A Fuzzy Ontology – Based Framework for Reasoning in Visual Video Content Analysis and Indexing

Nizar ELLEUCH  
REGIM  
University of Sfax  
National School of Engineers (ENIS)  
BP 1173, Sfax, 3038, Tunisia  
nizar.elleuch@ieee.org

Mohamed ZARKA  
REGIM  
University of Sfax  
National School of Engineers (ENIS)  
BP 1173, Sfax, 3038, Tunisia  
medzarka@gmail.com

Anis Ben AMMAR  
REGIM  
University of Sfax  
National School of Engineers (ENIS)  
BP 1173, Sfax, 3038, Tunisia  
anis.benammar@ieee.org

Adel M. ALIMI  
REGIM  
University of Sfax  
National School of Engineers (ENIS)  
BP 1173, Sfax, 3038, Tunisia  
adel.alimi@ieee.org

ABSTRACT
Multimedia indexing systems based on semantic concept detectors are incomplete in the semantic sense. We can improve the effectiveness of these systems by using knowledge-based approaches which utilize semantic knowledge. In this paper, we propose a novel and efficient approach to enhance semantic concept detection in multimedia content, by exploiting contextual information about concepts from visual modality. First, a semantic knowledge is extracted via a contextual annotation framework. Second, a Fuzzy ontology is proposed to represent the fuzzy relationships (roles and rules) among every context and its semantic concepts. We use an abduction engine based on a meta function as a membership function for fuzzy rules. Third, a deduction engine is used to handle richer results in our video indexing system by running the proposed fuzzy ontology. Experiments on TRECVID 2010 benchmark have been performed to evaluate the performance of this approach. The obtained results show consistent improvement in semantic concepts detection, when a context space is used, and a good degree of indexing effectiveness as compared to existing approaches.

Categories and Subject Descriptors
D.3.3 [Information Search and Retrieval]: Content Analysis and Indexing

General Terms

Keywords
Concept Detectors, Contextual Information, Fuzzy, Ontology, Semantic Indexing.

1. INTRODUCTION
In recent year, digital video collections are growing rapidly, accompanied with technological advances in the field of multimedia and computer science. In organizing this data, a strong demand emerges for tools to expand access for efficient exploitation. Indexing audiovisual documents with high-level semantic concepts, such as objects, scene, events, locations, activities or specific objects, is the main key to enable it.

During the last decade matching semantic concepts and visual data has attracted the attention of lot of research in order to facilitate semantic indexing and concept-based retrieval of multimedia contents. Generally, semantic concepts detection is most often carried out by using supervised learning from manually annotated image samples. In fact, the majority of approaches are almost exclusively focused on the independent development of concept detectors. The latter focuses on the extraction of low-level visual features such as color, texture, and shape from positive and negative samples in order to models the high level concept. However, the same semantic concept may appear in various contexts and its appearance may be very different according to these contexts. Therefore, indexing audiovisual documents based on concept detector is not optimal. It requires a rather big number of training examples in order to produce a generic indexing system, on the one hand, and ignores the fact that concepts always coexist together, on the other hand.

For example, the concept sky frequently co-occurs with the concept Airplane flying. Thus, using the contextual information from Airplane flying is expected to help detect sky.

The contextual information is an ambiguous term. It has been defined and interpreted in several ways in different domains. In the multimedia literature, visual context was introduced in [14-16], as an extra source of information for both object detection and scene classification.

Accordingly the context may provide an important cue in order to enrich the semantic interpretation and further enhance the performance of semantic indexing and multimedia retrieval content systems [2-7]. Thus, our approach leans on this reasoning.

We model the contextual information in order to exploit and understand efficiently the high-level content. The context information modeling is composed of three steps: semantic
knowledge representation/interpretation, semantic concept/context categorization and refinement process. The semantic knowledge representation focuses on building the context space that represents the relationships (roles and rules) among every context and its semantic concepts. Such information is extracted, represented and stored via a proposed fuzzy abduction engine mated with an inference engine. More specifically, the inference engine provided by fuzzy description logics (IDLs) is used for context ontology construction that links each context space with its semantic concepts. The second step, semantic concepts/contexts categorization, focuses on the construction of a small, compact, vocabulary that effectively discriminates concepts and contexts. Based on this vocabulary, semantic concepts/contexts models are trained via a SVM classifier. The third step, refinement process, aims to enrich and enhance the semantic interpretation of our video indexing system. Based on fuzzy rules defined in our fuzzy ontology, a deduction engine is used to handle new richer results.

