Republic of TUNISIA
Ministry of Higher Education and
Scientific Research

University of Sfax

National Engineering School of Sfax

Graduate School Sciences and
Technologies
PHD Thesis
Computer System Engineering

Serial N°: .....

THESIS

Presented in

The National Engineering School of Sfax

In order to obtain the

DOCTORATE

in

Computer System Engineering

by

Mohamed ZARKA

(Master degree in Software Engineering)

Fuzzy Reasoning for Multimedia Semantic Interpretation: Fuzzy Ontology Based Model for Video Indexing

---

defended on November 16^{th}*, 2016*

Jury

Prof.	Chokri Ben Amar,	University of Sfax, Tunisia	*President*
Prof.	Slim Kanoun,	University of Sfax, Tunisia	*Examinator*
Prof.	Fakhreddine Karray,	University of Waterloo, Canada	*Reviewer*
Prof.	Sami Faiz,	University of Manouba, Tunisia	*Reviewer*
Prof.	Adel M. Alimi,	University of Sfax, Tunisia	*Advisor*
Dr.	Anis Ben Ammar,	University of Sfax, Tunisia	*Co-advisor*

Abstract

Our thesis work deals with the video indexing based on semantic interpretation (an abstraction of objects or events that figure in a content), more particularly, the semantic indexing enhancement. Various approaches for semantic multimedia content analysis have been proposed addressing the discovery of features ranging from low-level features (color, histograms, sound frequency, motions, ...) to high-level ones (semantic objects and concepts). However, these earlier approaches failed to reduce the semantic gap and were not able to deliver an accurate semantic interpretation. Under such a context, exploring further semantics within a multimedia content to improve semantic interpretation capabilities, is a major and a prerequisite challenge.

Towards exploring further semantic information within a multimedia content (other than low-level and semantic concepts one), valuable information (mainly concepts interrelationships and contexts) could be gathered from a multimedia content in order to enhance semantic interpretation capabilities. Motivated by a kindred vision of human perception, yet targeting automated analysis of a multimedia content, the multimedia retrieval community addressed more attention to multimedia ontologies.

Aiming to contribute towards this direction, we focus on modeling an automated fuzzy context-based ontology framework for enhancing a video indexing accuracy and efficiency. Key dimensions of this inquiry constitute the main issues addressed by the use of ontologies for multimedia indexing, namely: (1) the knowledge management and evolution, (2) the ability to handle uncertain knowledge and to deal with fuzzy semantics, and (3) the scalability and the ability to process a growing multimedia content volume with a continuous request for a better machine semantic interpretation capacities.

What was accomplished in our study is a novel ontology management which is intended to a machine-driven knowledge database construction. Such a method could enable semantic improvements in large-scale multimedia content analysis and indexing.

In order to illustrate the semantic enhancement of concept detection introduced by our proposed scalable and generic ontology-based framework, we have conducted different experiments within three multimedia evaluation campaigns: TrecVid 2010 (within *Semantic Indexing* Task), *ImageClef 2012* (within *Photo Annotation and Retrieval* Task), and *ImageClef 2015* (within *Scalable Concept Image Annotation* Task).

Keywords: Video Indexing, Semantic Interpretation, Fuzzy Ontology, Fuzzy Reasoning, Hierarchical Concept Detector.

Dedicates

To my parents Salah and Monia,

To my wife Bochra and my daughters Mariem and Sarra,

To my sisters Rim and Mariem,

who have contributed to my work like nobody else.

Acknowledgements

I was once told that PhD is a long journey of transformation from a *novice* to *professional* researcher. To succeed, an individual can never do it alone; there must be someone who is there for this, providing all the helping hands and supports. I can't agree more, and as for my case, I have a lot of people to thank.

First thanks must be to my supervisor, *Prof. Adel M. Alimi*, who has been truly inspirational throughout my candidature. He always shows me a tireless enthusiasm to meet, discuss, listen and encourage. His invaluable pieces of advice and faith make all the difference, and I have been blessed to find in him all the good qualities of a supervisor.

Many thanks must go to my co-supervisor, Dr. *Anis Ben Ammar*, who has given all the extra guidance and help that prove to be precious in improving the quality of my work.

I would like to hold this opportunity to express my gratitude and respect to the reviewers *Prof. Sami Faiz* and *Prof. Fakhreddine Karray*, to the exterminator *Prof. Slim Kanoun* and to the chairman *Prof. Chokri Ben Amar*, for their acceptance to evaluate this research work and their valuable remarks and feedback.

A special thanks to Dr. *Nizar Elleuch* and Dr. *Issam Feki* who were always concerned to provide me the necessary conditions for the completion of my thesis.

The Regim-Lab has provided me nice working places with all the much needed facilities, and services, as well as financial supports including travel allowances. Thanks to all the people there who have helped me with many things.

I would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

I thank my wife *Bochra* who is always there for me and makes this whole journey so enjoyable.

Thanks to all my colleagues *Kais* and *Ghada* who have always given me tremendous supports and encouragement.

Last, but most importantly, I'd like to dedicate this thesis for my father *Salah* and my mother *Monia* to express my deepest gratitude. They are the best parents who are so willing in giving me the best in life without hoping for anything in return. I am also so blessed to have such really loving sisters *Mariem* and *Rim* who gave me so much support and encouragement to do well during this candidature.

Above all, I thank my stepfather *Mohamed Hechmi* and my stepmother *Fatma* who have always given me tremendous supports and encouragement. I would like also to thank my stepbrothers *Kais*, *Hamza*, *Mohamed Hédi*, *Mondher*, *Mohamed Salah* and *Abdelkarim* for supporting me spiritually throughout writing this thesis and in my life in general.

Thanks to all

Mohamed ZARKA

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Background and Motivation

Due to the advance of modern multimedia services and techniques, and the exponential growth of online multimedia data, images and video are overflowing our everyday life. FLICKR, as a photo sharing website, stores over 5 billion of images and records an uploading rate of over 3000 images per minute[1]. YOUTUBE records a rate of 300 hours of uploaded videos per minute[2]. Due to this huge amount of available information, and the diversity of their support and content (newscasts, sports, entertainment programs, movies, documentaries, video surveillance, ...), some problems and difficulties are raising, and a need for efficient tools to access to such amount of multimedia content was strongly stated.

The common information retrieval model consists in handling useful information through searching and retrieving from large document collections [Baeza-Yates et al. 1999, Manning et al. 2008, Grossman & Frieder 2012]. Basically, this model is based on two processes: (1) the indexing process which interprets and stores the content of documents, and (2) the search process which matches a user query and the stored document interpretations in order to evaluate their relevance and output relevant documents.

Image and video indexing is basically based on two different approaches: *Text Based* indexing [Chang & Fu 1980] and *Content Based* indexing [Smeaton et al. 2008]. While the *Text Based* means that a content is manually annotated by a text describing contained

[1] `http://blog.flickr.net/en/2010/09/19/5000000000`
[2] `https://www.youtube.com/yt/press/fr/statistics.html`

objects and events, the *Content Based* indexing means the extraction of features (like color and shape) in order to automatically detect contained semantic objects.

Authors in [Hauptmann, Yan, Lin, Christel, & Wactlar 2007, C. Wang et al. 2008, Hauptmann, Yan, & Lin 2007, Kompatsiaris & Hobson 2008, Darwish & Ali 2015, de Ves et al. 2015] pointed out that although the availability of many solutions to fill the so-called *semantic gap* [Hare et al. 2006] between the automatic interpretation and what human expects from the indexing task; none of them yields satisfactory results: the problem remains open for further research. In [Smeulders et al. 2000], the *semantic gap* is defined as:

Definition 1. *"... the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation."*

As illustrated in figure 1.1, the semantic gap manifests itself as a computational problem for the indexing process. At the low-level, the raw media refers to an image.The content-based indexing approaches extract features vectors that represent regions of the image or its whole content. At a higher level, the *objects* are detected through combinations of extracted features vectors. *Symbolic labels* may be assigned to these objects (*Crowd* and *Road* for the given example). Nevertheless, such a classical process (based on extracting and combining feature vectors) does not typically capture all the semantics in a given content. In fact, the image context and the relationships between the contained objects contribute to reach a high level of semantic representation and interpretation of the media content. Thus, more than the content analysis is requested for a better semantic interpretation: The knowledge structure components showed its capabilities to reach such a higher semantic interpretation. Moreover, the information retrieval is more and more supported by knowledge management processes.

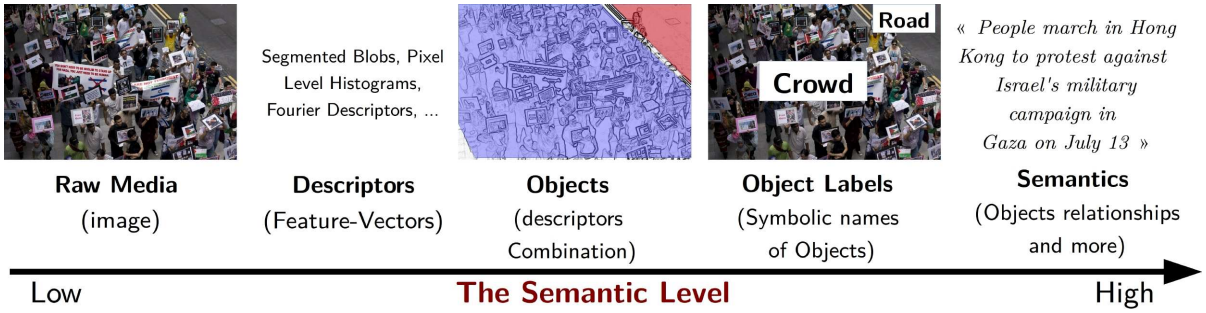


Figure 1.1: The Semantic Gap: Hierarchy of levels between the raw media and full semantics

In the current era of multimedia indexing, there is an emerging attention in leveraging massive exploration of knowledge management capacities in order to reduce the semantic gap. Indeed, the use of background knowledge could enrich and enhance a semantic interpretation about a content. As an example, the figure 1.2 displays four images where a classical automatic indexing process could detect the semantic object *"crowd"* and the semantic object *"road"* (for the images B, C and D). Such a basic object's recognition did not reflect the complete semantics of each one.

On the other side, the human perception uses a pre-established knowledge in order to better deduce newer information about that content. For the image A, one can deduce that the *crowd* attends a *Sport Game* (like football). The image B depicts also a *crowd* ahead a *road*, one can deduce also that this *crowd* of people attends a *car racing*. The image C depicts also a *crowd* ahead a *road*, but one can deduce that they participate to a protest since people take many slogans. Finally, the image D figures a *crowd* and a *road*, but we can not deduce nor a *protest* neither a *car racing*; just a *crowd* of people crossing a *road*.

The aforementioned examples incur a rather weak setting regarding the efficiency of the classical indexing process, establishing the need for a knowledge-based models and approaches. Shedding further light, such a knowledge-based approaches may allow not only for better semantic interpretation capabilities but also for the identification of open challenges and future directions and opportunities in multimedia retrieval.

Through answering many of aforementioned limitations, our motivation goes further toward exploring knowledge-based approaches for enhancing the indexing process capabilities.

1.2 Research Issues

Despite advanced IR capabilities, user still unsatisfied by their delivered search results. This discontent is caused commonly by non-relevant results that can be due to a non-good understanding of stored documents, and/or the non-apprehension of the user query. In fact, popular and widely-used search engines, like Google, Bing and Yahoo !, essentially rely on classical keyword matching techniques, and their relevance is often unsatisfying owing to inaccurate textural tag information.

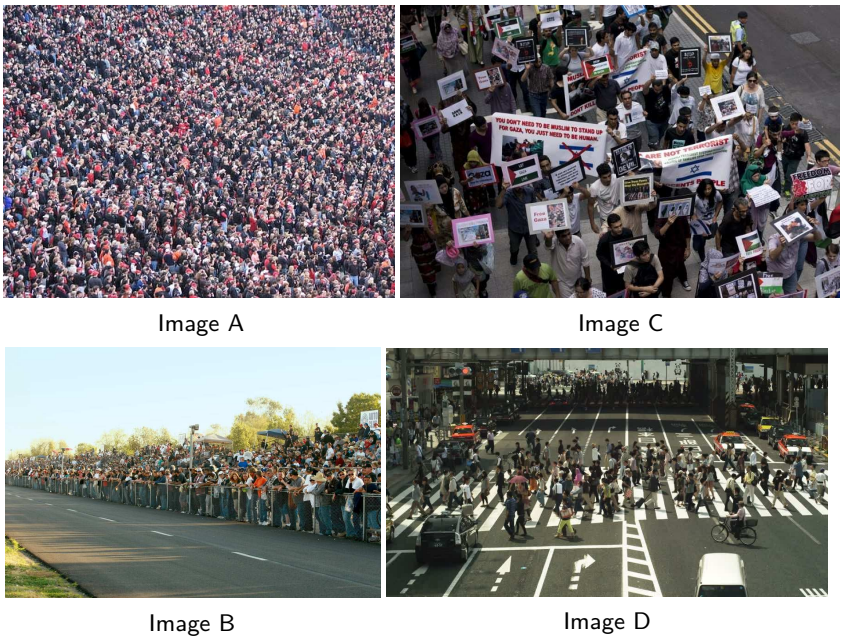


Figure 1.2: Different semantic interpretation for images figuring the semantic concept *crowd*

In such a situation, the multimedia retrieval community has looked for new approaches in order to make the availability of automated and efficient semantic interpretations for multimedia contents. Thus, a number of diverse approaches for semantic multimedia content analysis have been proposed addressing the discovery of multi-modal features [Marques & Furht 2012] like visual features (patterns, color, histograms, ...) and audio features. In particular, Content based image retrieval CBIR has attracted a lot of attention over the last two decades [Rui et al. 1999, Smeulders et al. 2000, Datta et al. 2005, Y. Liu et al. 2007, Huurnink et al. 2012]. However, these earlier approaches failed to reduce the semantic gap between the extracted features and the user's perception. Then, the next approaches focused on exploring these low-level features in order to detect high level-feature (semantic objects and concepts) [Datta et al. 2008, C. G. Snoek & Worring 2008, Egozi et al. 2011]. These semantic analysis based approaches proved to be highly effective and useful in many application areas, but, once again, failed to deliver an efficient semantic interpretation [C. G. Snoek & Smeulders 2010, Over et al. 2013]. Under such a context, exploring further semantics within a multimedia content is a major challenge.

To sum up, in our dissertation, we considered that a knowledge based approach should be more explored and tackled. We addressed the following research issues:

Issue 1: A novel knowledge based approach for video/image indexing other than low-level and high-level ones, further semantic information is gathered from a multimedia content in order to better interpret a semantic content. Motivated by a kindred vision of human perception, new approaches have investigated the engineering of knowledge-based approaches for multimedia retrieval, in particular, the ontologies [Mylonas et al. 2008, Hudelot et al. 2008, Mylonas et al. 2009, Elleuch et al. 2011, Paliouras et al. 2011a, Bannour & Hudelot 2014]. Yet, ontologies [Petridis et al. 2004, Wallace et al. 2005, Möller & Neumann 2008, Dou et al. 2015] (as a knowledge database) are powerful tools to design concepts/contexts and their interrelationships. In general, ontology-based approaches consist in defining a knowledge conceptualization and a reasoning process in order to handle and enhance a semantic interpretation. However, these recent approaches still facing some issues related to the construction of the ontology and the definition of its content.

Issue 2: dealing with uncertain knowledge and interpretations Multimedia content interpretation and the user perception give evidence to the uncertain nature of the retrieval task which can benefit from either fuzzy or probabilistic approaches. In order to model uncertain and imprecise knowledge for multimedia contents, many approaches were proposed. There are many discussions about these two approaches and their capabilities to handle and support uncertainty. In fact, endless arguments are defended by the knowledge extraction community about the effectiveness of fuzzy approaches and probabilistic ones [Gaines 1978, Bosko 1990, Sanjaa & Tsoozol 2007, Zadeh 2014; 2015]. In this dissertation, we focused on the use of a fuzzy approach. Thus, the use of fuzzy processing techniques to multimedia indexing and retrieval approaches has been extensively investigated in literature. Mainly, fuzzy retrieval models offer more flexibility to handle indexing terms, query terms and pertinent document ranking. The multimedia retrieval community took advantage by the use of fuzzy theory based models for knowledge representation and management.

Issue 3: Scalability Multimedia collections are increasing staggeringly. Thus, retrieving from large-scale datasets is a challenging task [Villegas et al. 2013, Villegas & Paredes 2014, Gilbert et al. 2015, Villegas et al. 2015]. The access to such an enormous content

has forced the multimedia retrieval community to look for advanced scalable approaches and techniques in order to make the availability of automated and efficient semantic annotation for such contents [F. Wang 2011, D. Zhang et al. 2012, Benavent et al. 2013, Sahbi 2013, Reshma et al. 2014].

Our present thesis work is part of the I-TV project led by the REGIMVID team. The latter project focuses on an intelligent and personalized access to a video data system via a complete process from low-level processing to viewing multimedia corpus. Our research works deal with issues of semantic understanding and reasoning for video/image contents. In the dissertation, we particularly focus on how to better understand and interpret an uncertain semantic content efficiently through the use of fuzzy ontologies. Our approach is based on defining a fuzzy knowledge dataset in order to enhance video/image semantic interpretation, and to overcome the scalability issue.

1.3 Aims and Contributions

As aforementioned, the aims of this thesis are threefold. (1) At a first step, we have attempted to suggest a knowledge based model to enhance the indexing accuracy within a multimedia retrieval system. (2) Subsequently, we have concentrated our contributions on an effective use of the knowledge based model in order to improve multimedia content analysis and indexing. (3) Finally, we have concentrated our contributions on the scalability issue when handling large-scale multimedia data and a considerable amount of semantic of their contents.

The figure 1.3 illustrates the general work-flow of the proposed approach for a knowledge based framework to enhance the indexing accuracy.

In this dissertation, we attempted to reach the following objectives:

Objective 1 (Obj_1) Dealing with a new knowledge based model for multimedia indexing. The objective is to develop a framework able to handle various information about a multimedia content, then to operate with this information in order to infer new information/knowledge through a reasoning process. Such a novel model has to define and highlight pertinent components for an efficient knowledge based indexing process [Zarka et al. 2011, Ksentini et al. 2012],

Objective 2 (Obj_2) Handling a fuzzy knowledge when reasoning with semantic interpretations in order to enhance and enrich them. This objective aims to define a semantic structure to model required knowledge, then to specify an automated ontology population from available annotated image/video datasets, and finally to handle ontology content evolving to further improve semantics capabilities through analyzing and revising inaccurate and irrelevant knowledge [Elleuch et al. 2011, Zarka et al. 2016],

Objective 3 (Obj_3) Ensure the capability to handle a large-scale multimedia indexing collection: the scalability While recent works focused on the use of semantic hierarchies to improve concept detector accuracy, this objective embodies the use of such hierarchies to reduce detector complexity and then, to handle efficiently large-scale datasets [Zarka et al. 2015].

Consequently, we denote that the outcomes of our thesis work will allow to address the following problems:

- Reducing the semantic gap through providing an effective knowledge based framework for enhancing a semantic interpretation,
- Reducing the uncertainty through the use of fuzzy reasoning framework with the aim of assessing the consistency of the indexing process,
- Ensuring the scalability aspect through handling hierarchical indexing process which scales well with large-scale multimedia content.

