may estimate, are as follows:

$Concepts = \{Cliff, Foliage, Mountain, Lamp-Post, Ground, Statue, Road, Sand, Trunks, Natural-Person, Snow, Cloud, Wave, Tree, Stone, Sea, Dried-Plant-Snowed, Sailing-Boat, Sky, Building\}$

Using these sets we defined the $RBox$ with the fuzzy role axioms and the $TBox$ with the fuzzy concept axioms that are presented in table 1.

The concept axioms of $TBox$ were defined in order to assist the multimedia extraction of higher concepts. Hence, new expressive concepts have been defined such as

Region	Concept	Degree
$region_6$	Sky	0.70
$region_6$	Cloud	0.65
$region_6$	Sea	0.56
$region_8$	Road	0.62
$region_8$	Sea	0.58
$region_8$	Cliff	0.55

Table 2. Fuzzy concept assertions from an excerpt of ABox

$PartOfBuildingComplex$ which represents a $Building$ that is either left or right of another $Building$. Additionally, axioms which correct mistaken estimations of analysis have been declared. A definition of concepts Sky and $Sand$ was made, using spatial relations that require Sky to be above Sea, and $Sand$ to be below Sea.

The segmented image produced by the knowledge-assisted module is used as the assertion component of the knowledge base ($ABox$). Individuals are represented by the resulting segments, participating in all the elements of $Concepts$ to different degrees. Relations are defined analogously, using the elements of $Roles$ and crisp degrees. Tables 2 and 3 present some concepts and role assertions for a specific example, examined in more detail in the following section.

5 Results

In this section we present some partial results of the system's modules and of the overall architecture, given in a walk-through description of the process flow. This includes the output of the SVM classifier and the application of fuzzy reasoning and semantic RSST algorithms. We use a prototype graphical user interface (Figure 2), which assists both in the integration of the KAA and FiRE modules, as well as in the visual inspection of both intermediate and final results.

Subj Region	Role	Obj Region	Degree
$region_6$	$\mathsf{above-of}$	$region_8$	1
$region_2$	$\mathsf{left-of}$	$region_3$	1

Table 3. Fuzzy role assertions from an excerpt of ABox

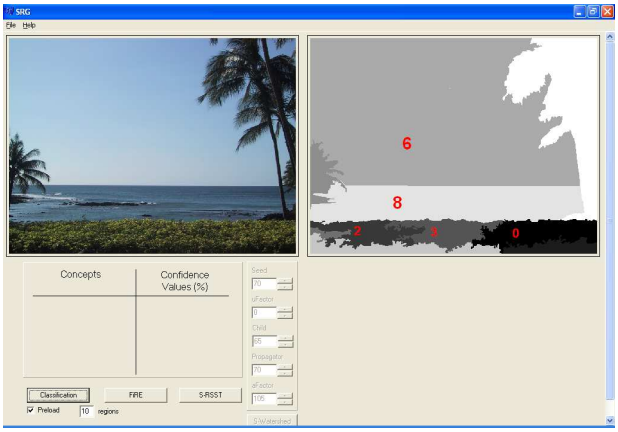

Figure 2. A beach photo, along with a segmentation mask

As described in section 3, an image is processed by the KAA module, which outputs a segmentation mask together with region-associated labels and degrees of confidence. For instance, Figure 3 depicts the classifier output for the blue highlighted region; it recognizes correctly this region as foliage with a degree of confidence of approximately 69%. The SVN classifier's output is in RDF format, and fuzzy degrees are inserted using the reification method. From the implementation point of view, a different format was selected, that of FiRE, which is more efficient and descriptive.

In order to demonstrate the importance of FiRE reasoning and its impact to the correction of regions' classification, take as example the sea region (region 8 in Figure 2). SVM misclassified it as road with a degree of 62% and as sea to a (smaller) degree of 58% (see Table 2). Invocation of FiRE and more specifically activation of the axiom that a sea region lies below a sky region (Table 3), assigned a new (higher) degree of confidence (70%) for the sea region.

The gain of semantic RSST algorithm is the refinement of the segmentation mask. In the specific example, it is evident that initial segmentation failed to merge the different foliage regions (regions 0, 2 and 3 in Figure 2), although recognition worked reasonably well (with degrees of confidence spanning between 53%-77%). Figure 4 depicts the result of the semantic RSST algorithm, where all foliage segments have been correctly merged, due to their common semantic concept.