The remainder of this paper is organized as follows: In section 2, we review the relevant work on building and utilizing visual context space. Section 3 describes the proposed framework for detecting semantic concepts. In section 4, we detail the implementation of the proposed fuzzy ontology. In section 5, we represent the detail of the obtained experimental results via TRECVID 2010 benchmark. We conclude with the direction of future works.

2. RELATED WORKS

Semantic concept detection has captured extensive research attention in multimedia indexing and retrieval, attributed to its promising role in bridging the semantic gap. In fact, many techniques have been developed and several image/video indexing systems have been built [11, 12, 19, 20]. These systems have a common approach: understanding the semantic in a multimedia document is basely performed by building concept detectors in order to annotate automatically video shots with respect to a set of semantic concepts.

Building concept detectors is often conducted in a diverse setting where the emphasis usually involves two problems: codebook, or bag-of-visual-words, modeling and machine learning on huge multimedia data sets [1].

The codebook is one of the most useful techniques for the modeling of image contents. It can be described as follows. First, a set of stable points or regions are sampled from the input images. These stable points or regions which carry important information are repeatedly found under transformations, including scale, rotation, and affine transformations. Next, feature descriptors are constructed using local image information in the neighborhood of the feature points (or regions). The set of features collected from each image is then clustered into K visual words. Thus, an image can be represented as a K-dimensional histogram by counting the visual words occurrence in the image. Based on these feature vectors which represent a low level knowledge, the state-of-the-art systems use machine learning to bridge the semantic gap between the latter and semantic concepts.

Though the concept detectors approach for semantic video indexing provides satisfactory performance for some concepts most semantic concepts are still not easily detected. Thus, the typical concept detector approach alone is not efficient for multimedia processing, especially video indexing. In fact, the major drawback of this approach is the choice of machine learning and its parameters. Moreover, another drawback of this approach is that the concept detectors are often developed independently, ignoring the fact that concepts always coexist together and the training samples are naturally multi-labeled. Therefore, much new research has involved the exploration of semantic knowledge among concepts for video indexing. Particularly, they aim to develop a context-based concept fusion (CBCF) framework to enhance the concept detection results [2-7]. These approaches fall into two categories.

The first category is based on exploration of pairwise concept correlation [2, 3, 4] that is generally determined by observation (e.g., from manual annotations or machine tagging of training data set). In [4], the concurrent matrix generates an explicit model based on the concurrent relation to refine the annotations. In [2], the Domain Adaptive Semantic Diffusion (DASD) exploits the semantic context (concept relationship) to refine concept detection scores via a function level graph diffusion process. The semantic context is modeled in an undirected and weighted concept graph, which is then used to recover the consistency and smoothness of video indexing results.

Based on the statistical principles and the manual annotation used to approximate pairwise concept relations, these previous approaches suffer from two major drawbacks. First, such approximations may not be generally consistent when we have limited training data. Furthermore, it is difficult to obtain accurate statistics involving different generic concepts in general videos collection. Therefore, the relationship to other concepts is generally ignored. Second, the manual annotation methods used for labeling semantic concepts are most often incomplete. Thus, such missing information can lead to inaccurate approximations and to misleading statistics.

The second category is based on learning techniques [5, 6, 7]. In [7], the contextual relationship, this is modeled by SVM. Firstly, the Discriminative Model Fusion (DMF) method generates a model vector aggregating the detection score of all the individual detectors. After that, an SVM is learned in order to refine the detection of the original concepts. In [5], an active CBCF method is proposed for extending the DMF model so as to incorporate active labeling from a user. Firstly, users are solicited to annotate a small number of samples. After that, a context-based SVM classifier is learnt. In [6], a set of 75 related concepts are firstly discovered through measuring the mutual information between their detection scores and pseudo-labels. Then, an SVM is learned to re-order and refine the initial detection result. All these approaches employ a two-layer learning structure, which uses the individual detector scores of the first layer as an input feature vectors for training detectors in the second layer in order to refine detector scores by modeling the proximity relation among concepts. Although performance improvement is reported in [5, 7], there are major drawbacks.

First, these approaches are fully supervised and require explicit knowledge of the target semantic concept and ground-truth labels such as the ontology hierarchy [8] and the Bayesian networks (manually constructed in most cases) [9]. Second, the number of correlated concepts for a given concept is generally small, compared to the un-correlated ones; thus using all the concepts as in [7] will significantly affect the performance. Third, some detectors may provide inaccurate probability estimation, especially for the concepts with very few positive training
samples. Therefore, such detectors will have a significantly detrimental impact on the learning relation.