What was accomplished in our study is a novel ontology management which is intended to a machine-driven knowledge database construction. Such a method could enable semantic improvements in large-scale multimedia content analysis and indexing. Thus, the main contributions for our thesis work are enumerated as follows (see figure 1.3):

Contribution 1 (C_1)**:** A fuzzy knowledge based framework for multimedia indexing to handle various interpretations about a video content. The framework operates with these initial interpretations in order to infer enhanced interpretations through a fuzzy reasoning process,

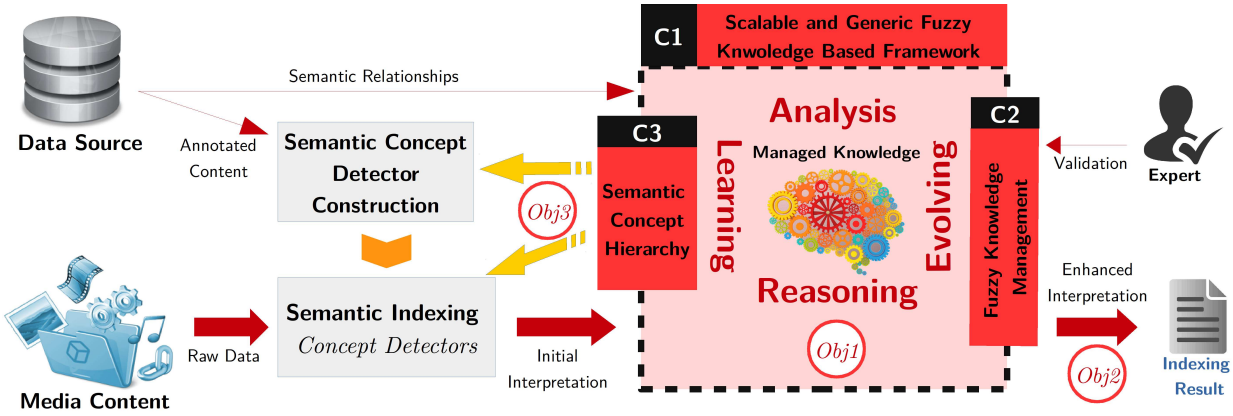


Figure 1.3: The proposed knowledge based framework: Handle a knowledge dataset to enhance the indexing process accuracy

Contribution 2 (C_2): An automated and scalable fuzzy ontology management approach (knowledge extraction, population, reasoning and evolving) for managing and enhancing fuzzy interpretation about a video content,

Contribution 3 (C_3): A scalable ontology-driven approach to construct hierarchical Semantic concept detectors. Fuzzy knowledge is used then at both learning and detection steps in order to enhance concept detectors accuracy, and to reduce the number of semantic concepts to be picked up within a content.

In order to illustrate the semantic enhancement of concept detection introduced by our proposed scalable and generic ontology-based framework, the following section enumerates different conducted experiments within multimedia evaluation campaigns.

1.4 Evaluation and Applications

In addition to the theoretical formulation of the proposed approach, we conducted some empirical evaluations over large-scale realistic data sets in the general domains. We first evaluated our approach within the standard benchmark TRECVID [Smeaton et al. 2006; 2009, Over et al. 2014]. We also applied our approach to real image data from different domains delivered by FLICKR data collections [J. Wang et al. 2009].

Different evaluation metrics, including *Mean Average Precision* (MAP) [Schoeffmann et al. 2012], *Inferred Average Precision* (infAP) [Yilmaz & Aslam 2008] and *Geometric Mean Average Precision* (GMAP) [Thomee & Popescu 2012] are used in our experiments.

Our evaluations also concern the scalability issue. We handled a large scale video dataset (about 8 000 hours from TRECVID) and a large-scale image dataset (up to 500 000 images from FLICKR).

Based on the proposed approaches and techniques, we participated to the development of the REGIMVID system [Elleuch, Zarka, et al. 2010, G. Feki et al. 2012, Ksibi et al. 2013]: a semantic video indexing, retrieval and visualization system.

Our conducted experiments can be epitomized within these tasks:

- *Semantic Video Indexing*: We first explored a multi-modal fuzzy fusion model to handle different semantic interpretation gathered by various modalities (text, audio, visual, ...). A practical system was designed to incorporate the latter fusion model. We used the developed system in order to enhance a video semantic indexing within the *Semantic Indexing* task of the TRECVID 2010 [Elleuch, Zarka, et al. 2010, Zarka et al. 2011, Over et al. 2011] evaluation campaign,

- *Photo Annotation*: After exploring knowledge database capabilities for enhancing a semantic interpretation, we proposed a fuzzy ontology based reasoning framework for managing valuable knowledge in order to open up a semantic interpretation about a content. The proposed framework was evaluated within the *ImageCLEF2012's Flickr Photo Annotation and Retrieval* [Thomee & Popescu 2012, Zarka et al. 2016] evaluation campaign,

- *Scalable Photo Annotation*: the aforementioned knowledge based framework being formulated and evaluated, we investigated a study on the semantic indexing scalability issue through considering an ontology based hierarchical image annotation. We evaluated, then, this framework within *ImageCLEF2015 Scalable Concept Image Annotation* [Gilbert et al. 2015, Villegas et al. 2015, Zarka et al. 2015] task.

1.5 Thesis Overview

At this point, many questions have arisen: what is an appropriate model/approach to introduce the knowledge management in an indexing system? Then, how to profit from a

knowledge database to enhance the indexing accuracy? Finally, how to cope with the scalability issue?

Keeping in mind our chronological progression in achieving the enumerated above objectives and contributions, the rest of the present dissertation is organized through three parts as follows:

i. ***Part I - Video Indexing: Background and Trends***

- **Chapter 2:** introduces a general overview of the Multimedia Information Retrieval. It describes also the main proposed models, the common assessment measures and the methodologies. The chapter ends with an enumeration of actual video indexing issues and trends.
- **Chapter 3:** introduces a survey of the use of fuzzy knowledge databases to solve problems discussed in the previous chapter. After enumerating actual related works, we show a discussion over the survey in order to motivate and introduce our proposed fuzzy ontology based model for video indexing

ii. ***Part II - Fuzzy Ontology Based Framework for Video Indexing***

- **Chapter 4:** displays a new approach for video indexing through the use of fuzzy ontologies. Our approach takes advantages of the fuzzy knowledge in order to enhance video indexing accuracy. Then, we focus on modelling and handling this fuzzy knowledge within an ontology. Therefore, an ontology management model is presented.
- **Chapter 5:** proposes to go further in the use of fuzzy ontologies models for enhancing video indexing accuracy. In particular, we address the following issues: how to extract valuable fuzzy knowledge? How to model this fuzzy knowledge within an ontology? How to reason with this fuzzy knowledge in order to improve a semantic interpretation? And finally, how to evolve the extracted knowledge? Therefore, we propose to automatically construct fuzzy ontologies to handle efficiently information about a video content.
- **Chapter 6:** explores the contribution of semantic hierarchies for image annotation in order to handle the scalability issue. Thus, we propose an ontology driven

hierarchical image annotation owing to reduce the number of semantic concepts to be detected within a content.

iii. ***Part III - Conclusions and Future Research Directions***

- **Chapter 7:** states a feedback on our contributions and discusses future directions that can be tackled by the presented research topics.

1.6 Publications

The aforementioned contributions in the fuzzy ontology based semantic enhancement for video indexing have been justified by the following scientific publications:

International Journal

- **Zarka, M.**, Ben Ammar, A., & Alimi, A. (2016). Fuzzy reasoning framework to improve semantic video interpretation. *Multimedia Tools and Applications*, 75(10), 5719–5750.
  Retrieved from `http://dx.doi.org/10.1007/s11042-015-2537-1`
  doi: `10.1007/s11042-015-2537-1`

International Peer-Reviewed Communications

- **Zarka, M.**, Ammar, A. B., & Alimi, A. M. (2011). Multimodal fuzzy fusion system for semantic video indexing. In *IEEE symposium on computational intelligence for multimedia, signal and vision processing, CIMSIVP 2011, Paris, France* (pp. 60–66). IEEE.

- Elleuch, N., **Zarka, M.**, Ammar, A. B., & Alimi, A. M. (2011). A fuzzy ontology: Based framework for reasoning in visual video content analysis and indexing. In Proceedings of the eleventh international workshop on multimedia data mining (pp. 1–8). New York, NY, USA: ACM.
  Retrieved from `http://doi.acm.org/10.1145/2237827.2237828`
  doi: `10.1145/2237827.2237828`

- Ksentini, N., **Zarka, M.**, Ammar, A. B., & Alimi, A. M. (2012). Toward an assisted context based collaborative annotation. In P. Lambert (Ed.), *10th international workshop on content-based multimedia indexing, CBMI 2012, Annecy, France, June 27–29, 2012* (pp. 71–76). IEEE.
  Retrieved from `http://dx.doi.org/10.1109/CBMI.2012.6269852`
  doi: `10.1109/CBMI.2012.6269852`

International Benchmarks

- Elleuch, N., **Zarka, M.**, Feki, I., Ammar, A. B., & Alimi, A. M. (2010). REGIMVID at TRECVID2010: semantic indexing. In P. Over et al. (Eds.), *TRECVID 2010 workshop participants notebook papers, Gaithersburg, MD, USA, November 2010*. National Institute of Standards and Technology (NIST).
  Retrieved from
  `http://www-nlpir.nist.gov/projects/tvpubs/tv10.papers/regim.pdf`

- **Zarka, M.**, Ben Ammar, A., & Alimi, A. (2015). Regimvid at imageclef 2015 scalable concept image annotation task: Ontology based hierarchical image annotation. In *Working notes for CLEF 2015 conference , Toulouse, France, September 8–11, 2015.*

Part I

Video Indexing: Background and Trends

Chapter 2

Multimedia Information Retrieval and Indexing

The Information Retrieval aims to define approaches and techniques to model, store and access information items [Baeza-Yates & Ribeiro-Neto 2011]. The information can take different forms: textual document, an audio sequence, a visual data (images) or a video clip. Thus, the multimedia information retrieval field engages many research areas as signal processing, machine learning, data mining, information theory, human-computer interaction and many other areas [Rueger 2010, Sheu et al. 2010]. Many researches have addressed an important interest to multimedia information retrieval in recent years. In fact, the raising of advanced multimedia devices together with the low cost storage devices have blown-up the production and the access to multimedia data content.

The present chapter describes basic concepts of multimedia retrieval and indexing which consist of the general architecture and the evolution of retrieval approaches from the text-based toward the concept-based approach. This chapter ends with an enumeration of open issues in semantic video indexing.

2.1 Multimedia Information

The multimedia content formats can vary according to the usage domain, for example web pages, video sequences, audio sequences, ... All these broadly different understandings of what a multimedia content help to interpret the semantics within a defined content. In

fact, for a given multimedia content, we can identify different basic data types (such as text, image, video, speech, audio), interactive elements (links, event on user action, mouse over, ...) and structural elements (text formatting, image and video location, ...). The syntax of a multimedia content is a collection of several different data types that provide rich information and enhanced experience for the reader [Stamou & Kollias 2005].

As depicted in figures 2.1 and 2.2, we conclude that humans segment a multimedia document into manageable blocks of information in order to construct a complete understanding of that content. Thus, in this thesis, we consider a multimedia content as an image and audio stream that carry some semantic information. The figure 2.2 illustrates a scene of car racing. While the first segment depicts some racing cars, the second segment depicts aligned cars at the start line. By identifying valuable objects and sounds within that content, we aim to generate a more complete and enhanced semantic interpretation. Video segmentation, visual objects detection and audio analysis are discussed in two other thesis works [I. Feki 2013, Elleuch 2015] carried out jointly with the present one.

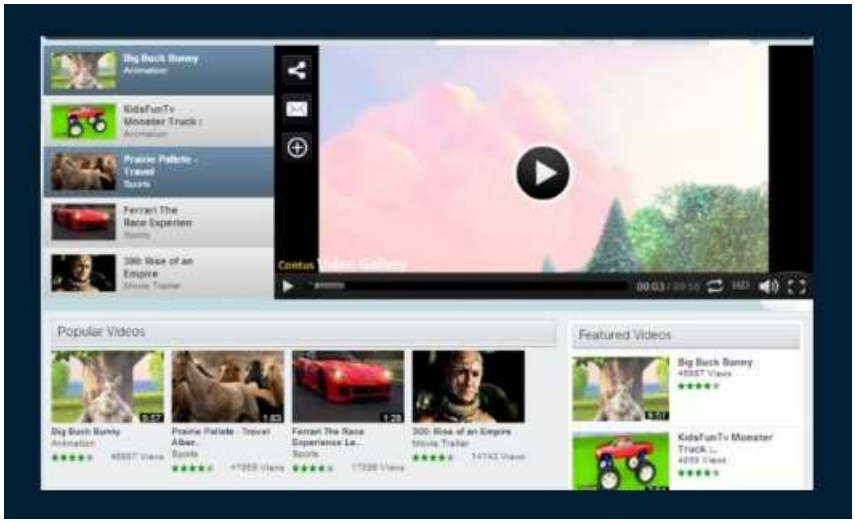


Figure 2.1: A Web page can be seen as a multimedia document

Unlike textual documents, a multimedia content contains more than symbols that a user can take on in order to express his need for information. At first, the multimedia content is rich in term of available semantic information: the visual content can expose a variety of messages and emotions, an audio content can also expose feeling and emotions, then, the structure itself (spatial-temporal information) also communicates valuable information. Thus, a multimedia content delivers a complete semantic interpretation to be communicated

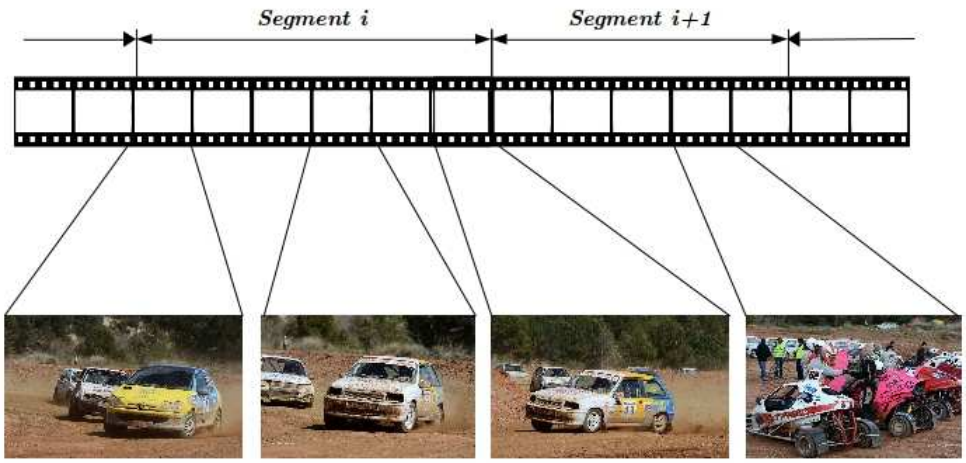


Figure 2.2: Video sequences can be seen as a multimedia document

to the user. Secondly, the computational system can process only mathematical and logic expressions, and not all what a human can express and interpret in term of ideas, emotions and feeling: the communication gap between the user and the system.

Many approaches were proposed in literature in order to empower the user with tools to express his query and achieve a better mapping between the expressed user need, the extracted multimedia information, and what the system can successfully match. Some of these techniques will be exposed in the following sections.

2.2 Multimedia Information Retrieval

In order to access a multimedia content, some techniques are used to interpret a human queries that translate its needs, and then to retrieve the closest match. As an example, a user searches its collection using the keyword or a phrase such as *car* or *car racing*, he will expect the information retrieval system to return all relevant items. Nevertheless, the user still disappointed with the given results. Such a disappointment can be rooted in two reasons. Firstly, the context in which the user has expressed its information need could be too vague and requires the user to further refine its query. Secondly, the weak and blurred link between the information representation schemes and the user semantic query. These two issues are called the *semantic gap* [Smeulders et al. 2000, Bahmanyar et al. 2015].

Low-level extraction approaches from visual contents (e.g., histograms, shapes, motions, ...) and audio content (e.g. volume, frequency, pitch, ...) are widely investigated, providing

a wide feature sets that could be explored for indexing and modeling a multimedia content. These low-level features rely on data-driven properties and characteristics, which may be misrelated to the concepts expressed in the semantic query.

The semantic information extraction process from multimedia contents is considered as an artificial intelligence problem. Nevertheless, Human perception is still being not well understood for imitating it in a computational system. In fact, an Information Retrieval (IR) system is considered as an artificial intelligence problem, on one hand, because the system has to mimic the human perception process in order to extract relevant semantics from a given multimedia content, and on the other hand, the system has to interpret the user query and match the relevant stored information. Then, the missing relationship between low-level features and human knowledge is the fundamental semantic gap problem when we need to search a multimedia collection by expressing some semantic queries.

Recent proposed multimedia analysis and indexing approaches rely mainly on manual annotations [F. Wang 2011, Lazaridis et al. 2013, Jin & Jin 2016]. Such approaches are flawed and costly, and can lead to a poor learning process. As depicted in figure 2.3, the user query is analyzed by the information retrieval system with the same system that can imitate human perception in order to process the query and match it to relevant information. The semantic gap exists in both: (1) multimedia content analysis, and (2) user query understanding.

In this scope, actual researches in multimedia retrieval are exploring new paradigms for semantic multimedia information analysis in order to deliver a higher degree for semantic information processing capabilities. In fact, the multimedia community addressed more intention on knowledge management based approaches for such a propose.

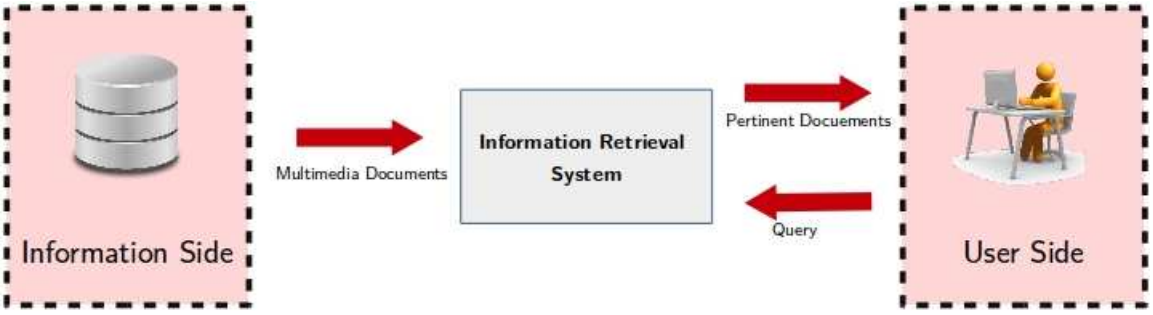


Figure 2.3: Information work-flow in Information Retrieval

We will continue, in the following, to display and identify the information retrieval main research directions.

The information system has been proposed since several decades. Most of proposed systems until mid 90^s handled only text data based documents [Manning et al. 2008, Marcia 2012]. From such an experience, a set of functional components were defined: (1) the analysis component that extracts a vocabulary from documents, (2) The indexing module that represents documents through its information symbols, (3) the query processing component that transforms a user information needs into information symbols, and (4) the retrieval component that ranks the stored documents representation according to the similarity measure between information symbols.

A multimedia information system, as displayed in figure 2.4, is similar to the traditional Information retrieval system detailed above, but with different algorithms: the multimedia analysis handles multimedia content unlike a text document. Thus, the information retrieval system architecture depicted in figure 2.4 can be addressed as a generic information retrieval system. In the following, we detail the components of this generic architecture.