6 Conclusions

In this paper, an architecture for multimedia reasoning was presented. An image is processed by knowledge-assisted analysis module which produces a segmentation

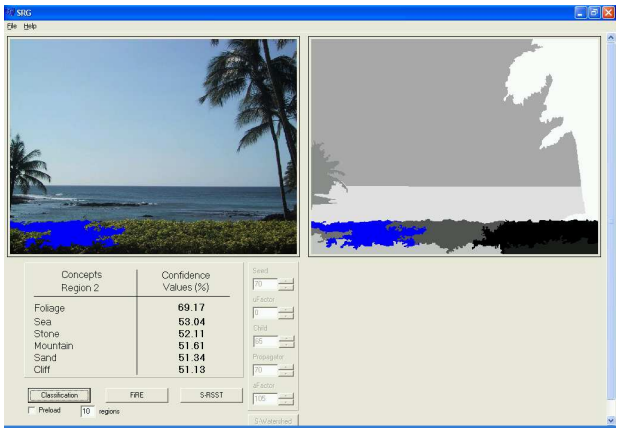

Figure 3. Knowledge-assisted analysis output. Concepts for the blue-highlighted region are displayed with their degree of confidence

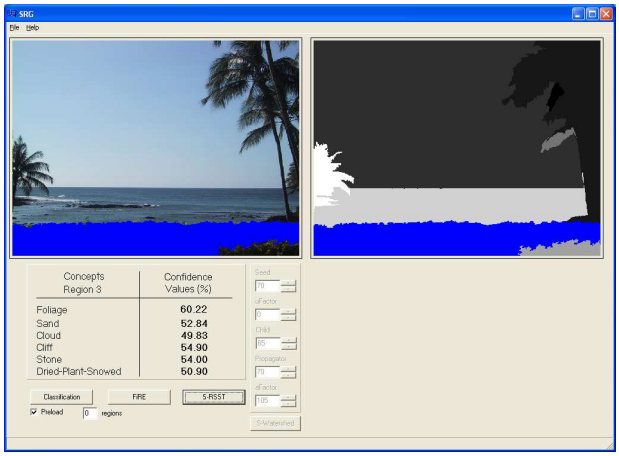

Figure 4. Semantic RSST output

mask and region-associated concepts. This mask is used by FiRE to extract implicit knowledge, producing in cooperation with KAA a new segmented image with higher-level concepts. The final results can be stored to a semantic repository and can be accessed by other modules. Though the results presented are restricted, they are representative of the architecture potential. Future work includes extension of this architecture using richer terminology and also the incorporation of text annotation. Furthermore this architecture will be examined using a large dataset (e.g TRECVID, Corel) and different domains.

References

[1] T. Adamek, N.O'Connor, and N.Murphy. Region-based segmentation of images using syntactic visual features. In *Proc. Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2005*, Montreux, Switzerland, April 13-15 2005.

[2] T. Athanasiadis, P. Mylonas, Y. Avrithis, and S. Kollias. Semantic image segmentation and object labeling. *IEEE Trans. on Circuits and Systems for Video Technology*, 17(3):298–312.

[3] S. Berretti, A. D. Bimbo, and E. Vicario. Efficient matching and indexing of graph models in content-based retrieval. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(12):1089–1105, Dec. 2001.

[4] I. K. G. Papadopoulos, V. Mezaris and M. Strintzis. Combining global and local information for knowledge-assisted image analysis and classification. *EURASIP Journal on Advances in Signal Processing, Special Issue on Knowledge-Assisted Media Analysis for Interactive Multimedia Applications*, accepted for publication.

[5] I. Horrocks, U. Sattler, and S. Tobies. Reasoning with Individuals for the Description Logic $\mathcal{SHIQ}$. In D. MacAllester, editor, *CADE-2000*, number 1831 in LNAI, pages 482–496. Springer-Verlag, 2000.

[6] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall, 1995.

[7] O. Morris, M. Lee, and A. Constantinides. Graph theory for image analysis: An approach based on the shortest spanning tree. *Inst. Elect. Eng.*, 133:146–152, April 1986.

[8] T. Sikora. The MPEG-7 Visual standard for content description - an overview. *IEEE Trans. on Circuits and Systems for Video Technology, special issue on MPEG-7*, 11(6):696–702, June 2001.

[9] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1349–1380, 2000.

[10] G. Stoilos, G. Stamou, V. Tzouvaras, J. Pan, and I. Horrocks. The fuzzy description logic f-shin. In *a*. International Workshop on Uncertainty Reasoning For the Semantic Web (2005), 2005.

[11] U. Straccia. Reasoning within fuzzy description logics. *Journal of Artificial Intelligence Research*, 14:137–166, 2001.