Considering these approaches and their main drawbacks, we propose, in the following, our framework for reasoning in visual content which overcomes the encountered limits.

3. PROPOSED FRAMEWORK

In this section, we present our model integrating a fuzzy ontology-based framework for reasoning in visual video indexing. The proposed approach involves three phases, namely semantic knowledge representation/interpretation, semantic concept/context categorization and refinement process.

Our contribution is elucidated within the a.m. phases, with more focus on modeling and building the context space and its exploitation to enhance video indexing system.

![Figure 1. Fuzzy abduction engine.](image)

Below is the description of semantic knowledge representation/interpretation based on context ontology construction that we perform.

3.1 Semantic Knowledge Representation / Interpretation

The semantic knowledge representation/interpretation based on a Fuzzy Abduction Engine aims to analyze, model and optimize context spaces with fuzzy roles and rules [21]. The latter help to discover, by using a deduction engine, further concepts and therefore enrich semantic interpretation. Thus, a context Ontology is constructed to first, model the relationships between each context and its semantic concepts, and then provide a deductive engine based on fuzzy rules.

3.1.1 Extracting the Contextual Space

The annotation process is generally used as a semantic knowledge extraction tool for assisting experts in identifying all entities (concepts) involved in specific domain knowledge (context space). In fact, the process of generating such semantic knowledge has in recent years been approached by organizing collaborative efforts. Thus, we propose an annotation framework to extract a semantic knowledge of each context space (see Figure 2). However, this system has to deal with three basic problems: how to unify the annotation, how annotate and how to get an objective annotation.

To reach these goals, our annotation system highlights the following services:

- Unified lexicon: In order to address the lack of convergence in user assigned free labels and therefore to unify the annotation, we adopt a fixed lexicon for annotation of concepts and contexts. In this field, the Large-Scale Concept Ontology for Multimedia (LSCOM) includes a unified set of concepts. Based on LSCOM lexicon definitions, our annotation framework of concepts (eg. Sky, Airplane, Road) and contexts (eg. Office, Airplane_Flying, Urban,) has been developed.
- Soft annotation: In order to capture all required information about the entities (semantic concepts) of each context space, a multi-labels annotation process is applied. Given the uncertainty of information about concept-context relationship, we integrate fuzzy rules. We attribute a membership relevancy degree to a semantic concept for a target context space. Therefore, as illustrated in figure2, we propose three relevance levels, named “Relevant”, “Not-Relevant” and “Not-Exist”. “Relevant” and “Not-Relevant” respectively indicate that the concept is present and semantically strong (weak) in the target context and “Not-Exist” their lack.
- Collaborative annotation: To get an objective annotation and disambiguate the complexity of semantic knowledge, we propose a collaborative annotation framework by multiple experts. Thus, we adopt the collaborative method proposed by A. Ksibi in [10].

Based on expert observations, the knowledge gained still remains uncertain. To overcome this, the appropriate solution is to incorporate fuzzy theory in ontology. Thus, integrating a fuzzy ontology to represent the uncertain contextual information allows the video indexing systems to handle the uncertainty of knowledge and enrich the semantic interpretations. These systems have to deal with two basic problems: how to build and represent the knowledge, and how to integrate context-concept information in video analysis to improve its effectiveness.

![Figure 2. A multi-label context annotation. The annotation process can be treated as an image in the contextual domain.](image)

In the following, the specification and design of the fuzzy ontology infrastructure are described.

1 http://www.ee.columbia.edu/~mvmm/lscom
We consider the “generalization” role between two concepts. For instance, “Snow” and “Mountain” are related in the “IsRelatedTo” relationship, which is denoted as: c1 \(\rightarrow\) c2.

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
<th>Meaning</th>
<th>TxT</th>
<th>CxC</th>
<th>CxT</th>
<th>TxC</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalization</td>
<td>(t_i : t_j)</td>
<td>The concept (c_i) is the generalization of the concept (c_j)</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>LSCOM</td>
</tr>
<tr>
<td>IsRelatedTo</td>
<td>(c_i \rightarrow c_j)</td>
<td>The concept (c_i) is related to the concept (c_j) within (t_k)</td>
<td>(X)</td>
<td></td>
<td></td>
<td></td>
<td>Learning</td>
</tr>
<tr>
<td>IsPartOf</td>
<td>({c_i} \in t_j)</td>
<td>A set of concept (c_i) is a part of the context (t_j)</td>
<td>(X)</td>
<td></td>
<td></td>
<td></td>
<td>Learning</td>
</tr>
<tr>
<td>Includes</td>
<td>(t_i \supset c_j)</td>
<td>The context (t_i) includes the concept (c_j)</td>
<td>(X)</td>
<td></td>
<td></td>
<td></td>
<td>Expert</td>
</tr>
</tbody>
</table>