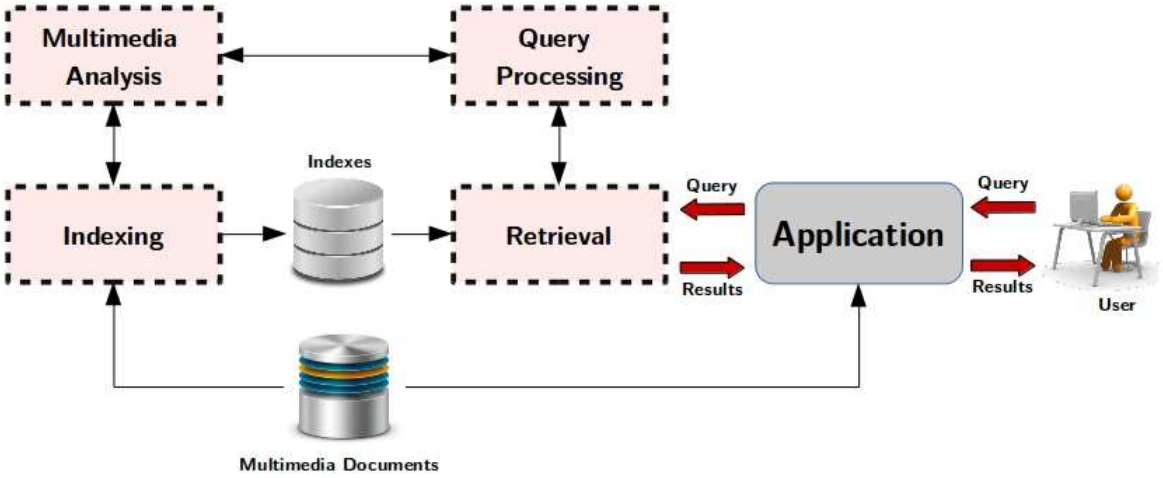


Figure 2.4: Classic Multimedia Information Retrieval Architecture

2.2.1 Multimedia Analysis

Information Retrieval systems analyze multimedia contents and extract features measuring the importance of information symbols. The extraction of these features aims: (1) to associate multimedia contents to meaningful symbols of information that a user can use to search for, and (2) to quickly find relevant document through information symbols indexing.

These information symbols re-extracted automatically, semi-automatically or manually [D. Zhang et al. 2012]. While the automatic information symbols extraction executes an analysis task without a human intervention, the semi-automatic one includes a human as part of the analysis task: some information can only be detected and added by a human, such as the name of a person, or the relation between two persons (e.g. friends). Many approaches were proposed, but more adequate to the considered information domain. As an example, GOOGLE [P.-I. Chen et al. 2011, S. Kumar et al. 2014] and FLICKR [Sigurbjörnsson & van Zwol 2008, Ginsca et al. 2014] page rank rely on some human intervention in order to edit information, and then improve search results. While FLICKR allows the user to tag images with a set of keywords that can be used later for searching images, GOOGLE relies on human edited links pointing to analyzed web pages in order to adjust its importance. We conclude that FLICKR adopts a semi-automatic approach, and GOOGLE adopts an automatic one since it relies on stored information.

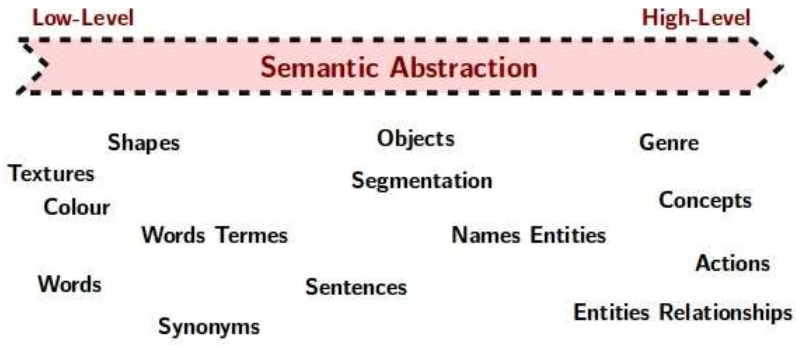


Figure 2.5: Relation between information symbols and semantic abstraction

The figure 2.5 displays some feature positions on a semantic abstraction level: from low-level features to high-level ones. While low-level features, such as color histogram, are easily extracted from a multimedia content through automatic methods, high-level features, such as content topic or objects existing in the content, require more complex approaches due to the high semantic dimension that they involve.

Classical text analysis approaches produced a small set of features (e.g. occurring words). However, multimedia analysis approaches generate scores, such as a probability, for all possible features (information symbols). Such a situation will induce to produce dense high-dimensional vectors that contain all features extracted from all multimedia documents, and

then cause several issues such as storage space. Thereby, the multimedia analysis approaches are important since they will impact the entire multimedia information system. So, many techniques and algorithms are provided to dress these issues.

- *Low-Level Analysis:* The low-level analysis extracts features from a multimedia content. These features are commonly related to human senses or language (e.g. image colors, image textures, audio rhythms, word, ...). Such features are well studied in literature and most of them have been developed in the area of data compression that exploits the characteristics of human vision and hearing senses [Förstner 1994, Nixon et al. 2012].

  These low-level features are used then to index multimedia contents.

- *High-Level Analysis:* The High-level analysis aims to extract inferred information from a multimedia content. This extracted information can be not explicitly detectable by a computer. Then, a prior-knowledge about the problem domain semantics should be involved. Such domain centered knowledge is formally described by a set of semantic concepts identified by keywords. The latter capture part of domain knowledge which can be used to detect the presence of a concept in a given multimedia content.

  The high-level features (concept/keyword) are used as index tokens that the information retrieval system uses to build the index of a semantic-multimedia content.

2.2.2 Indexing

Information features (symbols) extracted from a multimedia content are stored by the indexing component. While the analysis component impacts the effectiveness of the multimedia information system, the indexing one impacts the efficiency of the system. The main aim of the indexing algorithm is to manipulate an inverted-file index that enumerates all information symbols and all related documents that contain these symbols.

The indexing component uses a high-dimensional index to accommodate the high-dimensional multimedia content data nature (as the multimedia analysis component produces a score for all possible feature type and dimension). The efficiency of high-dimensional indexes is related to several aspects [Ai et al. 2013]: tree-structured indexes, hash-based indexes, compression of the index to reduce memory usage, ... All these techniques allow a more faster and

efficient look-up of the index table. In [Baeza-Yates et al. 1999, Baeza-Yates & Ribeiro-Neto 2011], the authors discussed the efficiency of the indexing component.

2.2.3 Query Processing

When the user communicates query to the information retrieval system, the latter analyses and transforms the query into an internal representation that used the same symbols for indexing multimedia contents.

The query processing component parses the user query according to a specific language, extracts information symbols and features, and proceeds it to the retrieval component to search index for the matching documents.

Text based information retrieval systems handle text queries as a query language support. In multimedia information systems, the queries can be expressed in many ways [Y. Liu et al. 2007, Yasmin et al. 2014]. In what follows, a brief enumeration of some of these queries methods.

- *Sketch Retrieval:* this is one of the first proposed methods to query a multimedia database [Cao et al. 2010; 2011]. The user query is in the form of a visual sketch of what the user expects to find. The information retrieval system proceeds then the extraction of features from that sketch in order to search the index for images that are visually similar.

- *Search by Example:* the user can submit an example image representing the information he is searching for. In such a case, the query processing extracts the low-level features for the query in order to look for similar stored documents with similar features.

- *Search by Keyword:* this is by far the most popular search query method: the user describes his information request with a set of keywords. Then, the system searches for multimedia content that corresponds to these keywords. The issue of this high-level query method is that the user can only use some predefined keyword vocabulary used also in multimedia content indexes.

- *Personalized/Adaptive Retrieval:* this is a refinement to all other search methods. It explores the history and profile of the user search [Magalhães & Pereira 2004]. This

extra-information can enhance the search experience by filtering information to particular domains, limiting certain document formats, ...

2.2.4 Retrieval

The retrieval component computes indexed document ranks according to their similarities to a user query. This component browses and sweeps the index according to the input query information symbols to search for most similar documents according to the relevance degree. The latter is computed by a function that determines the semantic relatedness degree between the query and the retrieved documents. Thus, the result delivered by the retrieval component is a set of ordered documents that are judged as similar documents to the given query [Memar et al. 2013].

The challenging problem in the ranking task is to find which documents are relevant and which are irrelevant. The relevance degree between the document and the user query is computed through various retrieval models. The information retrieval community proposed different retrieval models [Baeza-Yates et al. 1999, Croft et al. 2010] such as: Boolean, vectorial, probabilistic and fuzzy models.

- *Boolean Model:* is the oldest and the classical retrieval model. It is based on the Boolean algebra. With such a model, a query is defined as a Boolean expression, and the relevance of a document to a query is computed by the use of Boolean operators like *And*, *NOT* and *OR*. One limitation of the boolean model is that computed relevance weights are equal for all relevant documents: it will be difficult then to identify the most relevant documents between the less ones.

- *Vector Space Model:* is the most popular and used retrieval model. Both queries and documents are handled as vectors in the space. Then, the similarity between a query and a document is estimated by the use of direction and distance [Manning et al. 2008].

- *Probabilistic Model:* is a model that ranks document according to the probability of relevance to a given query. This model uses the probability ranking principal detailed in [Fuhr 1992]. Thus, a statistical distribution is used in order to measure the relevancy between relevant documents and irrelevant ones.

- *Fuzzy model:* Fuzzy logic plays an important role in many domains that handle imprecise and vague information [De Mantaras et al. 2015], such as text-mining, multimedia information system, machine learning, ... To handle uncertain information, the fuzzy logic allows intermediate truth values to be defined between conventional evaluations of true (1) and false (0). Fuzzy logic was used as a fuzzy retrieval model [Tahani 1976, Cross 1994, Miyamoto 2012].

2.3 Multimedia Retrieval Systems Evaluation

Information Retrieval has been widely addressed in research work, and it has been shown to be highly useful to compare different systems and approaches. In literature, two different measures are used: Firstly, an effectiveness metric is used to measure how well the system can satisfy the user information need. Secondly, the efficiency metrics measures both, the system responsiveness to the user query, and the system ability to cope with large-scale situations.

Effectiveness and efficiency measures are widely affected by the data that is used to test a system. In fact, a dataset can contain information with different complexities. Therefore, evaluation methodologies are now being investigated and standardized. In the following, we enumerate traditional metrics and resources used in the evaluation of information retrieval systems.

2.3.1 Effectiveness Evaluation

In order to introduce the retrieval effectiveness measure, we discuss the meaning of relevance. The relevance is a fundamental concept of information retrieval. The paper [Mizzaro 1997] is considered as the first work that focused on the relevance concept: *the relevance is claimed as a complex concept involving different aspects: methodological foundations, different types of relevance, beyond-topical criteria adopted by users, modes of expression of the relevance judgment, dynamic nature of relevance, types of document representation, and agreement among different judges.*

Therefore, many research areas adopt their own definition of relevance focusing more on their specific objectives: the information retrieval aims to identify documents that can be judged as the best answer to a given information query. Information retrieval relies on

document datasets where their relevance of a given query was judged by a human. Nevertheless, there is no general definition of what a relevant document is. In fact, the relevance of document is a diffuse information because a given document can have different meanings to different humans. This issue was widely discussed in literature that noticed a divergence between relevance judgment made by different human for a given document [Voorhees 2000, Volkmer et al. 2007]. This divergence is more observable in large multimedia datasets. In fact, a multimedia content is by nature subject of different interpretations and relevance annotation (judgment), also, human relevance judgment process is a very costly task and large-scale document datasets will be partially annotated (incomplete and inconsistent).

Precision and *recall* [Buckland & Gey 1994] are the popular information retrieval metrics. They are applied on a ranked list of both relevant and non-relevant documents for a given query. While the precision addresses the accuracy of the evaluated system, the recall addresses the completeness aspect. We enumerate, in the following, these two metrics and other derived ones.

- ***Precision (Prec):*** which measures the ability of a system to present only relevant items. Mathematically, the Precision ($Prec$) is defined as the number of true positive documents (T_p) over the number of true positives plus the number of false positives (F_p).

$$Prec = \frac{T_p}{T_p + F_p} \tag{2.1}$$

- ***Recall (Rec):*** which measures the ability of a system to present all relevant documents. Mathematically, the Recall (Rec) is defined as the number of true positive documents (T_p) over the number of true positives plus the number of false negative documents(F_n).

$$Rec = \frac{T_p}{T_p + F_n} \tag{2.2}$$

- ***F-measure (Harmonic mean):*** is a measure which assesses the trade-off between precision and recall. The *F-measure* is computed as follows:

$$F = \frac{2}{\frac{1}{Prec} + \frac{1}{Rec}} \tag{2.3}$$

- ***Average Precision (AP):*** is obtained after each relevant document is retrieved. Let k be the number of relevant retrieved documents, the average precision expression is:

$$AP = \frac{\sum_{k \in \{r | r \text{ is rank of relevant docs}\}} Prec@k}{\text{Number of relevant Documents}} \tag{2.4}$$

- ***Mean Average Precision (MAP):*** is a metric which summarizes the overall system retrieval effectiveness into a single value as the mean of all keywords average precision,

$$MAP = \frac{1}{|Q|} \sum_{q \in Q} AP_q \qquad \text{where } Q \text{ is a set of queries} \tag{2.5}$$

These metrics measure the effectiveness of a given system in fixed evaluation scenario. Thus, the obtained measures are valid only for a specific scenario and cannot be generalized to other situations.

2.3.2 Efficiency Evaluation

The efficiency measurement in multimedia information system concerns the extra computational complexity over conventional information retrieval systems. This extra complexity is due to the extra processing required to analyze a multimedia content and extract the contained semantics.

In the context of our thesis, we solely focus on the runtime complexity of the analysis component. In fact, and as discussed in previous sections, the information indexing involves the indexing and the storage of extracted low-level and high level features. The efficiency of the indexing task measures both:

- *Time complexity:* which measures how many documents/concepts per second can the indexing task processes. Thus, the time complexity measures the system responsiveness.

- *Space complexity:* which measures the amount of memory required to process a document for the entire vocabulary. The space complexity measures the scalability aspect.

These two complexities define how well a system scales with several simultaneous requests.

2.3.3 Collections

In the assessment of semantic multimedia retrieval systems, multimedia collections are used as research tools that provide a common test environment in order to evaluate and compare different approaches and techniques.

Collections are used to evaluate a variety of techniques, such as shot-boundary detection, low-level visual features, story segmentation, ... In this thesis, we address the problem of indexing multimedia content by their semantic content. One aspect is required to be present in a collection that will be used for the assessment of multimedia indexing: keywords that correspond to concepts present in the collection content. They describe what meaningful concepts are present within a given multimedia content.

Multiple benchmarking initiatives have been proposed which aim the assessment of multimedia retrieval system performances with standardized test collections and tasks, such as in ImageClef [Villegas et al. 2013, Villegas & Paredes 2014, Villegas et al. 2015] and TrecVid [Smeaton et al. 2006, Over et al. 2013; 2014]. In what follows, we enumerate the multimedia collections/benchmarks addressed in this thesis work.

i. **ImageClef:** ImageClef is an evaluation forum for the cross-language image retrieval. The main goal of that benchmark is to provide the required infrastructure for the evaluation of visual information retrieval systems operating in monolingual, cross-language and language-independent contexts. Thus, ImageClef allows multilingual users accessing the growing visual multimedia data, and creates a public resources for benchmarking information retrieval systems and approaches.

   Since its start in 2003, ImageClef has been a track in the *Cross Language Evaluation Forum* (Clef). The latter is one of the major forums for research in information retrieval. The main goal is to create collections and topics in order to offer to the information retrieval community an opportunity to evaluate and to compare theirs approaches.

   At first, ImageClef focused on text-based image retrieval. But since 2003, the focus has shifted towards combining visual and textual features for multi-modal image retrieval on general images collections, in particular medical ones.

ImageClef is addressing the barriers between research interests and real-world requirement by offering application-driven evaluation tasks. One of these tasks that this thesis focused on is the *Photo annotation* one.

In ImageClef 2012, a collection of 25 000 images was provided: 15 000 images for the development, and 10 000 image for the test. The development dataset was manually labeled with 94 semantic concepts. But in ImageClef 2015, the collection has shifted to 500 000 and 240 concepts, but with a very small development dataset (about 1 500 images).

ii. **TrecVid:** Toward an efficient and effective management of multimedia collections, many research works interest arising focusing on the combination of multimedia interpretation, extraction, retrieval and management. This growing requirement has initially resulted in the creation of a video retrieval track TrecVid within the Trec conference. Actually, the TrecVid becomes a workshop in its own right.

Founded in 2003, the TrecVid aims to promote and encourage research works in content-based video retrieval and indexing through providing large test collections, realistic system tasks, uniform procedures, and a forum for different researchers to compare their results.

TrecVid provides a several number of tasks, and in this thesis work, we focused on the *semantic indexing task*. The latter provides a collection of 400 hours of video sequences: 200 hours for the development dataset, and another 200 hours for the test one. In 2010, TrecVid defined 130 semantic concepts, and in 2015, it defined 500 concepts.

2.4 Open Issues in Semantic Multimedia Retrieval

In the past few years, remarkable advances have been done relatively to the video indexing at low-level as at high and semantic levels. Nevertheless, many open research issues still be open and need to be more addressed in order to make an efficient use of video retrieval systems [Hole & Ramteke 2015, Tunga et al. 2015]. In what follows, we identify some open challenges.

High-Dimensional Indexing The dimensions of extracted feature vectors used in most indexing systems to represent video contents, are quite high (1024 dimensions in [Elleuch et al. 2015]). There is a practical requirement to reduce the intrinsic dimension of these feature vectors.

Similarity Matching The matching process within video retrieval requires similarity measurement for evaluation feature vectors similarities. Classical retrieval systems use *Euclidean* and *cosine* distance measure which fails to successfully simulate human perception similarity for a video content [Juneja et al. 2015]. Some similarity functions have been proposed in literature [Shen & Cheng 2011]. Fuzzy similarities functions enable more accurate similarity measurement [Y. Chen & Wang 2002, Chaira & Ray 2005, Baccour et al. 2013; 2014, Kraft et al. 2015], but fail to deliver fast retrieval and to guarantee a better scalability [Shen & Cheng 2011].

Relevance feedback First video retrieval systems were focused on fully automatic processing, but they were unable to really produce promising and satisfactory results. One of the occurred issues was the fact that user satisfactions of given results were not taken into consideration. Thus, these systems included the use feedback in order to learn from the user intervention and, then, to regulate the system process for generating more satisfactory results. The relevance feedback can either be taken from the user directly, or it can come from intelligent approaches that capture and trace the user trends and profile.

Low-level/high-level semantic gap Many of signal processing and computer vision contributions have been mostly explored in literature for enhancing the video indexing and retrieval accuracy [Brunelli et al. 1999, Y. Liu et al. 2007]. More recent research works were focused on the high-level description and retrieval. Actually, research works that bridge the semantic gap between signals (as an example: pixels for images and frequency for audio) and implicit semantic information are taking a growing interest. Knowledge management and reasoning are really taking more intention and are opening many semantic barriers.

Performance Evaluation and Standard Video indexing and retrieval is an area which highly requires standardization and uniform evaluation criteria in order to judge how well the system is performing and how it performs better than other systems. Many evaluation campaigns (like TrecVid [Smeaton et al. 2006, Over et al. 2013; 2014] and *ImageClef* [Villegas et al. 2013, Villegas & Paredes 2014, Villegas et al. 2015]) proposed both development and test dataset, and precision and recall based metrics.