### 3.1.2 Ontology Structure

For specifying the proposed ontology to describe semantic knowledge, less expressive fuzzy description logic is applied to facilitate fast computations. In the following, we detail how we constructed a contextual knowledge model for semantic interpretations. In this field, our fuzzy ontology \(O_f\) is modeled as:

\[
O_f = \{T, C, R_{tc}, R_{ct}, R_{ctc}, Q\}
\]

- \(T = \{t_1, t_2, \ldots, t_n\}\) is a set of \(n\) contexts
- \(C = \{c_1, c_2, \ldots, c_m\}\) is a set of \(m\) concepts
- \(R_{tc}: T \times C \rightarrow [0, 1]; i \in [0, \ldots, n]\) and \(j \in [0, \ldots, m]\); is a fuzzy rule that the context \(t_i\) performs for the concept \(c_j\)
- \(R_{ct}: C \times T \rightarrow [0, 1]; j \in [0, \ldots, m]\) and \(i \in [0, \ldots, n]\); is a fuzzy rule that the concept \(c_j\) performs for the context \(t_i\)
- \(R_{ctc}: C \times C \times T \rightarrow [0, 1]; j \in [0, \ldots, m]\) and \(k \in [0, \ldots, n]\); is a fuzzy rule that the concept \(c_j\) performs for the context \(t_k\)
- \(Q\) is a set of fuzzy qualifier. In \(O_f\), we define two qualifiers: “weak” and “strong”.

We have also defined some roles between concepts and contexts {Generalization, IsRelatedTo, IsPartOf, Includes}. Their interpretation is rather simple and detailed in Table I.

- **Definition 1 (Generalization Role)**
  We consider the “generalization” role between \(t_i\) and \(t_j\) if \(t_i\) is a sub-context of \(t_j\), which is denoted as: \(t_i : t_j\).
  For instance, “Ground_Vehicle” and “Vehicle” are related in the “Generalization” relationship. Ground_Vehicle : Vehicle indicates that all relevant video shots for sub-context “Ground_Vehicle” must also be relevant to context “Vehicle”. Actually, the generalization relationship is the most common relation used to build ontology hierarchy, which can be exploited to enhance concept detectors. Moreover, the LSCOM ontology, dealing only with this relationship, provides a ready enumeration of generalizations between all defined concepts.

- **Definition 2 (IsRelatedTo Role)**
  We consider the “IsRelatedTo” role between \(c_i\) and \(c_j\) if \(c_i\) is related to \(c_j\) within \(t_k\), which is denoted as: \(c_i \rightarrow c_j\).
  For instance, “Snow” and “Mountain” are related in the “IsRelatedTo” relationship, which suggests that all relevant video shots to concept “Snow” could be relevant to concept “Mountain”.

- **Definition 3 (IsPartOf Role)**
  We consider the “IsPartOf” role between \(c_i\) and \(t_j\) if \(c_i\) is part of \(t_j\), which is denoted as: \(c_i \in t_j\).

For instance, {“Sky”, “AirPlane”} and “AirPlane_Flying” are related in the “IsPartOf” relationship. Sky, AirPlane \(\in\) AirPlane_Flying lead to all relevant video shots to concept “Sky” and “Airplane” could be relevant to context “AirPlane_Flying”.

- **Definition 4 (Includes Role)**
  We consider the “Includes” role between \(t_i\) and \(c_j\) if \(t_i\) includes \(c_j\), which is denoted as: \(t_i \supset c_j\).
  For instance, “CarRacing” and “Car” are related in the “Includes” relationship. CarRacing \(\supset\) Car suggests that all relevant video shots to context “CarRacing” could be relevant to concept “Car”.

In order to make the above roles able to be applied in real-world situations, we introduce for each one a degree of confidence \(\alpha\) \in [0,1]. In addition, for each role, a \(\mu\) function is defined that aims to computes respectively for each related pairwise \(<c_i, c_j>\), <\(c_i, t_j>\) and <\(t_i, c_j>\) a degree that \(c_i\) supplied for \(c_j\), \(c_i\) supplied for \(t_j\) and \(t_j\) supplied for \(c_j\). \(\alpha\) and \(\mu\) are generated automatically through Abduction Engine based on \(\beta\) function. Generally, the \(\beta\) function is defined as follows:

\[
\beta(x) = \begin{cases} 
\frac{x - x_0}{x - x_0} & \text{if } x \in [x_0, x_1] \\
0 & \text{otherwise}
\end{cases}
\]

Where:
- \(p > 0, q > 0\);
- \(x_0\) and \(x_1\) are real parameters;
- \(x_\alpha = \frac{(px_1 + qx_0)}{(p+q)}\)

According to relevance degrees proposed in our context annotation framework, our fuzzy ontology \(O_f\) employs two linguistic labels, or qualifiers \((Q = \{"Weak", "Strong"\})\), in order to provide a fine- tuning of degrees of confidence. Thus, as shown in Fig 3, each rule is “Strong” qualified if its degree of confidence is greater than 0.5 and “Weak” qualified if its degree of confidence is less than 0.5.

![Figure 3. Two Fuzzy \(\beta\) function to represent \(O_f\) qualifiers.](image-url)
3.1.3 Building Ontology through Abduction Engine

In order to detect and extract further rules within concepts and contexts, we use the Multi-Agent Genetic Algorithm for the Design of Beta Fuzzy Systems (MAGAD-BFS), proposed in [13], as an Abduction Engine.

Based on genetic algorithm (GA), MAGAD-BFS allows us to optimize a fuzzy logic system (FLS) with Beta membership functions [17]. It consists of minimizing the number of Beta fuzzy rules \( N_{\text{fl}} \), which are formulated according to the Eq 2, while adjusting Beta function \( p \) and \( q \) parameter’s (\( \text{obj}_{\text{fun}} \)) of each rule until a desired precision \( \varepsilon \).

\[
R_j : \{ R^{f}_{\text{ct}}, R^{f}_{\text{cel}}, R^{f}_{\text{ct}} \}; \quad \text{IF} \ (X \text{ is } Q^j) \text{ THEN } (Y = f_j(X)) \quad (2)
\]

Where:
- \( X = \cup T \) is a input variable;
- \( Y = \cup T \) is an output variable;
- \( Q^j \) is a linguistic qualifier of input variable;
- \( f_j \) is the output of the \( j \)th fuzzy rule.

The objective function (\( \text{obj}_{\text{fun}} \)) to be minimized is defined as follows:

\[
f^*(X) = \sum_{j=1}^{N_{\text{fl}}} f_j(X) \beta_j(X) \quad (3)
\]

\[
\text{obj}_{\text{fun}} = \frac{\sum_{t=1}^{m+n} \left[ Y_t - f^*(X_t) \right]^2}{\sum_{t=1}^{m+n} Y_t^2} \times \left( \frac{N_{\text{fl}} - N_{\text{min}}}{N_{\text{max}} - N_{\text{min}} + 1} \right) \quad (4)
\]

Where:
- \( N_{\text{min}} \) and \( N_{\text{max}} \) are respectively the minimum and the maximum number of fuzzy rules allowed in the final Beta fuzzy system;
- \( \mu_j = \beta_j \) is a Beta function that activates the \( j \)th fuzzy rule.

Accordingly, each video shot \( V_{sk} \) is ranked with a probabilistic measure \( P(V_{sk}|c_i) \) or \( P(V_{sk}|t_j) \).

3.3 Deduction Engine

Based on probabilistic scores of concepts and contexts provided by the indexing process, a fuzzification step is performed. The latter aims to handle the imprecision and inexactness of concepts and contexts detectors, on one hand, and generate the fuzzy inputs required by fuzzy rules on the other hand. Thus, we consider a concept \( c_1 \) or a context \( t_j \) “Relevant” in a video shot \( V_{sk} \) if \( P(V_{sk}|c_i) \) respectively \( P(V_{sk}|t_j) \) is greater than 0.7. However a concept \( c_1 \) or a context \( t_j \) is qualified by “Not-Relevant” in a video shot \( V_{sk} \) if \( P(V_{sk}|c_i) \) respectively \( P(V_{sk}|t_j) \) is between 0.3 and 0.7.

Based on these fuzzy inputs, the deduction engine explores all defined rules in order to infer the most appropriate one and thus generates an optimal score for the target rule output. In this fields, two cases arise: when a fuzzy rule is “Strong” qualified or “Weak”.