Generic and multi-modal approaches Expected video indexing and retrieval systems have to handle information and video contents from different fields and disciplines. Moreover, they should take into consideration and analyze all information from all the available modality of a given content (text, visual, audio, ...). Video information retrieval field is compelled to integrate multi-modal signal processing, computer vision, Artificial Intelligence, knowledge management, ...

Scalable Indexing and Retrieval The high dimensions of the extracted features from a video content, and the contained rich semantics represent massive data to be handled and computed. Efficient and scalable video indexing and retrieval systems have to avoid overwhelming tasks. Scalable indexing and retrieval techniques have to handle non overwhelming algorithms and tasks. Thus, Scalable systems should ensure that the computing cost does not scale exponentially with the amount of video data content and the semantics to be handled.

In spite of many research efforts to compact video extracted features [Douze et al. 2010, Baroffio et al. 2013, S. Wang et al. 2015, Z. Li et al. 2015], and to propose scalable indexing and retrieval techniques and approaches [Caputo et al. 2014, Gilbert et al. 2015], real scalable information retrieval systems still far from concretization.

2.5 Conclusion

As denoted above, the multimedia retrieval community is tackling more robust indexing systems in terms of effectiveness (that concerns in term of high semantic capabilities) and efficiency (that concerns the scalability and high-dimensional indexing abilities). In the chap-

ter, we show how the knowledge based approaches are considered as compelling direction to overcome the above issues.

Chapter 3

Video Indexing: From signal processing to knowledge reasoning

Since multimedia resources are playing an increasingly outstanding role in our lives, the previous two decades have marked a great move of research for semantic analysis of multimedia contents in order to enable computational interpretation and processing of such resources. In fact, video analysis and retrieval topic have attracted considerable interest from both industry and academia, in which the key technology copes with data overload problem. The overriding interest of this topic is to model, represent and organize the large multimedia data for further efficient use and accessibility. The main aim can be stated as to provide pertinent multimedia contents that are relevant to a user inquiry [Jaimes et al. 2005, Lew et al. 2006, D. Feng et al. 2013].

The main focus of this chapter is to explore the research achievements in multimedia information retrieval and indexing, mainly in the enhancement of large-scale video content analysis and indexing [Dasiopoulou et al. 2005, C. G. Snoek & Worring 2008, Gani et al. 2015]. Thus, the present chapter sheds light on actual trends and opportunities to enhance multimedia content indexing efficiency through the use of ontology based approaches.

3.1 From low-level to knowledge based indexing

Ontologies, as knowledge database, have been emerged from an interesting conceptualization paradigm to a very promising modeling technology for multimedia retrieval. Ontologies

enable meaning driven retrieval process through a machine-understandable form of a content description.

Earlier research in multimedia feature extraction focused mainly on the visual modality and involved frame-based and object-based approaches [Puri & Chen 2000, Deb 2004, C. G. M. Snoek & Worring 2005]. Frame-based approaches deal with low-level features (such as histograms, colors, textures, motion, ...) in order to detect a shot or to query by example [Brunelli et al. 1999, Antani et al. 2002, Kang 2003, Smith et al. 2003]. These approaches became standard procedures and were easy to compute, however, they were not suitable to detect fine-grained semantics in a multimedia resource. Thus, frame-based features were not able to get detailed semantic interpretation, but basically macro-grained (particularly for detecting the subject of a multimedia resource like sports, news, ...). Object-based approaches were proposed in order to overcome low-level based features limits for the description of a multimedia content and to fill the gap between perceptual properties and semantic meaning of a multimedia content. In fact, object-based approaches deal with low-level features for an individual region instead of the whole content. This means that object-based approaches are suitable to detect high-level features like *"table"*, *"chair"*, *"car"*, *"person"*, ... [C. G. M. Snoek et al. 2006, Lew et al. 2006, Spyrou & Avrithis 2008]. Many models have been proposed to identify semantic concepts in images [Jurie & Triggs 2005, Yang et al. 2007, Z. Wang et al. 2010] and audio [You 2010, I. Feki et al. 2011, Rawat et al. 2013]. However, all these models raised one fundamental question: *Which semantic concepts should these models focus on and deal with?*

In order to promote researches on multimedia analysis and to deliver a common set of semantic concepts, the *Moving Pictures Expert Group* (MPEG) have proposed the MPEG-7 [Salembier & Sikora 2002] standard for describing a multimedia content. The aim of MPEG-7 is to address a wide variety of media types and to describe audiovisual information through providing a rich set of standardized tools that generate and understand audiovisual features. Hence, the MPEG-7 standard defined more than 140 semantic concepts that can describe a multimedia content. However, as thoroughly elaborated in [Naphade et al. 2006], some practical obstacles have hindered the emerging research works and efforts in the multimedia content analysis field, and the MPEG-7 received a weak attention from multimedia research community. Firstly, many semantic concepts defined in MPEG-7 were not suitable for an

automated detection. For an example, it was very hard to multimedia content analyzers to discover a semantic concept like *"remarkable people"* (defined in MPEG-7). Secondly, most of multimedia retrieval approaches were based on statistical machine-learning techniques [Deb 2004]. These latters use annotated datasets to build models and classifiers. However, datasets (and particularly training ones) were insufficient and non-standardized to promote researches on multimedia semantics.

To address these two issues, and to supply large annotated multimedia datasets that support a common set of semantic concepts, the LSCOM ontology (for *Large Scale Concept Ontology for Multimedia*) was proposed [Kennedy & Hauptmann 2006, Naphade et al. 2006]. At first, a set of 1 000 semantic concepts were defined and 80 hours of broadcast news video were manually annotated aiming at providing a valuable resource for the multimedia research community. Actually, LSCOM provides more than 2 500 concepts. Many concept detection researches began to be used successfully by exploring the LSCOM ontology: *VIREO-374* [Jiang et al. 2007; 2010] and *Columbia374* [Yanagawa et al. 2007] are able to detect up to 374 semantic concepts in a visual content, and *Mediamill101* [C. G. M. Snoek et al. 2006] is able to detect up to 101 semantic concepts. LSCOM is used also as a basis of many video annotation tools [Garnaud et al. 2006, Worring et al. 2006, Ksentini et al. 2012].

Evaluation campaigns have played a significant role for the progress in semantic concept detection within a multimedia content; The evaluation campaign TRECVID [Smeaton et al. 2009, Over et al. 2013] has played the most significant role [C. G. Snoek & Smeulders 2010] by exploring the LSCOM ontology resources. TRECVID intends to benchmark search engines and to promote the content-based retrieval via open metrics-based evaluation. As outlined in [C. G. Snoek & Smeulders 2010, Over et al. 2013], experiments in TRECVID have led to conclude that available semantic concept detection approaches cannot be generalized to any semantic concept. Indeed, these approaches focus on identifying objects in a content without dealing with implicit information: concepts co-occurrence and the context in which an object is defined [L. Feng & Bhanu 2012; 2016].

Concepts co-occurrence Earlier object detection approaches considered that a detector is modeled for a single semantic concept. It also means that for detecting a set of semantic concepts in a content, a set of detectors are simultaneously taken into consideration.

Nevertheless, a semantic relationship between this set of concepts could be explored. Thus, authors in [Naphide & Huang 2001] displayed a probabilistic model to explore inter-concept relationships. Many other works [L. Feng & Bhanu 2012, Zheng et al. 2013] (to cite a few) followed this promising track by computing similarities between detected concepts from an annotated multimedia dataset (a training dataset). As an example, when the semantic concepts *"sand"* and *"sky"* are detected in a content with a certain probability, a chance to consider that the concept *"desert"* is present should be increased (even if this concept was not detected), and a chance to consider that the concept *"Penguin"* is present should be decreased. Therefore, the co-occurrence of concepts is taken a serious consideration in multimedia retrieval community.

Contexts In general, a multimedia content interpretation is an outcome of a defined context in which contained semantic objects are defined [Dumitrescu & Santini 2009]. Therefore, the *"context"* emerged as a great opportunity to contribute to multimedia analysis enhancement [Elgesem & Nordbotten 2007, Jiang et al. 2009, Hori & Aizawa 2003, Fauzi & Belkhatir 2014, L. Zhang et al. 2014, Schoeffmann et al. 2015]. Many context-based multimedia retrieval systems used the context approach with an informal definition [Cioara et al. 2009, Nguyen 2010, Parsons et al. 2009, Perpetual Coutinho et al. 2012, Hamadi et al. 2014]: In fact, contexts are defined manually by authors. Consequently, these approaches are based on manual list of contexts to consider and a set of semantic concepts that are defined under each context.

In summary, context based approaches and semantic co-occurrence exploration are capturing the attention of the multimedia retrieval community and are being considered as promising research trend toward better semantic interpretation capabilities for multimedia contents. Yet, these new approaches are moving increasingly toward availing knowledge management capabilities (like ontology) in order to handle concepts, contexts and their relationships.

3.2 Semantic Multimedia Indexing: Towards Knowledge-Based Approaches

In this section, we exhibit an overview on some research works that targeted to explore knowledge databases in order to enhance the multimedia analysis accuracy. We display how these databases are modeled, then we discuss their expressiveness levels, and after enumerating related works, we talk over our motivation to propose a knowledge-based approach in order to enhance semantic interpretations.

3.2.1 Ontology Modeling

Many research works expose different models for developing and managing ontologies [Sure et al. 1999, Fernández-López 1999, Noy & Mcguinness 2001, Davies et al. 2005, Gargouri & Jaziri 2010, Terkaj & Urgo 2014, Zablith et al. 2015]. Commonly, ontology modeling consists in defining some steps in order to represent the main tasks to build ontologies starting from an existing knowledge source. The most important steps are: (1) the ontology structure, (2) the ontology population, (3) the reasoning process, and (4) the ontology evolution.

Ontology structure The ontology structure is the process of knowledge organization by defining concepts and their expected relationships. *OWL* (*Web Ontology Language*) is a standard used for modeling and exchanging ontologies and is designed to support the *Semantic Web* [Staab & Studer 2009]. Semantically, *OWL* is based on *Description Logics* (*DL*) [Baader et al. 2003]. *OWL* ontologies are categorized into three types depending on their expressive level: from *OWL-Lite* to *OWL-DL*. These two expressive levels differ in their complexity and may be used depending on required inference simplicity or formality of descriptions.

Ontology Population Ontology Population is the process of knowledge acquisition by analyzing and transforming unstructured, semi-structured and/or structured source data into ontology instances. This process looks for identifying instances of concepts and interrelationships of an ontology. Manual population by a domain expert is a costly and time consuming task, then, automatic/semi-automatic approaches are considered [Song et al. 2009, Faria & Girardi 2011].

Inference and Rules Rules are of the form of an implication between an antecedent (*body*) and consequent (*head*). So, when combining ontologies and rules for domain conceptualization and inference modeling, ontologies can enhance and enrich the amount of knowledge that it can represent.

Ontology Evolution The ontology content has to evolve continuously throughout its life cycle in order to be able to answer different change requirements. The Ontology evolution seeks to grow the background knowledge in order to better enrich its semantic capabilities. This evolution process consists in updating and validating concepts and semantic relationships [Gargouri & Jaziri 2010, Paliouras et al. 2011b, Petasis et al. 2011]. Current ontology evolution works give more attention to the way to enrich the ontology content: from fully automatic to fully manual. Typically, a content validation by a domain expert is considered as the most common process to ensure a valuable ontology content.

Ontologies are powerful tools to specify concepts and their interrelationships. Several research areas have focused on ontologies for knowledge management. More particularly, information retrieval systems use ontologies in order to enhance a machine ability to understand the document semantic contents. Nevertheless, the very specific nature of these systems requires to handle ontologies with specific considerations.

3.2.2 Ontology Expressiveness

As discussed above, ontologies could have a varying degrees of expressiveness: *OWL-Lite* and *OWL-DL*. We discuss in the following these two levels of expressiveness:

Light Ontologies:

are modeled as direct graphs in which nodes represent semantic concepts. The relationships between these nodes depict links that associate corresponding concepts and that express semantic nearness [Reimer et al. 2012].

These ontologies were used under a varying shapes in order to improve multimedia retrieval and indexing. For instance, concept hierarchies [Naphade et al. 2006, Deng et al. 2009], visual taxonomies [Fei-Fei & Perona 2005, Griffin & Perona 2008, Yao et al. 2010], and

semantic hierarchies [Fan et al. 2008, L.-J. Li et al. 2010] have been used for visual content annotation.

Light ontologies have been used to narrow the semantic gap. In fact, many works have reported an improvement in visual multimedia (image) indexing and annotation accuracy [Tousch et al. 2008, Yao et al. 2010, Martinet et al. 2011]. Nevertheless, light ontology based approaches are just used to model explicit knowledge without exploring implicit one: it is leading importance to explore the inter-concepts correlation (explicit knowledge) in order to discover and deduce newer (implicit) knowledge [Bannour & Hudelot 2011, Dingli 2011]. In fact, these approaches have used neither the expressiveness nor the reasoning capabilities provided by ontologies.

Formal Ontologies

The description language (DL) [Baader et al. 2003] are a formal knowledge representation formalisms that are used to model the terminological knowledge of application domain in a structured manner. DL have been used in various application domains, but their most noted success is the adoption of the DL-based language OWL [Horrocks et al. 2003] as the standard ontology language for the semantic web [Berners-Lee et al. 2001].

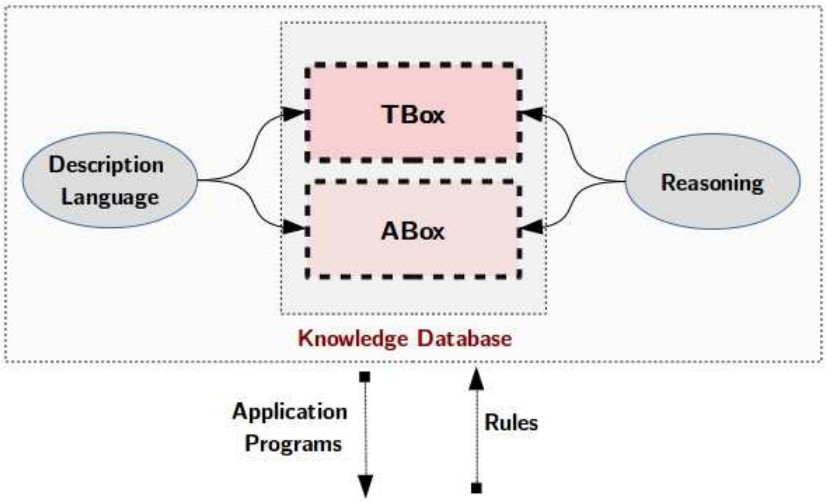


Figure 3.1: A knowledge Representation System Architecture of a knowledge representation system based on Description [Baader et al. 2003]

Description Language-based knowledge base $\mathcal{K}$(or an ontology) can be defined as a pair $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ where $\mathcal{T}$ (called the *TBox*), is a set of *concept axioms* and *role axioms*, and $\mathcal{A}$ (called the *ABox*), is a set of *assertional axioms* (see figure 3.1).

The *concept axiom* has the form $C \sqsubseteq D$ where C and D are concept expressions. The *role axiom* has the form $R \sqsubseteq S$ where R and S are role expressions. Finally, the *Assertional axioms* have the form $C(a)$ where C is a concept and a is an individual name, or that have the form $R(a, b)$, where R is a role and a and b are individual names.

In the Description Language, an interpretation I is defined as follows: $I = \mathcal{I} = \langle \triangle^{\mathcal{I}} =, \cdot^{\mathcal{I}} \rangle$ that consists of a non-empty set $\triangle^{\mathcal{I}}$ and an interpretation function $\cdot^{\mathcal{I}}$, which maps from individuals, concepts and roles to elements of the domain, subsets of the domain and binary relations on the domain, respectively.

For a given interpretation $\mathcal{I}$, we can say that $\mathcal{I}$ satisfies a concept axiom $C \sqsubseteq D$ (respectively $R \sqsubseteq S$) if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ (respectively $R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$). Moreover, $\mathcal{I}$ satisfies a concept assertion $C(a)$ (respectively $R(a, b)$) if $a^{\mathcal{I}} \in C^{\mathcal{I}}$ (respectively $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$).

An interpretation $\mathcal{I}$ is called a model of an ontology $\mathcal{K}$ if and only if it satisfies each axiom in $\mathcal{K}$. A concept name C in an ontology $\mathcal{K}$, is unsatisfiable if for each model $\mathcal{I}$ of $\mathcal{K}$, $C^{\mathcal{I}} = \varnothing$. An ontology $\mathcal{K}$ is incoherent if there exists an unsatisfiable concept name in $\mathcal{K}$. An ontology $\mathcal{K}$ is inconsistent if and only if it has no model.

DL languages are identified with the concept constructors that they allow. For instance, the minimal propositionally closed language allowing for the constructors $\sqcap$ (conjunction), $\sqcup$ (disjunction), $\neg$ (negation), $\forall$ (value restriction) and $\exists$ (existential restriction) is called $\mathcal{ALC}$. In [Baader et al. 2003], the author enumerates all the constructors used to identify various Description Logic formalisms ($\mathcal{N}$, $\mathcal{Q}$, $\mathcal{O}$, $\mathcal{F}$, $\mathcal{U}$, ...).

Once a description of application domain using DL, the latter can make inference through deducing implicit consequences from the explicitly represented knowledge. The subsumption is the basis inference on concept descriptions: given two concept descriptions C and D, the subsumption problem $C \sqsubseteq D$ is the problem of checking whether the concept description D is more general than the concept description C. Then, the subsumption problem tries to determine whether the first concept denotes in every interpretation a subset of the set denoted by the second one. C is subsumed by D, with respect to a *TBox* $\mathcal{T}$, if in every model of $\mathcal{T}$,

D is more general than C, i.e., the interpretation of C is a subset of the interpretation of D. This can be denoted as $C \sqsubseteq_{\mathcal{T}} D$.

Description language based ontologies have been successfully used to achieve some interesting reasoning capabilities in the context of semantic multimedia analysis. In fact, such an ontology formalism enables an intensive use of inference rules on the managed knowledge to perform advanced multimedia semantic analysis, attaining thus a semantic interpretation that a human perception and cognition can deliver. In order to achieve such an inference tasks, the description language is based on a set of concepts and relationships between them. The lattes are called roles (denoted with $\mathcal{R}$). Axioms are used to capture the conditions that need to be met by consistent and coherent interpretations (the states of the domain). Therefore, the knowledge management tasks within an ontology could be formalized as:

- ***Deduction:*** is a task where the interpretation is an instantiation of formal knowledge consistent with evidence about the real-world domain [Hartz & Neumann 2007, Hudelot et al. 2008, Dasiopoulou, Kompatsiaris, & Strintzis 2009].
- ***Abduction:*** is a task where the interpretation is an instantiation of formal knowledge which allows to deduce the evidence [Shanahan 2005, Peraldi et al. 2007, Atif et al. 2014].

The deduction and the abduction reasoning tasks were introduced as inference standards. For the deduction reasoning, if $\sum$ is a logical theory and α a set of facts, through deduction is verified whether φ is logically entailed, that is whether $\sum, \alpha \models \varphi$. For the abduction reasoning, given $\sum$ and φ, the abduction consists in looking for an *explanation* α so that the entailment $\sum, \alpha \models \varphi$ is true.

In [Neumann & Möller 2008, Möller & Neumann 2008, Dasiopoulou & Kompatsiaris 2010, Bannour & Hudelot 2014], a comprehensive survey on the use of formal (description language based) ontologies for multimedia semantic analysis is exposed. In the following section, we enumerate some of these works.