In the first case, the deduction engine proceeds as follow: Let \( R^f_k \) a fuzzy rule defined as: \( R^{f}_{k}; \ c_i \text{ is Strong RelatedTo } c_j \) within \( t_k' \), and let \( P(V_{sk}|c_i) \) and \( P(V_{sk}|t_k') \), respectively, a score detection of concept \( c_i \) and context \( t_k' \) in the same video shot \( V_{sk} \). The optimal score, or the deduced relevance degree, of the fuzzy rule \( R^f_k \) outputs, denoted as \( a^f_k(c_i) \), is computed as follow:

\[
\alpha^f_k(c_i) = \mu_\delta \left( \max \left( P(V_{sk}|c_i), P(V_{sk}|t_k') \right) \right) \times \mu_{\text{Strong}}(\alpha_k) \quad (5)
\]

Where \( \mu_\delta \) and \( \alpha_k \) are, respectively, the Beta membership function and the confidence degree of the \( k \)th fuzzy rule according the rule “IsRelatedTo”.

In the second case, the deduction engine applies the following equation.

\[
\alpha^f_k(c_i) = \mu_\delta \left( \min \left( P(V_{sk}|c_i), P(V_{sk}|t_k') \right) \right) \times \mu_{\text{Weak}}(\alpha_k) \quad (6)
\]

The same approach is built by the deduction engine for the other rules according the rule “IsPartOf”, “Includes” and “Generalisation”.

4. IMPLEMENTATION OF THE PROPOSED FUZZY ONTOLOGY

As aforementioned, fuzzy contextual ontology is a formal and explicit representation of semantic knowledge in visual domain. In this section, we will detail its implementation process.

4.1 Knowledge Extraction

When building the context space, we need a large scale corpus for generalizing the contextual information and the relationship between contexts-concepts. To attempt this challenge, various knowledge sources are developed in multimedia literature such as Flickr images, LSCOM, WordNet, TRECVID data sets and etc. Accordingly, our contextual annotation framework explores actually the development data set IACC.1.tv10.training\(^2\) which is composed of 119685 shots. Each shot is manually assigned to a predefined context specifying the meaning of its contents.

---

recognized by experts. In addition, it is given a set of concepts with three relevance degrees as presented in section 3.

Once expert’s collaborative observations have been completed, as shown in Fig 5, a statistical analysis is performed. The objective is to learn the correlation $m_{ij}$ between concepts $c_i$ and $c_j$ in their context space. $m_{ij}$ is given by:

$$m_{ij}(c_i, c_j) = \frac{\sum_{y \in Y} P(y_i, y_j) \log \frac{P(y_i, y_j)}{P(y_i)P(y_j)}}{\log \frac{\sum_{y \in Y} P(y_i, y_j)}{\sum_{y \in Y} P(y_i)P(y_j)}}$$

Where $m_{ij}(c_i, c_j)$ is the mutual information between concepts $c_i$ and $c_j$. $y_i \in \{+1, +0.5\}$ and $y_j \in \{+1, +0.5\}$ respectively, indicate the relevance degrees (“Relevant” and “Not-Relevant”) of $c_i$ and $c_j$ for $t_k$ shots. The probabilities $P(y_i)$, $P(y_j)$ and $P(y_i, y_j)$ can be estimated from IACC.1tv1010 training data set. Then, we normalize the mutual information by adopting the marginal entropy minimization strategy [18] which is defined by the Eq 8. The latter scaled the mutual information into the interval $[0, 1]$.

$$\text{Norm}_{m_{ij}} = \frac{m_{ij}(c_i, c_j)}{\text{min} \{H(c_i), H(c_j)\}}$$

Where $H(c_i)$ is the marginal entropy of the concept $c_i$. It is defined as follow:

$$H(c_i) = -\sum_{y \in \{+1, +0.5\}} P(y_i) \log P(y_i)$$

4.2 Fuzzy Rules Abduction

In order to discover the fuzzy rules relating to the roles “Includes”, “IsPartOf” and “IsRelatedTo”, the abduction engine is trained based on the semantic knowledge. Thus, for every output of the above roles, feature vectors are firstly generated. A feature vector is a string of numerical values whose dimension is $n + m$ that correspond to the number of concepts and contexts. A 1 or 0.5 or 0, at $i^{th}$ position, indicates, respectively, whether the $i^{th}$ concept or context is “Relevant” (1), “Not-Relevant” (0.5) or “Not-Exist” (0) for the expected output. Then, the abduction engine is consecutively learned and provides fuzzy rules by estimating the degree of confidence $\alpha$ and the beta membership function $\mu$, as shown in Table II.