3.2.3 Related Work on Knowledge-Based Approaches for Multimedia Analysis

In the last fifteen years, ontologies have been emerged from an interesting conceptualization paradigm to a very promising modeling technology for multimedia retrieval. Ontologies

enable meaning driven retrieval process through a machine-understandable form of a content description. In the following, we enumerate some multimedia ontologies outlining main characteristics.

The *Harmony* ontology [Hunter 2001], the *aceMedia* ontology [Petridis et al. 2004] and the *Rhizomik* [García & Celma 2005] ontology are first initiatives to attach formal semantic to MPEG-7. These ontologies have been explored to support semantic image /video analysis and annotation, addressing many content domains, including pancreatic cell images (for *Hamony*), soccer video (for *Rhizomik*), ... More recent MPEG-7 based ontologies, like *SmartWeb* [Oberle et al. 2007], *DS-MIRF* [Tsinaraki et al. 2007], *COMM* [Arndt et al. 2007] and *Boemie* [Dasiopoulou, Tzouvaras, et al. 2009], focused on a fully translation/mapping of the complete MPEG-7 specification into *OWL*.

These ontologies were mostly used in analyzing and annotating sport oriented multimedia contents. In their majority, enumerated ontologies have been constructed manually, and their knowledge structures are rather focused on low-level and high-level features (as classes), and their spatial-temporal relationships for particular content domains. Thus, these approaches are limited to provide a formalism that allows to use ontologies as repositories for storing knowledge. So, there is no correspondence between the expressive power provided by the adopted representation language, and the constructed ontology definitions. Hence, the ontologies were not fully exploited for multimedia retrieval. The key issue in enhancing multimedia retrieval is to focus to inherent semantics in addition to extracted ones through exploring reasoning and deduction capabilities of ontologies. Indeed, and contrary to the aforementioned ontologies, recent ontologies are addressing issues related to actual research works handling semantic contexts and concept co-occurrence.

Indeed, in [Mylonas et al. 2009], the authors proposed an ontology based approach to improve concept detection through visual thesaurus and semantic contexts. The authors introduce thus the use of semantic contexts in order to refine the confidence values for regions before taking decision. The latter deals with a proposed ontology that specifies fuzzy semantic relations among concepts/contexts *Location*, *Property*, *Part*, *Similar*, ...

In [Simou et al. 2008], the authors proposed a knowledge based framework for enhancing an initial set of over-segmented regions. Spatial relations and neighborhood information are managed within an ontology. The latter is defined by $\mathcal{SHIN}$ description language formalism.

The authors proposed also a deduction engine called FIRE in order to extract additional implicit knowledge.

In [Hudelot et al. 2010], the authors proposed an ontology for spatial relations in order to facilitate image interpretation. Such relations were considered as crucial for the semantic concept detection task. The proposed ontology was defined by the $\mathcal{ALC}(\mathcal{D})$ description language formalism.

In [Bannour & Hudelot 2014], the authors proposed a framework for building and using structured knowledge models for visual content analysis and annotation. The authors proposed thus an ontology to build explicit and structured knowledge models dedicated to image annotation. The defined ontology used the $\mathcal{SROIQ}(\mathcal{D})$ description language formalism. The built ontology consists in conceptual, contextual and spatial knowledge about an image (including relationships like *isAnnotatedBy*, *hasAppearedLeftOf*, *hasAppearedCloseTo*, ...). The authors proposed also reasoning framework in order to check about the consistency of the annotation efficiency.

Another research works focused on improving the indexing efficiency through the use of the ontologies capabilities. Indeed, in [Leite & Ricarte 2008, Cheng & Xiong 2012, Mukesh et al. 2013], the authors proposed an ontology based framework in order to enhance a semantic interpretation. These works mainly focused on defining how to manage knowledge in an ontology (concept and relationships). Also, other retrieval aspects were dealt: in [Rodríguez-García et al. 2012], the authors detailed how to evolve the ontology content using the DBPEDIA as an external data source. Also, in [Mustafa et al. 2008], the authors refined the semantic level by analyzing contexts and concepts interrelationships in a particular context.

In [Reshma et al. 2014], an ontology was generated and used both: (1) in training phase to select images that should be used for optimizing classifiers, and (2) in testing phase for deducing new annotations through concept inter-relationships.

3.2.4 Discussion

Recent research works are focusing on using ontologies for multimedia retrieval in order to allow semantic interpretation and reasoning over extracted descriptions. However, much remains to be done in order to achieve less human aid ontology modeling approaches.

Firstly, almost all existing approaches for modeling ontologies still relying on manual knowledge population (knowledge defined by experts) and there is no explicit method proposed for an automated ontology population. Such manual approaches are always costly, not always relevant and incomplete.

Secondly, using various relationships between concept/context has led to diversify the semantic capabilities of an ontology (as proved by aforementioned approaches), but it reduces their capacities to cover more multimedia content domains. Within an ontology structures, many semantic relationships between concepts were defined: *is-a* and *has-part* in WordNet [Fellbaum 2010], *is-a* in LSCom [Fel 2006], *IsPartOf*, *Location*, *Property*, *Part*, *Similar*, ... in [Mylonas et al. 2008; 2009]. We suppose that the generic aspect of an ontology strongly depends on its ability to model any semantic relationship between concepts/contexts.

Thirdly, the scalability should be considered in modeling ontologies, particularly, for the ontology structure effectiveness and reasoning computational cost. While in [Simou et al. 2008, Hudelot et al. 2010, Paliouras et al. 2011a, Bannour & Hudelot 2014], the ontology structure is defined by a highly expressive language, a simpler structure and less expressive language is used in [Fel 2006, Vallet et al. 2007, Mylonas et al. 2008; 2009, Fellbaum 2010]. Expressive language level is tweaked by paying attention to the huge amount of computational tasks needed for video interpretation. In fact, a simple ontology structure could be a good accommodation between a video analysis speed and performance.

Finally, the context of a content could provide an important cue for enhancing a semantic interpretation [Fauzi & Belkhatir 2014]. Yet, the definition of a context remains unclear and there is no computational method to define it. Contexts are defined manually in all works listed above. We consider that a computational definition of a context is a crucial step for an automated ontology construction.

The ontology used in the indexing process should also address issues related to actual research works dealing with the indexing process. Indeed, the use of *context* and handling the uncertain aspect of a semantic interpretation of video contents are also substantial. In the next section, we discuss handling uncertain information and knowledge in multimedia content analysis.

3.3 Uncertain Knowledge Management

In literature, there is several valuations about the selection between a fuzzy or probabilistic approaches regarding handling uncertain knowledge. Furthermore, Knowledge managing and ontology engineering attracted many research communities, but in multimedia retrieval one, some specific considerations have to be taken into account.

3.3.1 How to manage the uncertainty

Knowledge discovery is related to the analysis of large contents in the purpose of extracting valuable, meaningful, unknown and unexpected relationships. Real world data is characterized by the vagueness and the uncertainty of its content.

In multimedia retrieval, discovering knowledge from annotated multimedia contents are considered as an important task in order to handle semantics efficiently. Yet, there are many cases in which annotated contents display uncertain situations. In literature, two approaches attracted the attention of many researchers in order to handle uncertain knowledge: the probability theory and the fuzzy logic.

As discussed above, these annotated video datasets comprise naturally many uncertain situations. So, we adopted a fuzzy logic based approach [Zadeh 1979; 2008, Singh et al. 2013] to handle such situations because we believe that such approach fits better than probabilistic ones for managing uncertain knowledge, and it was widely adopted by many works that deal with uncertainty in multimedia retrieval [Bloch 2005, Hudelot et al. 2008, Simou et al. 2008, Dasiopoulou, Tzouvaras, et al. 2009, Bannour & Hudelot 2014].

Indeed, the probabilistic approaches are based on estimating probability that a data belongs to a class. On the other hand, fuzzy approaches attribute for a given data different degrees of membership to classes. Thus, the fuzzy logic handles a type deterministic uncertainty describing the data class ambiguity. And unlike the probabilistic approaches which answer to the question if a data belongs or not to a class, the fuzzy ones compute the degrees to which a data belongs to a set of classes. Then, the fuzzy logic considers that a data could belong to a set of classes at the same time with different membership degrees.

When analyzing a multimedia content, we intend generally to annotate that content by a set of semantic concepts. For each one, a degree of membership is computed in order to

describe if that semantic concept figures or not in the content. This degree is ranging from 0 to 1, and there are some situations where it is hard to say if the concept really exists (1) or not (0). Furthermore, this degree is generally confounded with the probability (between 0 and 1) that a semantic concept exists in a content. So, even the probability and the fuzzy membership are ranging from 0 to 1, they are fundamentally different.

To conclude, and in literature, there are many discussions about these two approaches and their efficient capabilities to handle and support uncertainty. In fact, many arguments are defended by the knowledge extraction community about the effectiveness of fuzzy approaches and probabilistic ones [Gaines 1978, Bosko 1990, Sanjaa & Tsoozol 2007, Zadeh 2014; 2015]. In our work, we focused on the use of a fuzzy approach not for proving that such an approach is better than the probabilistic ones, but because we believe that fuzziness is more suitable for our case.

3.3.2 Fuzzy DLs

Despite the expressiveness of DLs, they lack the ability to handle vague and uncertain semantics which is a real requirement in multimedia content indexing [Simou et al. 2008]. In fact, as stressed in this dissertation, multimedia retrieval faces the problem of handling uncertain information about multimedia contents.

In order to cope with the uncertainty issue in multimedia indexing, *Fuzzy-DL* [Straccia 1998; 2006, Stoilos et al. 2007, Ma et al. 2013] is considered as an interesting formalism for representing a multimedia ontology. In fact, the fuzzy Description Logics is considered as a very interesting logical formalism as it can be used in numerous domains like multimedia and information retrieval [Meghini et al. 2001] to provide ranking degrees and to cope with vague concepts like *"near"*, *"far"* and many more.

Fuzzy logic deals with vagueness and imprecision using fuzzy predicates. Therefore, the fuzzy logic offers a considerable foundation for description language generalization in order to deal with such vague concepts. Thus, fuzzy DL allows expression of the form ($\langle C(a) \geq n \rangle$ where $n \in [0..1]$, ie ($\langle Far(distance) \geq 0.7 \rangle$, with intended meaning "the membership degree of individual a being an instance of concept C is at least n".

In the two last decades, many fuzzy description logic formalisms were discussed and proposed. The fuzzy aspect integration into DL does not concern only adding role hierarchies

and number restrictions, but also the decidability proof and decision procedures for the knowledge base satisfiability and consistency.

For instance, in [Straccia 1998], the authors proposed the fuzzy integration within the DL $\mathcal{ALC}$ formalism with taking an attention on a nice trade-off between computational complexity and expressive power of DLs.

In [Stoilos et al. 2005b; 2007], the fuzzy $\mathcal{ALC}$ is extended to the fuzzy $\mathcal{SHIN}$ with transitive role axioms ($\mathcal{S}$), inverse roles ($\mathcal{I}$), role hierarchies ($\mathcal{H}$) and number restrictions ($\mathcal{N}$). The main contributions in such a work are a detailed reasoning algorithms, and the decidability proof of the fuzzy DL fuzzy-$\mathcal{SHIN}$. In [Simou & Kollias 2007, Simou et al. 2008], fuzzy-$\mathcal{SHIN}$ was used as a formalism to construct and manage an ontology for a multimedia content analysis process.

In [Straccia 2006], the authors proposed a fuzzy extension of the OWL Description language formalism $\mathcal{SHOIN}$. The latter was also extended to in [Stoilos et al. 2006] to fuzzy $\mathcal{SHOIQ}$ investigating several properties of the semantics of transitivity, qualified cardinality restrictions and reasoning capabilities.

In [Bannour & Hudelot 2014], fuzzy-$\mathcal{SROIQ}$ was proposed in the aim to provide both a set of constructors allowing the construction of new concepts and roles. This formalism includes $\mathcal{ALC}$ standard constructors (i.e. negation, conjunction , disjunction, full existential quantification, and value restriction) extended with transitive roles ($\mathcal{S}$), complex role axioms ($\mathcal{R}$), nominals ($\mathcal{O}$), inverse roles ($\mathcal{I}$), and qualified number restrictions ($\mathcal{Q}$). The proposed fuzzy DL formalism is used to manage an ontology that supports an image annotation framework.

To sum up, many extensions are raising more and more on the crisp DL in order to enable handling uncertain and vague knowledge through the use of fuzzy logic theory. Thus, a great variety of fuzzy DLs can be found in the literature [Garcı et al. 2010, Cerami & Straccia 2013]. Nevertheless, it has been shown that several fuzzy Dl formalisms face the undecidable reasoning issue [Baader & Peñaloza 2011]. In fact, the fuzzy extensions of DL do not have the finite model property, then the proposed reasoning algorithms are neither correct nor complete: the knowledge base satisfiability is then an undecidable problem [Cerami & Straccia 2013]. Despite many efforts to proof the decidability of fuzzy DL, multimedia spatial-temporal reasoning based on description logic is well known as undecidable. This is due to

specific nature of spatial temporal information to be handled. In fact, this information is cyclic and transitive, and within this particular situation, the fuzzy DL formalisms are undecidable.

3.3.3 Discussion

As discussed in chapter 1, we intend to go deeper in the exploitation of ontology semantic capabilities in order to improve the video analysis process, and particularly, the indexing one. Our objective is twofold. Firstly, we aim to define an approach to extract, by an automatic manner, valuable knowledge from a video annotated dataset. Secondly, we show how we construct an ontology that includes conceptual and contextual knowledge and populated by the extracted knowledge.

We focus on solving the main issues related to managing knowledge in multimedia retrieval: uncertainty and the large amount of data to be treated. As aforementioned, a fuzzy approach is adopted to handle uncertainty. On the other hand, we intend to propose an approach that could enable a machine to discover new knowledge from large-scale multimedia datasets, and to reason about a multimedia semantic interpretation in order to enhance it. However, analyzing large-scale video datasets to explore and reason with their contents is a critical task particularly when using *Description Logics* (DL) as a formal description of ontology content and reasoning [Stoilos et al. 2005a, Dasiopoulou & Kompatsiaris 2010, Bannour & Hudelot 2014]. Then, in our dissertation, we aim to define an alleviated ontology structure using *fuzzy-DL*, and a tweaked reasoning engine in order to enhance video content descriptions with an acceptable computing cost.

3.4 Multimedia Retrieval Scalability

3.4.1 The scalability issue

Nowadays, multimedia analysis tools and applications require a real-time processing, in particular when embedded within interactive environments such as smart-phones and home entertainment systems. A such real time processing needs fast and low-latency approaches for multimedia analysis. With such an intensive computing demands, the multimedia retrieval

is faced with the problem: how do the multimedia analysis efficiently process the growing amount of data and how to define scalable approaches that could meet user requirement?

In order to enable multimedia retrieval scalability to the huge amount of multimedia data, the literature proposed three different research directions: The first direction deals with the use of cloud and distributed programming frameworks such as *Multi-Threaded Processing* [Amit et al. 2006], *Hadoop* [White 2012, Landset et al. 2015], *MapReduce* [Dean & Ghemawat 2008; 2010] and *Apache Spark* [Meng et al. 2016, Zaharia et al. 2016]. The second direction deals with deep neural networks based methods and techniques [Long et al. 2016] in order to address the scalability. The third direction deals with semantic hierarchies [Deng et al. 2010, Zhou & Fan 2014, Ordonez et al. 2015] in order to reduce the complexity of managing large scale data through the use of some heuristics.

In what follows, we enumerate some research works that aimed to solve the scalability issue through these three different directions.

3.4.2 Cloud/distributed-Based Approaches

Cloud and distributed based approaches provide powerful technology and techniques to perform massive-scale and complex computing [Furht 2010, J. Zhu 2010]. Due to the time demanding multimedia analysis task that requires a large computational framework, the multimedia community addressed cloud and distributed approaches [W. Zhu et al. 2011].

In [Mohamed & Marchand-Maillet 2012], the authors proposed an adapted structure of *MapReduce* programming model for a scalable multimedia indexing. With an experiment contacted on *XML* text and images from ImageNet, they achieved a good speed compared to a sequential implementation.

In [Guðmundsson et al. 2012], the authors described a study where the *Hadoop* parallel and distributed run-time environment is used to speed up the construction of a large high-dimensional index for multimedia contents. They achieved then a speed-up of about 400% compared to a classical index construction.

In [Mourão & Magalhães 2015], the authors proposed a *Hadoop* distributed search engine framework. They considered that such a framework is flexible enough to support and handle several millions of multimedia contents.

In [W.-N. Chen & Hang 2008, Luo & Duraiswami 2008, Lopresti et al. 2012, Oh et al. 2015, Osipyan et al. 2015], the authors exposed different research works on the use of *GPU* (Graphical Processing Unit) in order to parallel the heavy computing process for different multimedia analysis tasks: face detection, features extraction and classification, ...

In [P. Kumar 2015], the authors focused on modeling a high performance video data processing technique. The latter uses a GPU based parallel implementation of object detection algorithm.

3.4.3 Deep Learning-Based Approaches

Deep learning [G. E. Hinton & Salakhutdinov 2006, G. E. Hinton et al. 2006, Bengio 2009] can be defined as a set of machine learning algorithms that can learn the data representation and feature extraction with many layers of non-linear transformations. A typical deep learning architecture consists of artificial neural network with many layers of non-linear processing units [Haykin & Network 2004, Jaeger 2016].

Actually, we observe that Deep learning is a thriving field with many practical applications and research topics, including semantic object detection [G. E. Hinton et al. 2006, Bengio et al. 2007, Ciregan et al. 2012, Krizhevsky et al. 2012] and information retrieval [Salakhutdinov & Hinton 2009, G. Hinton & Salakhutdinov 2011].

In conjunction with significant gain in machine performances (particularly with GPUs based acceleration), the massive amount of labeled data available for supervised training allowed the deep neural networks to significantly improve the machine learning abilities for a variety of applications [Long et al. 2016]. The amount of multimedia data has grown to an extend that classical multimedia processing and analysis techniques are unable handle data effectively [Jiang 2015]. In fact, some deep neural networks based methods and techniques were proposed in order to address the scalability issue and to handle large amount of multimedia data. [Q. Wu et al. 2015, Druzhkov & Kustikova 2016] display a comprehensive survey on deep learning methods and software tools for video segmentation, image classification, object detection, audio analysis, ...

In recent literature, a growing number of research work are being the effectiveness of deep neural networks based methods in handling the scalability issue, in particular *Deep Convolutional Neural Networks* (CNN).

Indeed, in [Jiang 2015] used CNN for large scale image feature extraction and classification. In [Sainath et al. 2015], CNN are used for large scale speech analysis. In [Girshick et al. 2014], CNN are also efficiently used for image region detection, feature extraction, object detection and semantic segmentation. In [Tong et al. 2015], a CNN based framework for large scale video shot boundary detection and semantic annotation is detailed.

3.4.4 Semantic hierarchies-Based Approaches

Despite the significant progress shown by the above enumerated works to achieve a scalable multimedia analysis frameworks and approaches, and in particular deep neural networks based ones, some other works attempt to use the reasoning power of the ontologies for the semantic multimedia content interpretation and analysis. In fact, a formal model of a given knowledge can be used in order to help and guide the semantic multimedia analysis, and then to alleviate its computational cost.