The ontology filling process consists in applying two sub-processes which the complexity of each one is $o(n^t)$.

5. EXPERIMENTS AND RESULTS

In the following, we present the experimental results. These are conducted on TRECVID 2010 datasets which are widely employed for the evaluation of semantic systems indexing. The main goal is to test the improvement of using context space for semantics concepts detection and check the effectiveness of the proposed approach as compared to existing techniques. Hence, we first examine the effectiveness of using a semantic knowledge induced by the proposed fuzzy contextual ontology to enhance the detection of semantic concepts and to better enrich the semantic interpretation of multimedia content. We use the inferred average precision (inAP), the precision (P) and the recall (R) as the performance metric.

We ran experiments on the subsets defined in Fig 5. The obtained results are reported in Table III.

Looking at the results, we make the following comments:

- Table II lists a part of the fuzzy rules generated by the Abduction Engine. We can see that these fuzzy rules are able to infer more reasonable predictor concepts in several cases, e.g., concepts {Snow, Mountain} within Landscape context strongly leads to the detection of the concept Sky. These fuzzy rules represent the ground truth of domain knowledge.

- The effectiveness of video indexing systems is clearly improved when a knowledge-based approach is integrated. In fact, when the LSCOM ontology is incorporated, the precision improvement of semantic concept detection is the order of 11%. However, we have obtained 21% via our ontology $O^I$. This variation is mainly due to the hierarchical roles of each one. The LSCOM ontology, based on “Generalization” roles, provides enrichment only for the concepts of a higher level. However, the $O^I$ ontology expounds other roles such as “IsPartOf”, “Includes” and “IsRelatedTo”. These allow us to highlight the relation between a context and its concepts and concept-concept within a target context space.

- The proposed approach improves not only the precision of contexts detection, but also concepts detection. In fact, our ontology $O^I$ performs best for 16 (6 context and 10 concepts) for 17 high level feature. This result is rather obvious: the proposed ontology $O^I$ tries to represent the concept space; with 4 roles (“Generalization”, “IsPartOf”, “Includes” and “IsRelatedTo”); by using an Abduction Engine. The latter automatically generates fuzzy rules and optimizes them. These fuzzy rules, that represent the ground truth, further improve the effectiveness of video shot results, which will improve the Inferred Average Precision.

- The context-based concept fusion framework enhances the high level feature detection. In fact, the recall is improved for 5 (Outdoor, Vegetation, Vehicle, Ground_Vehicle, Airplane_Flying) out of 17 high level feature. We can see that the enrichment has only targeted the context. Although this recall improvement
Table 2. A Partial view of the abducted Fuzzy rules

<table>
<thead>
<tr>
<th>Name</th>
<th>The abducted fuzzy rule</th>
<th>Qualifier</th>
<th>β membership function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalization</td>
<td>Airplane_flying : Vehicle</td>
<td>Strong</td>
<td>-</td>
</tr>
<tr>
<td>IsRelatedTo</td>
<td>Airplane</td>
<td>Airplane_flying → sky</td>
<td>Strong</td>
</tr>
<tr>
<td></td>
<td>Snow</td>
<td>Landscape → Mountain</td>
<td>weak</td>
</tr>
<tr>
<td></td>
<td>Sno</td>
<td>Snow, Mountain</td>
<td>Landscape → Sky</td>
</tr>
<tr>
<td></td>
<td>Person</td>
<td>Studio, News → Anchorperson</td>
<td>Strong</td>
</tr>
<tr>
<td>IsPartOf</td>
<td>{Sky, Mountain} ∈ Landscape</td>
<td>Strong</td>
<td>p = 8, q = 0.01</td>
</tr>
<tr>
<td></td>
<td>{Building, Sky, Road, Car} ∈ Urban</td>
<td>Strong</td>
<td>p = 9, q = 0.02</td>
</tr>
<tr>
<td>Includes</td>
<td>Landscape ⊇ Snow</td>
<td>weak</td>
<td>p = 2, q = 2</td>
</tr>
</tbody>
</table>

Table 3. Concept retrieval performance (Inferred Average Precision infAP, Precision P and Recall R) for different Concept detection methodologies applied on TRECVID 2010 data set.

<table>
<thead>
<tr>
<th>High level feature</th>
<th>Concept Detector REGIMVID with LSCOM</th>
<th>REGIMVID with O²</th>
<th>CBCF with SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>InfAP P R</td>
<td>infAP P R</td>
<td>P R</td>
</tr>
<tr>
<td>Outdoor</td>
<td>- 0.52 0.59</td>
<td>- 0.88 0.77</td>
<td>0.9 0.82</td>
</tr>
<tr>
<td>Vegetation</td>
<td>- 0.74 0.68</td>
<td>- 0.74 0.68</td>
<td>0.93 0.87</td>
</tr>
<tr>
<td>Landscape</td>
<td>- 0.6 0.79</td>
<td>- 0.6 0.79</td>
<td>0.7 0.82</td>
</tr>
<tr>
<td>Sky</td>
<td>- 0.6 0.9</td>
<td>- 0.6 0.9</td>
<td>0.85 0.95</td>
</tr>
<tr>
<td>Trees</td>
<td>- 0.6 0.72</td>
<td>- 0.6 0.72</td>
<td>0.73 0.82</td>
</tr>
<tr>
<td>Mountain</td>
<td>- 0.6 0.8</td>
<td>- 0.6 0.8</td>
<td>0.83 0.85</td>
</tr>
<tr>
<td>Vehicle</td>
<td>0.016 0.22</td>
<td>0.103 0.77</td>
<td>0.72 0.76</td>
</tr>
<tr>
<td>Ground_Vehicle</td>
<td>0.043 0.66</td>
<td>0.18 0.73</td>
<td>0.69 0.75</td>
</tr>
<tr>
<td>Road</td>
<td>- 0.43 0.6</td>
<td>- 0.43 0.6</td>
<td>0.88 0.9</td>
</tr>
<tr>
<td>Car</td>
<td>0.075 0.64</td>
<td>0.17 0.73</td>
<td>0.79 0.83</td>
</tr>
<tr>
<td>Bus</td>
<td>- 0.5 0.73</td>
<td>- 0.5 0.73</td>
<td>0.52 0.73</td>
</tr>
<tr>
<td>Bicycles</td>
<td>0.142 0.92</td>
<td>0.185 0.97</td>
<td>0.83 0.97</td>
</tr>
<tr>
<td>Emergency Vehicle</td>
<td>- 0.9 0.83</td>
<td>- 0.9 0.83</td>
<td>0.9 0.83</td>
</tr>
<tr>
<td>Building</td>
<td>0.022 0.22</td>
<td>0.1 0.43</td>
<td>0.55 0.45</td>
</tr>
<tr>
<td>Truck</td>
<td>- 0.35 0.37</td>
<td>- 0.35 0.37</td>
<td>0.35 0.37</td>
</tr>
<tr>
<td>Airplane Flying</td>
<td>0.102 0.78</td>
<td>0.102 0.78</td>
<td>0.83 0.79</td>
</tr>
<tr>
<td>Airplane</td>
<td>- 0.5 0.6</td>
<td>- 0.5 0.6</td>
<td>0.71 0.69</td>
</tr>
<tr>
<td>Total:</td>
<td>0.071 0.66</td>
<td>0.134 0.71</td>
<td>0.74 0.77</td>
</tr>
</tbody>
</table>

(about 2%), the precision improvement has declined. The origin of this degradation is its requirement for an explicit knowledge about contextual space that is provided manually based on intuition and human knowledge. In addition when a detector provides inaccurate probability (e.g. Sky, truck, bus, road), the effectiveness is low (the precision improvement of Vehicle is 10%). In these fields, O² reached a rate of about 50%.

6. CONCLUSION AND FUTURE WORKS

Our current research efforts indicate clearly that high level concepts can be efficiently detected when a knowledge-based approach is incorporated within a video indexing system. Thus the core contribution of this work has been the implementation of a fuzzy contextual ontology. The latter aims to represent semantic’s knowledge of concepts, which are extracted by a contextual annotation framework. The representation of such knowledge was performed by an abduction engine via fuzzy rules. Then, we applied our contextual ontology to enhance and further refine concept detection over TRECVID 2010 corpus. Its effectiveness in terms of precision and recall is proved on diverse concepts.

Future work will consist of extending others knowledge sources such as Flickr images for our fuzzy contextual ontology. In addition, it would be very interesting to integrate the spatial coherence between the context space with the exploration of cross-concept correlation and inter-shot dependency.
7. ACKNOWLEDGMENTS
The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program. Also, the authors are grateful to NIST and the TRECVID coordinators for the benchmark organization’s effort.

8. REFERENCES