Semantic hierarchies [Simou et al. 2005, Fan et al. 2008, Deng et al. 2009] were proposed to construct semantic relationships between semantic concepts. Such a hierarchy could be used then to alleviate the multimedia analysis process, and then to be able to handle large-scale multimedia contents.

In [Dasiopoulou et al. 2008, Dasiopoulou, Kompatsiaris, & Strintzis 2009], the authors proposed an approach for reasoning on the output of a statistical classifiers. In fact, the extracted descriptions about a content are used within a reasoning process in order to look for extra semantic concepts that were not detected by the statistical classifiers. Likewise, in [Hudelot et al. 2010, Bannour & Hudelot 2014], the authors proposed to use a built knowledge model in a framework for reasoning over the outputs of machine learning algorithms.

3.4.5 Discussion

To sum up, we think that the literature exposed many research works to enable the multimedia indexing capabilities for handling large-scale data. In this dissertation, we are more oriented towards semantic hierarchies based approaches rather than distributed and parallel ones. In fact, we think that giving a built knowledge database about semantic concepts/contexts relationships, the semantic hierarchy could be used to enhance the semantic analysis efficiency.

Furthermore, we observe that semantic hierarchy based approaches focused on late guiding the semantic concept detection. In fact, these approaches handle the output of semantic concept detectors, and not the detectors themselves. Thus, we think that the valuable knowledge stored within an ontology could contribute at the enhancement of multimedia analysis efficiency through guiding the construction of the concept detectors.

The chapter 6 exposes our contribution for a scalable multimedia content indexing through a framework that constructs semantic concept detectors based on knowledge reasoning.

3.5 Evaluation of Literature Review

Actually, there is yet standard approaches which can be used for indexing a video content. The latter can be indexed based on either the low-level (perceptual) features or high-level (semantic) annotation. As already discussed in the previous chapters, such approaches fail to alleviate the semantic gap issue. This dissertation aims to show the benefits of integrating fuzzy knowledge based approaches in the video indexing process. Nevertheless, such new approaches must be standardized, generic and robust enough to be applied for different user and application requirements.

By drawing on the semantic assets provided by the two proposals (namely the concepts co-existence and contexts), the multimedia community investigated the knowledge engineering for multimedia retrieval. As a knowledge database, the ontologies are considered as powerful tool to design and handle concepts/contexts and their interrelationships. Ontology-based multimedia retrieval approaches focus on defining a knowledge conceptualization and a reasoning process in order to analyze and improve semantic interpretation of a multimedia content. Ontology-based approaches for semantic multimedia retrieval displayed promising results [Kannan et al. 2012]. However, new issues appeared and further researches on ontology engineering for multimedia analysis have to be more addressed.

Based on the literature review, the tables 3.1, 3.2 and 3.3 expose some main characteristics of the three video indexing approaches.

In this dissertation, we are focusing on the semantic video indexing through exploring ontologies structures and semantic capabilities. The latter are used to improve the multimedia

Table 3.1: Literature Review on Knowledge-based Multimedia Analysis

Approaches	Advantages	Limits
Earlier approaches: Mapping low-level features to an ontology	Provided a rich set of standardized tools that generate and understand multimedia content.	Focused only on mapping low-level features to an ontology without availing its reasoning capabilities.
Lite ontologies for semantic analysis	- Model knowledge as classification, categorization and taxonomies for many application domains. - Model semantic concept inter-relationships and semantic contexts in order to provide a more semantic level for the interpretation of a multimedia content.	Handle explicit knowledge without exploring implicit one: the reasoning capabilities of an ontology are not availed.
Formal ontologies for semantic analysis	- Provide a powerful expressiveness and formal knowledge representation. - Supply many reasoning capabilities to infer implicit knowledge. - Contribute to enhance multimedia indexing through inferring inherent semantics in addition to extracted ones through exploring reasoning and deduction capabilities of ontologies.	- The proposed approaches rely on manual knowledge population: a costly task and not always relevant and complete. Semantic concepts inter-relationships and contexts are defined manually. - Many semantic concepts/contexts inter-relationships are defined depending on the application domain to handle. So, no generic knowledge structure was proposed. - No real discussion about the scalability of the proposed formalisms and their capabilities to deal with real large-scale data. - No generic multimedia analysis oriented framework proposed to manage the complete knowledge work-flow within an ontology: knowledge abduction, population, reasoning and evolving.

Table 3.2: Literature Review on Fuzzy-DL Formalisms to Handle Uncertain Knowledge

Approaches	Advantages	Limits
Fuzzy-DL formalisms for Multimedia ontologies	- Fuzzy logic deals with vagueness and imprecision using fuzzy predicates. Therefore, the fuzzy logic offers a considerable foundation for description language generalization in order to deal with such vague concepts. - A variety of fuzzy-DL formalisms is proposed, and many reasoning engines were developed. - Widely emerged in recent knowledge-based multimedia semantic analysis. Thus Many fuzzy-DL formalisms are being used to construct and manage an ontology for a multimedia content analysis process.	- Fuzzy-DL faces the undecidable reasoning issue. In fact, semantic concepts/contexts inter-relationships are often cyclic, transitive which make the reasoning task undecidable. - No real discussion about the scalability of the proposed formalisms and their capabilities to deal with real large-scale data.

Table 3.3: Literature Review on Scalable Multimedia Indexing approaches

Approaches	advantages	Limits
Parallel based approaches	- A significant and spectacular speed-up for multimedia analysis frameworks and approaches through parallelizing the computing task. - *MaprReduce*, *Hadoop* and in particular CUDA are getting more and more engendered to accelerate heavy computing works for many research fields.	The parallelization is defined manually by the developers and generally considered as a complex task.
Knowledge based approaches	- The reasoning power of the ontologies are potential formalism that could help and guide the semantic multimedia analysis, and then alleviate its computational cost. - Semantic concepts/contexts inter-relationships could be potential information that could generate some heuristics for the detection of large set of semantic concepts within a multimedia content.	The semantic concepts hierarchies are used to enhance the outputs of semantic concept detectors. But few works are interested in integrating of such hierarchies within the construction process of these detectors.

indexing process, and to build a scalable indexing tool. Specifically, our concerns in this dissertation are as follows:

Firstly, we look for a generic knowledge-based framework that handle multi-modal interpretations in order to deduce a more complete semantic representation for a multimedia content through ontologies capabilities (basically abduction and deduction).

Secondly, we focus more on the ontology management side, and we define a complete automatic knowledge manager: from extracting valuable knowledge, defining the ontology structure by taking into consideration semantic context and concept co-occurrence, and the deduction process that generates an enhanced interpretation over ones delivered by classical semantic concept detectors, to the knowledge evolving. The proposed ontology profits from the expressiveness and reasoning power of fuzzy-DL formalisms with paying attention to the scalability issue.

Thirdly, we explore more the ontology while focusing on exploring their content to build scalable semantic concept detectors. Subsequently, we aim to use knowledge model in order to produce a semantically consistent and scalable multimedia indexing process.

3.6 Conclusion

In the present chapter, we exposed a comprehensive overview on the recent work for multimedia retrieval, and particularly the multimedia indexing. Our aim was not to provide a complete survey of the state of the art approaches, but to show the importance of knowledge based approaches for multimedia indexing problems and to shed light on the benefits of each of them. In fact, exploring ontologies seems to be essential to improve the multimedia indexing, and then to narrow the semantic gap.

The next chapter will present our first contribution that is revealed by the defintion of a generic framework for a multimedia semantic indexing.

Part II

Fuzzy Ontology Based Framework for Video Indexing

Chapter 4

Multimodal Fuzzy Fusion Framework for Semantic Video Indexing Improvement

In this chapter, we detail and discuss the first contribution C_1. In fact, we propose a general multimodal fuzzy approach to enhance video semantic indexing accuracy. The latter approach was developed incrementally through some consecutive research works. For that, and in the section 4.2, we first introduce our preliminary reflections for a knowledge based framework to index video content efficiently. Then, in section 4.3, and based on what was obtained through an experimental study, we present a framework improvement, in particular, handling fuzzy knowledge and corresponding reasoning ability. Finally, in section 4.4, we present our proposed video annotation tool in order to generate valuable data source about video content. First of all, section 4.1 displays first afterthought for a knowledge based video indexing framework.

4.1 Context and Motivation

Multi-modal fusion has gained much attention by the multimedia community [Atrey et al. 2010]. In fact, a video indexing task involves processing of multi-modal data in order to obtain valuable insights about the content. Such a task could use sensory data (such as audio and visual analysis) as well as non-sensory data (such as meta-data). Thus, the fusion of multiple

modalities can provide complementary information and increase, then, the accuracy of the indexing process.

In [Lucien 1999], the data fusion is defined as follows:

Definition 1. *"a formal framework in which are expressed means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality".*

A discussed in [Waltz et al. 1990, Esteban et al. 2005, Blasch et al. 2006, Guerrero et al. 2009], a fusion process consists of five levels: The *level 0* (named *pre-processing*) and *level 1* (named *object refinement*) cover respectively signal processing and data alignment. The *level 2* (named *situation refinement*) attempts to construct a complete picture from incomplete information provided by the lower levels (*level 0* and *level 1*). The *level 3* (named *threat refinement*) interprets the results from *level 2* in terms of the possible opportunities for operation. A *process refinement*, referred to *level 4*, loops around these earlier levels to monitor performance and optimize the fusion process. Earlier data fusion applications were initially used for military reasons. But currently, data fusion model is extended for several other areas [Liggins II et al. 2008].

By drawing inspiration from this fusion model, we propose a fusion framework to fuse multi-modal video content extracted semantics (concepts). Thus, we adopt the JDL/DFS data fusion model [Waltz et al. 1990].

After an experimental study carried out within TRECVID2010 evaluation campaign, we concluded that the multi-modal fusion process is important, but a big challenge occurred on how to efficiently manage and explore fuzzy knowledge for better semantic indexing enhancement. From this observation, we decided to tackle in depth the knowledge management raised issue. Thus, we proposed a preliminary fuzzy ontology based framework to explore knowledge reasoning capabilities in video indexing. A light ontology structure was proposed, as well as an abduction engine to extract fuzzy knowledge from valuable data sources, and a deduction engine able to leverage from the extracted knowledge in order to generate newer knowledge filling out a semantic interpretation about a video content.

Yet again, the experimental study showed that the use of contextual ontologies has improved the indexing process performances, however good, a simple set of manually defined contexts were used. Aiming to go further in the use of contextual information, we proposed

afterwards a fuzzy video collaborative annotation tool that helps to detect potential semantic concept inter-relationships. The latter are important to define semantic contexts.

In the following section, we display in details main works that have been conducted in order to realize our first contribution C_1.

4.2 A Multimodal Fuzzy Fusion Framework

In this section, we will introduce our fusion system and describe the multi-modal fuzzy fusion process.

The proposed fusion architecture, based on the JDL/DFS model, is shown in figure 4.1. After extracting unimodal semantic interpretation (level 0), the level 1 is called to search and eliminate conflicting situations. Then, level 2 is applied on level 1 results aiming at finding further concepts based on abduction and deduction engines. These engines analyses relationships between concepts. The level 4 is used to control the whole fusion process. Below is the description of the different integrated levels and the proposed fusion architecture.

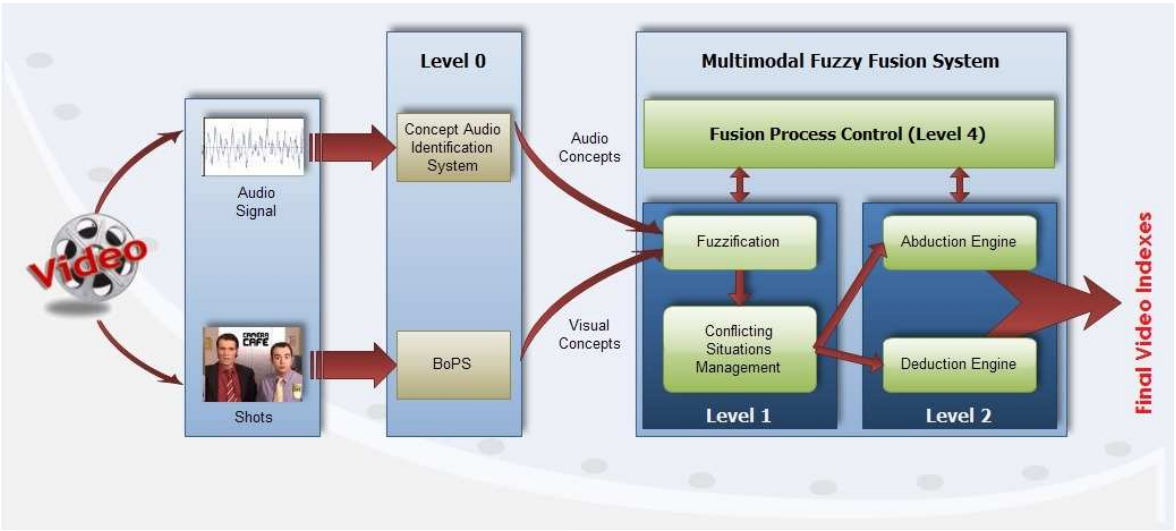


Figure 4.1: Overview of the proposed Multimodal Fuzzy Fusion System

4.2.1 Object refinement (level 1)

The *level 1* deals with mixed unimodal semantic interpretations. The latters are structured through this data format: every concept has a list of indexed video content sorted by their descending pertinent ranks. These ranks are fuzzified then analyzed.

Concept ranks fuzzification The purpose of this step is to calculate the fuzzy membership degree of a concept to a video content. This is given by a normalization function (see equation 4.1).

Let r be the rank of a concept for a video content, and R is the highest rank of the same concept for all video contents. We seek for a transformed rank called r_N as follows:

$$r_N = \left(\frac{(\epsilon - 1)}{(R - 1)} * (R - r)\right) + 1 \tag{4.1}$$

Where ϵ is a positive integer.

Conflicting situations handling Sometimes, certain conflicting situations can be found in aggregated semantic interpretations. Since we are using fuzzy logic, two contradictory semantic interpretations can coexist for the same video content. This situation is permitted until the equation 4.2 is verified.

Let c_1 and c_2 be two contradictory concepts. And Let $\mu(c_1)$ and $\mu(c_2)$ be respectively the relevance degree of the first and the second concepts.

$$\Big((1 - \mu(c_1)) - \epsilon\Big) \leq \mu(c_2) \leq \Big((1 - \mu(c_1)) + \epsilon\Big) \tag{4.2}$$

Where ϵ is a positive integer.

If the equation is not verified, the concept, extracted from the less confident modality, is deleted. This trust degree is established for each modality by the fusion control process (*level 4*).

In other situations, we find that the concept relevance analysis in a video content varies from one modality to another. This conflict over the relevance degree for the same concept is solved using the equation 4.3. (Same equation used in [Vrochidis et al. 2010]).

Let c a concept. And let $\mu_1(c)$ and $\mu_2(c)$ the relevance degrees computed repectively from the first and the second modalities.

$$\mu(c) = \alpha\mu_1(c) + \beta\mu_2(c) \tag{4.3}$$

Where α and β are trust degrees respectively for the first and second modalities fixed by the fusion process control (*level 4*), and $\alpha + \beta = 1$.

We can recognize that level 1 has a great importance since it eliminates any irrelevant information. However, several actual indexing systems do not account well for this component (as in [Vrochidis et al. 2010], [C. G. M. Snoek et al. 2006] and [Ayache et al. 2007]).

The set of fuzzyfied and filtered concepts are finally passed to *level 2*.

4.2.2 Situation refinement (*level 2*)

The purpose of this level is to look for new concepts by analyzing available interpretations. To do this, we propose the use of two different intelligent techniques:

Deduction engine

Using a Mamdani fuzzy system [Mamdani & Assilian 1975], a deduction engine infers new concepts using fuzzy rules extracted from the LSCOM ontology [Kennedy & Hauptmann 2006]. This ontology is based on generalization relationships. Figure 4.2 illustrates how we extract automatically fuzzy rules from the LSCOM ontology.

Since the concept relevance degrees are already fuzzifyed, we don't need to use the fuzzification (first component of the Mamdani fuzzy system). The defuzzification component is also not used.

Our deduction engine is similar to the network of operators described in [Ayache et al. 2007]. The main difference between the two systems is that our system extracts the rules automatically from the analysis of an ontology, but the network of operators is set manually. The second difference is that our system deals with fuzzy interpretations.

Figure 4.3 shows an overview of the deduction engine using extracted fuzzy rules.

Abduction engine

The abduction engine is based on a learning system that allows analysis of a learning video database in order to find possible relationships between concepts as input and expected concepts as output. These relationships are then considered as fuzzy rules used for the deduction of new concepts based on extracted ones from a test video database. Figure 4.7 illustrates an overview of the abduction engine.

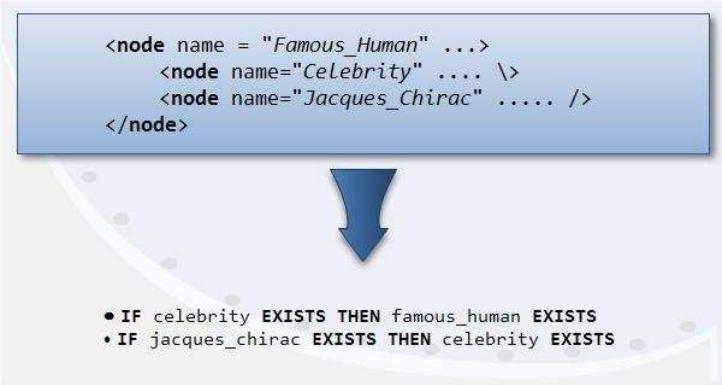


Figure 4.2: Extracting Fuzzy Rules from LSCOM Ontology

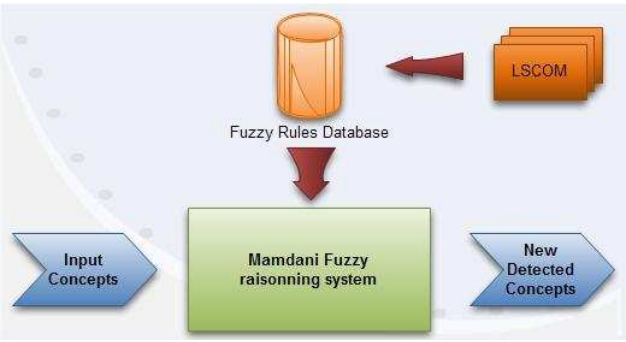


Figure 4.3: Deduction engine for situation refinement

This engine uses the βeta fuzzy systems (BFS) based on a multi-agent genetic algorithm [Kallel & Alimi 2006]. Thus, we use the equation 4.4 for minimizing objective function obj_fun and then, optimizing the learning process.

Let C be a set of n concepts: $C = \{c_1, c_2, \ldots, c_n\}$. And let $\mu'(c_i)$ and $\mu''(c_i)$ be two relevant degrees of the concept c_i to a video content respectively for generated concepts as output and the expected ones. We compute the objective function obj_fun by this equation:

$$obj\_fun = \frac{\sum_{i=1}^{n} \left(\mu''(c_i) - \mu'(c_i)\right)}{\sum_{i=1}^{n} (\mu''(c_i))^2} \tag{4.4}$$

A similar abduction engine is described in [C. G. M. Snoek et al. 2006] using a supervised SVM classifier. This latter consists in analyzing a set of concepts to discover potential relationships.

We note that the output of *level 2* is the fusion of two outputs from the deduction and abduction engines. Eventual conflicting situation are treated in the same way as used in the

level 1. The degree of trust for each reasoning engine is fixed by the fusion control process (*level 4*).

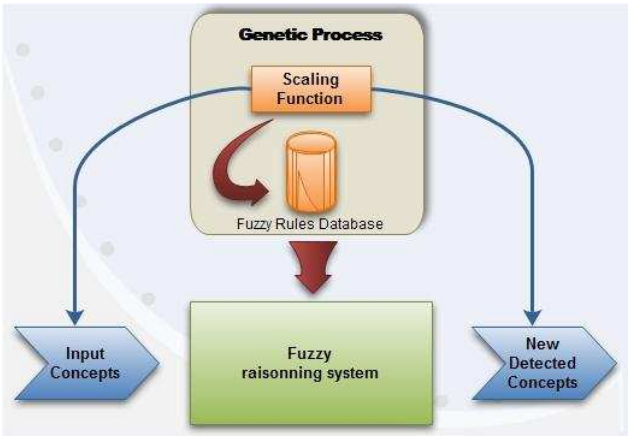


Figure 4.4: Abduction engine for situation refinement

4.2.3 Fusion Process Control (*level 4*)

This level aims at controlling the fusion process by manipulating trust degree of each modality (text, sound and images) and for each reasoning engine (abduction and deduction). These trust degrees are determined according to a supervised learning.

We can conclude that the proposed multimodal fuzzy fusion is characterized by its ability to deliver a rich and coherent semantic interpretation. This is mainly due to:

- the use of fuzzy logic and fuzzy reasoning,
- a maximum compatibility with the *JDL/DFS* data fusion model,
- operating a maximum information extracted from a video content (multimodality).

However, the quality of this fusion system is still dependent on the quality of initial semantic interpretations delivered by unimodal analyzers.

4.2.4 Experimental Study

In order to illustrate the semantic enhancement of concept detection introduced by our proposed fuzzy fusion and ontology-based framework, we have conducted two preliminary experiments within two multimedia evaluation campaigns. Hence, we expose the experimental

setup and the obtained results of our framework within TREC Video Retrieval Evaluation 2010 (TRECVID 2010) [Over et al. 2010] at the *Semantic Indexing* task.

Datasets Description

We used datasets provided in TRECVID 2010 benchmark. The TRECVID 2010 *Semantic Indexing* task provides two datasets proposed by the National Institute of Standards and Technology (NIST): a test and a development dataset. The development dataset (*IACC.1.tv10.training*) contains 3200 Internet Archive videos ($50GB$, $200h$), while the test dataset (*IACC.1.A*) contains approximately 8000 Internet Archive videos ($50GB$, $200h$). The development dataset is annotated by 130 semantic concepts.

Evaluation metrics

In order to be able to compare different indexing approaches, various system effectiveness metrics have been used. These metrics are commonly based on precision (which is defined as the number of relevant answers as a part of the total number of retrieved ones), and recall (which is defined as the number of relevant answers as part of total relevant ones in the collection). In our experiments, we consider the following evaluation measures: the *Inferred Average Precision* (*infAP*), the *precision* (P) and the recall (R) as a performance metric for the TRECVID 2010 *Semantic Indexing* task.

Experiments with TrecVid 2010 dataset

As an earlier experiment, we participated in TRECVID 2010 with two runs $Regim_4$ and $Regim_5$ [Elleuch, Zarka, et al. 2010]. The first run integrates only a visual semantic concept detector using key-points detection and visual features extraction tools [Elleuch, Ben Animar, & Alimi 2010]. And the second run aims at enhancing indexing effectiveness of the first one through a rule based deduction reasoning engine. Then, only the deduction process is used in the $Regim_5$ run. The abduction engine is ignored since a set of rules [1] were defined manually from the LSCOM ontology [Kennedy & Hauptmann 2006]. This rule list is based on two potential relationships between two concepts: *implies* (111 rules) and *excludes* (5 rules). For instance, the rule *"Sky **implies** Outdoor"* depicts that if the concept *Sky* exists in a shot,

[1] http://www-nlpir.nist.gov/projects/tv2010/tv10.semantic.indexing.relations2.txt

Table 4.1: Sample rules extracted from the LSCOM ontology

LSCOM Generalization Relationships		Extracted Rule
Advocate ***is a*** *Person*	$\Longrightarrow$	IsRelatedTo(Advocate,Person)=1
Airport_Terminal ***is a*** *Building*	$\Longrightarrow$	IsRelatedTo(Airport_Terminal,Building)=1
Backpack ***is a*** *Luggage*	$\Longrightarrow$	IsRelatedTo(Backpack,Luggage)=1

then the concept *Outdoor* exists too. Similarly, the rule *"Single_Person* ***excludes*** *Crowd"* depicts that if the concept *Single_Person* is detected in a shot, then the concept *Crowd* doesn't exist. In our case, we considered only ***implies*** relationships based rules.

The results of both $regim_4$ and $regim_5$ are shown in figures 4.5 and 4.6. The $regim_5$ run presents a concept detection enhancement over a classical visual concept detector for concepts 6, 15 and 126 (respectively the concepts: *Animal*, *Boat_Ship* and *Vehicle*). Further, the $regim_5$ run achieved an enhancement of 4.8% for the mean inferred average precision (***infAP***=0.089) among the $regim_4$ run (***infAP***=0.085). Thus, having regard to a weak set of rules used in the deduction engine, a knowledge based concept detection enhancement looks promising. Additionally, not all the concept set (130) were considered in the rule list. We would like also to point out that the TRECVID 2010 evaluation process gave concept detection performances for only 30 concepts (among 130). Thus, handling more concepts in the rule list should lead to a better semantic enhancement.

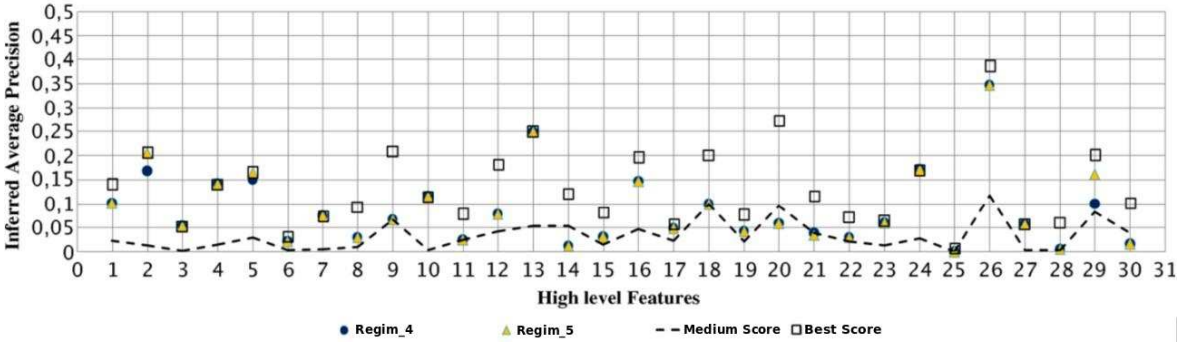


Figure 4.5: TRECVID 2010: $regim_4$ and $regim_5$ runs evaluations

With such results, we extended our experimentation on the same dataset by using more rules for the deduction engine. Indeed, the LSCOM ontology is built on a set of semantic concepts interrelated by the generalization relationship (*"is a"*). We transformed this relationship into rules. In total, 4383 rules were defined. The table 4.1 shows some of these generated rules.

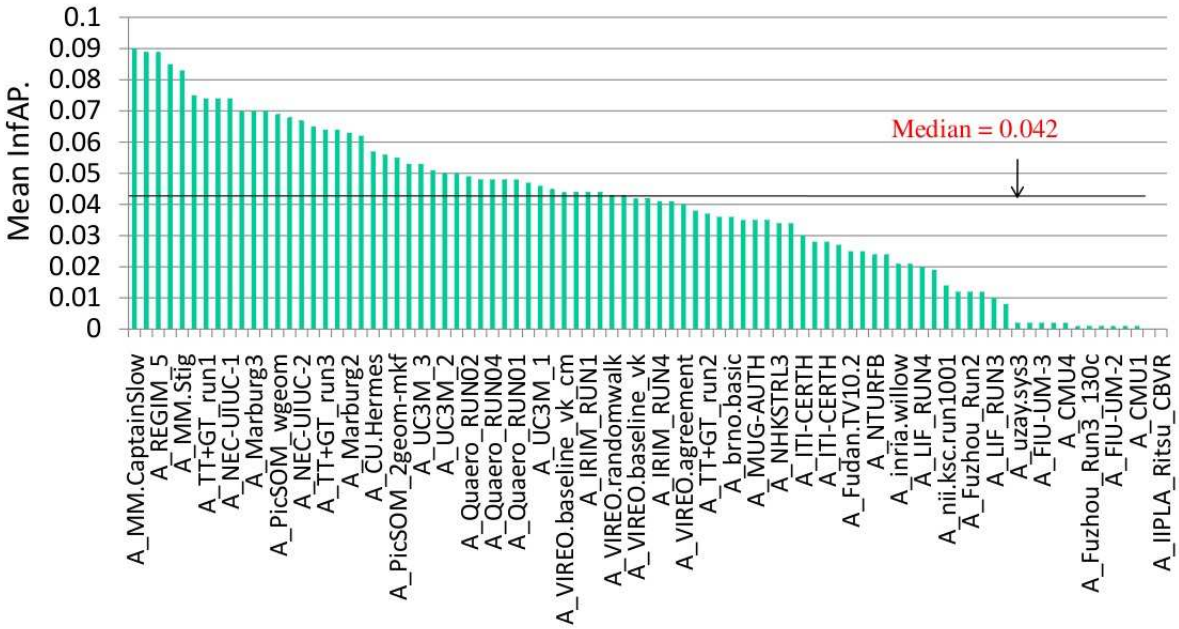


Figure 4.6: TrecVid 2010: $regim_5$ ranking in TrecVid 2010 Semantic Indexing Task (SIN)

Obtained results are displayed in table 4.2. Relying on these results, we can conclude that an indexing system effectiveness can be clearly improved through a knowledge-based approach. In fact, extracting and using rules from LSCOM ontology enable precision enhancement of about 18%. The recall is improved also of about 8%.

4.2.5 Discussion

As discussed in the previous chapters, semantic concepts detection is based on the use of supervised learning from manually annotated images samples. In fact, main approaches are almost exclusively focused on an independent development of concept detectors which focuses on the extraction of low-level visual features from positive and negative samples in order to model the high level concept. Nevertheless, one semantic concept may appear and exist in many different contexts, further, its appearance and meaning may be different according to these contexts. Thus, we think that concept based video indexing is not optimal. Effectively, the indexing process requires a very big amount of training examples in order to produce a generic indexing system, on the one hand, and slights the fact that concepts could exist together, on the other hand. As an example, the semantic concept *airplane_flying* coexists

Table 4.2: TrecVid 2010: Concept detection enhancement

Semantic Concepts	**Visual Concept Detector**			**LSCOM**		
	InfAP	***P***	***R***	***infAP***	***P***	***R***
Outdoor	-	0.52	0.59	-	**0.88**	**0.77**
Vegetation	0.1	0.74	0.68	0.1	0.74	0.68
Landscape	-	0.6	0.79	-	0.6	0.79
Sky	-	0.66	0.9	-	0.66	0.9
Trees	-	0.62	0.72	-	0.62	0.72
Mountain	-	0.68	0.8	-	0.68	0.8
Ground_Vehicle	0.043	0.3	0.66	**0.18**	**0.6**	**0.73**
Road	-	0.43	0.6	-	0.43	0.6
Car	0.075	0.42	0.64	**0.17**	**0.58**	**0.73**
Bus	-	0.52	0.73	-	0.52	0.73
Bicycles	0.142	0.67	0.92	**0.185**	**0.82**	**0.97**
Emergency Vehicle	-	0.9	0.83	-	0.9	0.83
Building	0.022	0.18	0.22	**0.1**	**0.5**	**0.43**
Truck	-	0.35	0.37	-	0.35	0.37
Airplane Flying	0.102	0.8	0.78	0.102	0.8	0.78
Airplane	-	0.5	0.6	-	0.6	0.6
Total	0.071	0.53	0.66	**0.134**	**0.64**	**0.71**

with the semantic concepts *sky* and *airplane*. Then, we can define the semantic concept *airplane_flying* as context which makes a relationship between the concepts *sky* and *airplane*.

The term *context* is ambiguous. It has been defined in several ways. For the multimedia community, a *context* is introduced as an extra information for both concept detection and scene classification [Mylonas & Avrithis 2005, Torralba et al. 2010].

To go further toward semantic enhancement, and considering both obtained results within the TrecVid 2010 and the research works focused on the semantic *context*, we attempted to introduce the *context* in our proposed framework. Thus, we model contextual information in order to better explore and understand a video content. In addition, we have opted for a standardized knowledge model for representing fuzzy relationships between semantic concepts. We have used then the ontologies to model the fuzzy knowledge used to improve video indexing.

As a next research step, we attempted to propose a context-based framework for video indexing.

4.3 Ontology based Framework for Video Content Indexing

The second revision of our fuzzy multi-modal framework leans on the introduction on the *context*, and the exploration of ontologies facilities to handle fuzzy relationships between concepts. And for the new defined framework, we focused on: 1) semantic knowledge representation and interpretation, and 2) the refinement process.

Fuzzy knowledge representation intends to build the contextual space that handles relationships among every context and its related semantic concepts. Such knowledge is extracted, modeled and populated through an abduction engine mated with an inference engine. The latter is provided by fuzzy description logics [Straccia 2006, De Mantaras et al. 2015]. The fuzzy knowledge being extracted and populated within an ontology, the refinement process is used to enhance the semantic interpretation of video indexing. Such a process is based on fuzzy rules used by a deduction engine in order to infer new interpretation for a given analyzed video content. In the following, we detail our context based framework.

4.3.1 Framework Overview

In this section, we present the proposed fuzzy ontology-based framework for reasoning in video indexing. The proposed approach involves two steps, namely semantic knowledge representation/interpretation and refinement process. Our contribution is focused on modeling and building the context space and its exploitation to enhance video indexing system.

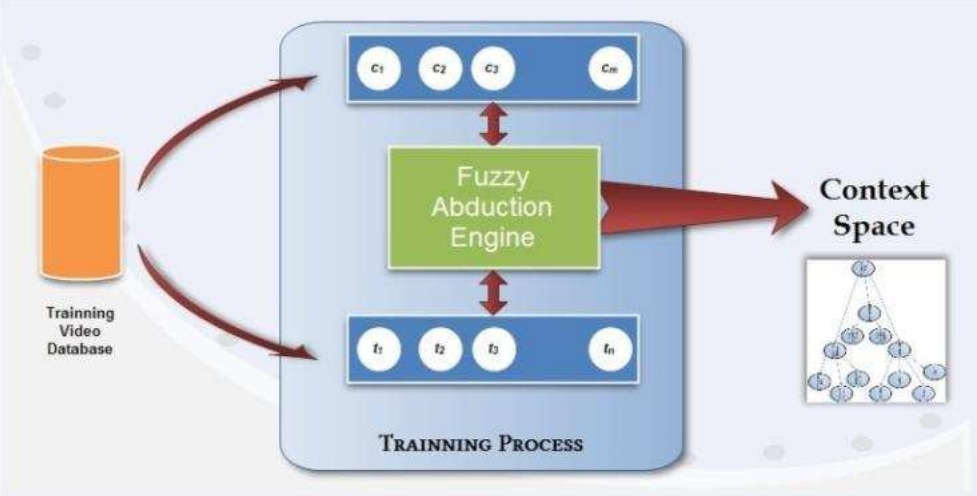


Figure 4.7: The Context Based Fuzzy Abduction Engine

In what follows, we display a description of both the semantic knowledge representation/interpretation, then the refinement process through the deduction engine.

4.3.2 Semantic Knowledge Representation/Interpretation

Based on a fuzzy Abduction Engine, the semantic knowledge representation and interpretation aims to analyze and model context spaces with fuzzy roles and rules (by the same way as the first proposed framework [Zarka et al. 2011]). These roles and rules are used then, through firing a deduction engine, in order to discover further concepts and therefore enrich semantic interpretation.

Therefore, a context-based ontology is, firstly, constructed by populating different relationships between each context and its semantic concepts, and secondly, providing a deductive engine based on fuzzy rules in order to infer newer knowledge about a video content.

Extracting the Contextual Space

The annotation process is used as semantic knowledge extraction tool to assist experts to identify and define all semantic concepts involved in a domain knowledge (context space). Generating such semantic knowledge has in recent years been approached by collaborative efforts within multimedia retrieval evaluation campaigns.

Thus, we propose an annotation approach to extract a semantic knowledge for each context space. The following considerations are taken into consideration for the proposed annotation approach:

Unified lexicon: We adopt a fixed lexicon for annotation of concepts and contexts in order to guarantee a convergence in user assigned free labels. The LSCOM ontology includes a unified set of concepts. We used then the LSCOM provided lexicons to define concepts (**eg.** *Sky, Airplane, Road*) and contexts (**eg.** *Office, Airplane_Flying, Urban*),

Soft annotation: Concept/context relationships can be considered as uncertain. Fuzzy relationships should be then used for the annotation approach. A membership relevance degree is then attributed to a semantic concept and a target context. We propose so three relevance levels, namely *"Relevant"*, *"Not-Relevant"* and *"Not-Exist"*. *"Relevant"*,

"Not-Relevant" respectively indicate that the concept is present and semantically strong (weak) in the target context and *"Not-Exist"* their lack.

Ontology Structure

A less expressive fuzzy description logic is used to describe the semantic knowledge. Such a choice is argued to facilitate fast computation.

In the following, we display how we model and populate a contextual knowledge for semantic interpretation.

Our fuzzy ontology is modeled as: $O^f = \{T, C, R^f_{tc}, R^f_{ct}, R^f_{cc}, Q\}$, where :

- $T = \{t_1, t_2, ..., t_n\}$ is a set of n contexts;
- $C = \{c_1, c_2, ..., c_m\}$ is a set of m concepts;
- $R^f_{t_i c_j} : TxC \rightarrow [0,1], i \in \{0, ..., n\}$ *and* $j \in \{0, ..., m\}$ is a fuzzy role between context t_i and concept c_j;
- $R^f_{c_i t_j} : CxT \rightarrow [0,1], i \in \{0, ..., m\}$ *and* $j \in \{0, ..., n\}$ is a fuzzy role between concept c_i and context t_j;
- $R^f_{c_i c_j} : CxC \rightarrow [0,1], \quad i, j \in \{0, ..., m\}$ is a fuzzy role between concept c_i and concept c_j ;
- Q is a set of fuzzy qualifier. In O^f, we define two qualifiers: *"weak"* and *"strong"*.

We define too sub-roles between contexts and concepts: {*Generalization, IsRelatedTo, IsPartOf, Includes*}. The interpretation of these roles is detailed in the Table 4.3.

Table 4.3: Semantic Relationships Between Concepts and Contexts

Name	***Symbol***	***Meaning***	***Type***	***Definition***
Generalization	$t_i : t_j$	The concept c_i is the generalization of the concept c_j	TxT	LSCOM
IsRelatedTo	$c_i \| tk \rightarrow c_j$	The concept c_i is related to the concept c_j within t_k	CxC	Learning
IsPartOf	$\{c_i\} \in t_j$	A set of concept c_i is a part of the context t_j	CxT	Learning
Includes	$t_i \supset c_j$	The context t_i includes the concept c_j	TxC	Expert

- ***Generalization Role***: The *"generalization"* role between t_i and t_j is defined if t_i is a sub-context of t_j, which is denoted as: $t_i : t_j$. As an example, *"Ground_Vehicle"*

and *"Vehicle"* are related as *"Generalization"* relationship. $Ground\_Vehicle : Vehicle$ indicates that all relevant video shots for sub-context *"Ground_Vehicle"* must also be relevant to context *"Vehicle"*. The generalization relationship is the most common relation used to build ontology hierarchy, which can be exploited to enhance concept detectors. The LSCOM ontology, dealing only with this relationship, provides a ready enumeration of generalizations between all defined concepts.

- ***IsRelatedTo Role*** The *"IsRelatedTo"* role between c_i and c_j is defined if c_i is related to c_j within t_k , which is denoted as: $c_i|tk \rightarrow c_j$. As an example, *"Snow"* and *"Mountain"* are related as *"IsRelatedTo"* relationship. $Snow|Landscape \rightarrow Mountain$ suggests that the all relevant video shots to concept *"Snow"* within the context *"Landscape"* could be relevant to concept *"Mountain"*.

- ***IsPartOf Role*** The *"IsPartOf"* role between c_i and t_j is defined if c_i is part of t_j, which is denoted as: $\{c_i\} \in t_j$. As an example, *"Sky"*, *"AirPlane"* and *"AirPlane_Flying"* are related as *"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"*.

- ***Includes Role*** The *"Includes"* role between t_i and c_j is defined if t_i includes c_j, which is denoted as: $t_i \supset c_j$. As an example, *"CarRacing"* and *"Car"* are related as *"Includes"* relationship. *"CarRacing"* $\supset$ *"Car"* suggests that the all relevant video shots to context *"CarRacing"* could be relevant to concept *"Car"*.

In order to enable handling real world situations, we introduced for every defined role a degree of confidence α where $\alpha \in [0, 1]$. In addition, for each role, a μ function is defined that aims to compute respectively for each related pairwise $< c_i, c_j >$, $< c_i, t_j >$ and $< t_j, c_i >$ a degree that c_i supplied for c_j, c_i supplied for t_j and t_j supplied for c_i. α and μ are generated automatically through Abduction Engine based on βeta function [Aouiti et al. 2003]. Generally, the βeta function is defined as follows:

$$\beta(x) = \begin{cases} (\frac{(x-x_0)}{(x_c-x_0)})^p(\frac{(x_1-x)}{(x_1-x_0)})^q & if\ x \in [x_0, x_1] \\ 0 & otherwise \end{cases} \tag{4.5}$$

Where:

- $p > 0, q > 0$;
- x_0 and x_1 are real parameters;
- $x_c = \frac{(px_1+qx_0)}{(p+q)}$.

According to relevance degrees proposed in our context annotation framework, our fuzzy ontology O^f employs two qualifiers ($Q = \{Weak, String\}$), in order to provide a fine-tuning of degrees of confidence. Thus, each rule is *"Strong"* qualified if its degree of confidence is greater than 0.5, else *"Weak"* qualified.

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 βeta Fuzzy Systems (MAGAD-BFS), proposed in [Kallel & Alimi 2006], as an Abduction Engine.

Based on genetic algorithm (GA), MAGAD-BFS allows optimizing a fuzzy logic system (FLS) with Beta membership functions [Alimi et al. 2000]. It consists of minimizing the number of βeta fuzzy rules N_{Rf}, which are formulated according to the equation 4.6, while adjusting βeta function p and q parameter's (obj_{fun}) of each rule until a desired precision ϵ.

$$R_j : \{R^f_{ct}, R^f_{cc}, R^f_{tc}\} : \ IF\ (X\ is\ Q^j)\ THEN\ (Y = f_i(X)) \tag{4.6}$$

Where:

- $X = C \cup T$ is an input variable;
- $Y = C \cup T$ is an output variable;
- Q^j is a linguistic qualifiers of input variable;
- f_j is the output of the j^{th} fuzzy rule.

The objective function (obj_{fun}) to be minimized is defined as follows:

$$f^*(X) = [\sum_{j=1}^{N_{Rf}} f_j(X)\beta_j(X)] \tag{4.7}$$

$$obj_{fun} = \frac{\sum_{i=1}^{m+n}[Y_i - f^*(X_i)]^2}{\sum_{i=1}^{m+n} Y_i^2} * (\frac{N_{R^f} - N_{min}}{N_{max} - N_{min} + 1}) \tag{4.8}$$

Where:

- N_{min} and N_{max} are respectively the minimum and the maximum number of fuzzy rules allowed in the final βeta fuzzy system;
- $\mu_j = \beta_j$ is a βeta function that activates the j^{th} fuzzy rule.

Deduction engine

In the indexing process, each video shot V_{S_k} is ranked with a probabilistic measure $P(V_{S_k}|c_i)$ or $P(V_{S_k}|t_j)$. Based on the latter scores, 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_i or a context t_j *"Relevant"* in a video shot V_{S_k} if $P(V_{S_k}|c_i)$ respectively $P(V_{S_k}|t_j)$ is grant than 0.7. However, a concept or a context is qualified by *"Not-Relevant"* in a video shot V_{S_k} if $P(V_{S_k}|c_i)$ respectively $P(V_{S_k}|t_j)$ 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 field, two cases arise: when a fuzzy rule is *"Strong"* qualified or *"Weak"*.

In the first case, the deduction engine proceeds as follow: Let R_k^f a fuzzy rule defined as :R_k^f: c_i is *Strong RelatedTo cj* within $t_{k'}$ and let $P(V_{S_i}|c_i)$ and $P(V_{S_i}|t_{k'})$, respectively, a score detection of concept c_i and context $t_{k'}$ in the same video shot V_{S_i}. The optimal score, or the deduced relevance degree, of the fuzzy rule R_k^f outputs, denoted as $\alpha'_k(c_j)$, is computed as follow:

$$\alpha'_k(c_j) = \mu_k * (max\{P(V_{S_k}|c_i), P(V_{S_i}|t_{k'})\}) * \mu_{Strong}(\alpha_k) \tag{4.9}$$

Where μ_k and α_k are, respectively, the βeta membership function and the confidence degree of the k^{th} fuzzy rule according the role *"IsRelatedTo"*.

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

$$\alpha'_k(c_j) = \mu_k * (min\{P(V_{S_k}|c_i), P(V_{S_i}|t_{k'})\}) * \mu_{Weak}(\alpha_k) \tag{4.10}$$

The same approach is built by the deduction engine for the other rules according the role *"IsPartOf"*, *"Includes"* and *"Generalisation"*.

4.3.3 Fuzzy Ontology Construction

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.

Knowledge extraction

In order to build the context space, a large-scale corpus is requested for generalizing contextual information and relationships between contexts and concepts. Thus, we explored the development data set provided by the evaluation campaign TrecVid IACC.1.Tv10.Training2 which is composed of 119 685 shots. Each shot is manually assigned to a predefined context specifying the meaning of its contents recognized by experts.

Fuzzy Rules Abduction

In the aim to discover fuzzy rules in the form of *"Includes"*, *"IsPartOf"* and *"IsRelatedTo"*, the abduction engine is trained by the use of the semantic knowledge. Thus, for every output of the above enumerated 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 α and the β membership function μ, as shown in Table 4.4.

4.3.4 Experiments

In this section, we discuss the obtained results for an experiment that we conducted on TrecVid 2010 dataset. The latter dataset is widely employed for the evaluation of video semantic indexing accuracy.

Table 4.4: A Partial view of the abducted Fuzzy rules

Name	The abducted fuzzy rule	Qualifier	βeta membership function
Generalization	$Airplane\_flying : Vehicule$	Strong	-
Generalization	$Airplane \mid Airplane\_flying \longrightarrow sky$	Strong	$p = 10,\ q = 0.01$
	$Snow \mid Landscape \longrightarrow Moutain$	Weak	$p = 0.01,\ q = 1$
	$\{Snow, Mountain\} \mid Landscape \longrightarrow Sky$	Strong	$p = 11,\ q = 0.1$
	$Person \mid Studio\_News \longrightarrow Anchorperson$	Strong	$p = 8,\ q = 0.03$
IsPartOf	$\{sky, Trees\} \in Landscape$	Strong	$p = 8,\ q = 0.01$
	$\{Building, Sky, Road, Car\} \in Urban$	Strong	$p = 9,\ q = 0.02$
Includes	$Landscape \supset Snow$	weak	$p = 2,\ q = 2$

The main goal of our experiment is to evaluate the use of the context space for semantic concept detection and to evaluate the effectiveness of our proposed knowledge based approach compared to existing techniques.

At first, we evaluate the effectiveness of the use of a semantic knowledge induced by the proposed fuzzy contextual ontology to enhance the detection of semantic concepts and to better enhance the semantic interpretation of a video content. Thus, we use the following metrics: *inferred average precision* (*infAP*), the *precision* (P) and the *recall* (R).

By the use of the contexts defined in figure 4.8, we obtained results reported in table 4.5.

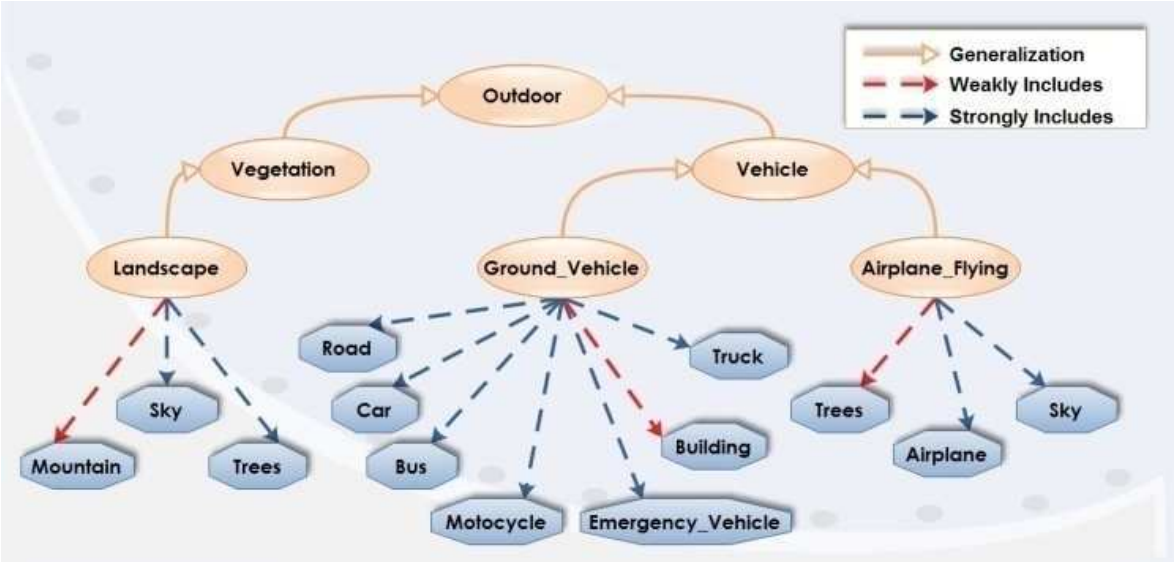


Figure 4.8: Partial view of concept distribution generated by contextual experts annotation

As displayed in table 4.5, the video indexing accuracy is clearly improved when a knowledge-based approach is used. In fact, when the LSCOM ontology is used, the precision improvement

Table 4.5: TrecVid 2010: Concept retrieval performance for different Concept detection methodologies

Semantic Concepts	**Concept Detector**			**LSCOM**			O^f	
	infAP	P	R	infAP	P	R	P	R
Outdoor	-	0.52	0.59	-	**0.88**	**0.77**	**0.9**	**0.82**
Vegetation	0.1	0.74	0.68	0.1	0.74	0.68	**0.93**	**0.87**
Landscape	-	0.6	0.79	-	0.6	0.79	**0.7**	**0.82**
Sky	-	0.66	0.9	-	0.66	0.9	**0.85**	**0.95**
Trees	-	0.62	0.72	-	0.62	0.72	**0.73**	**0.82**
Mountain	-	0.68	0.8	-	0.68	0.8	**0.83**	**0.85**
Ground_Vehicle	0.043	0.3	0.66	**0.18**	**0.6**	**0.73**	**0.69**	**0.75**
Road	-	0.43	0.6	-	0.43	0.6	**0.88**	**0.9**
Car	0.075	0.42	0.64	**0.17**	**0.58**	**0.73**	**0.79**	**0.83**
Bus	-	0.52	0.73	-	0.52	0.73	0.52	0.73
Bicycles	0.142	0.67	0.92	**0.185**	**0.82**	**0.97**	**0.83**	0.97
Emergency Vehicle	-	0.9	0.83	-	0.9	0.83	0.9	0.83
Building	0.022	0.18	0.22	**0.1**	**0.5**	**0.43**	**0.55**	**0.45**
Truck	-	0.35	0.37	-	0.35	0.37	0.35	0.37
Airplane Flying	0.102	0.8	0.78	0.102	0.8	0.78	**0.83**	**0.79**
Airplane	-	0.5	0.6	-	**0.6**	0.6	**0.71**	**0.69**

of semantic concept detection in the order of 11%. However, we obtained 21% through the use of O^f ontology. This improvement 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^f 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^f performs best for 16 (6 context and 10 concepts) for 17 high level feature. This result is rather obvious: the proposed ontology O^f tries to represent the context 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 indexing systems. In addition, we note that using the deduction engine has improved the ranking 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 tar-

geted the context. Although this recall improvement (about 2%), the precision improvement has declined.

4.3.5 Discussion

The experiments that we conducted indicate clearly that semantic concepts could be efficiently detected when a knowledge-based approach is incorporated within a video indexing system. Thus, the core contribution of this work is the implementation of a fuzzy contextual ontology.

In fact, the first experiment dealt with the LSCOM as a knowledge back-end for the deduction engine. The obtained result showed that such knowledge-based approach could deduce and detect further semantic concepts. Then, the proposed context-based fuzzy ontology O^f defined fuzzy semantic relationships between semantic concepts and semantic contexts. This ontology showed better and promising result.

Nevertheless, the latter aims to model knowledge of concepts which are extracted from a valuable data-source through an abduction engine. Generally, video annotation tools provide as outputs valuable information about semantic interpretation for video content [Dasiopoulou et al. 2011]. In literature, the available annotation tools do not support contextual information during the annotation process. In the next section, we display our proposed context-based video annotation tool.

4.4 Collaborative Annotation

The next proposition aims to improve the video annotation results through the contextual information [Ksentini et al. 2012]. Then, we focus on the collaborative annotation through sharing the past annotations in the aim to manage conflict situations. Also, we introduce semantic contexts in the annotation process in order to provide answers to the sensorial problem: one concept can present different meaning within different contexts.

4.4.1 Collaborative Annotation

The collaborative aspect of the proposed annotation tool aims to promote annotation sharing between annotators while being guided by visual tools a better annotate images. In order

to assist the annotator for better video semantic comprehension, we integrate the following tools.

Detection shots of key frames: is a tool that makes it possible to add useful information for image annotation, and to better identify the context in which the objects appeared, either by listening to the sound track or the follow-up of object movements. From the video descriptions, we recovered the temporal position of the key frame to annotate, the time beginning of the representative plan and its duration.

Sharing passed interpretations by other annotators: we think that it is very important to manage conflicts which can occur. For instance, for a given image, an annotator can refer to the previous annotations presented in chronological order in order to take idea and then to correctly annotate the image. We estimate inter-annotator agreement in order to manage conflicting situations.

Automatic suggestion of concepts: The annotation tool suggests to the user some concepts related to choses ones. The suggestion is made by a statistical study that we conducted on annotated datasets delivered by TrecVid2010. In fact, these statistical studies look for inter-concepts co-occurrence. Then, and when the annotator chooses a concept for a given image, the annotation tool looks for co-occurred concepts in order to suggest them to the annotator.

Ontology driven annotation: generally, the annotation tools use informal annotation (either binary or free texts). In our proposed annotation tool, we propose to integrate an annotation controlled using concepts from the LSCOM ontology in order to have a formal and standard annotation. The ontology used is presented by a tree structure which allows to the annotators to traverse the concept list and select pertinent ones efficiently.

The figure 4.9 illustrates the proposed annotation tool.

4.4.2 Conceptual Relationship Mining

In order to estimate the conceptual relationships between the concepts detected during the annotation process, we represent them by characteristic vectors. Then , we calculate the inter-

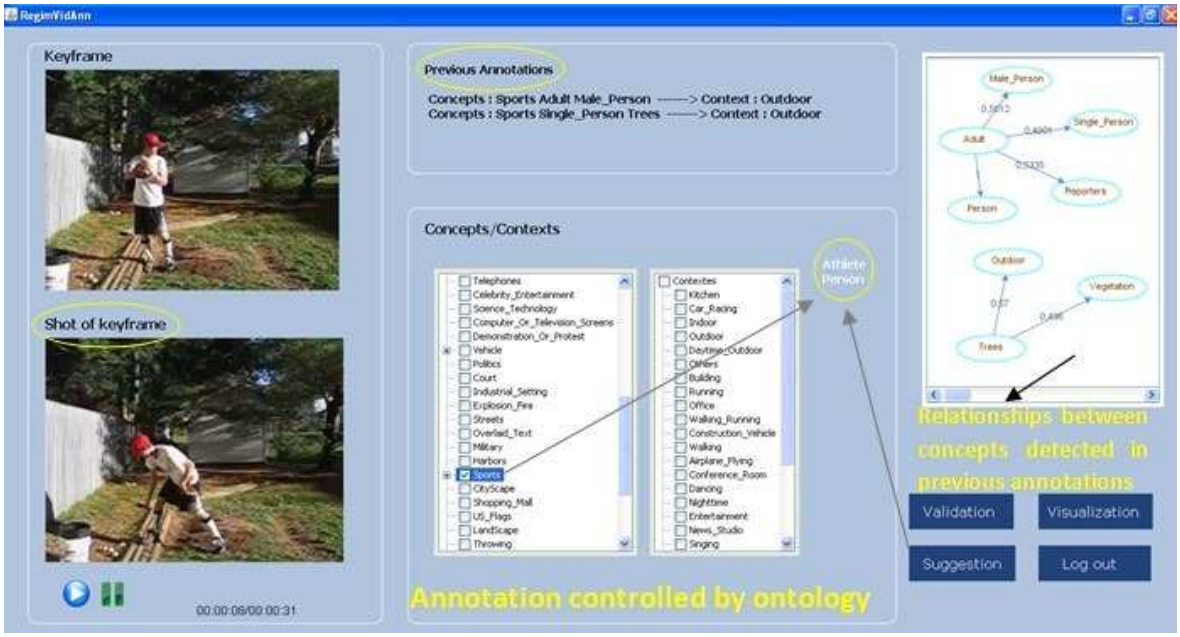


Figure 4.9: Overview of the proposed Collaborative Annotation Tool

concepts similarities. These vectors are defined by analyzing the final results of the video annotation process. Thus, we define a dynamic matrix whose lines represent the annotated key-frames, and the columns of the annotated concepts.

As concepts have various appearances according to the context in which they appear, we propose to add the notion of context in our calculations of conceptual relationships. Since the similarity between two concepts C_i and C_j varies from a context to another, we extract from the initial matrix sub-matrices. Each one of the latters represents the frequencies of concept appearances in the images in a well-defined context.

Once the vectors are defined, we calculate the similarity between the concepts by adopting the similarity measure *Cosine* Similarity. This latter is frequently used as a measurement of resemblance between two objects. Thus it is defined as follows:

$$cosine(c,d) = \frac{\vec{C_i}.\vec{C_j}}{|C_i|.|C_j|} \tag{4.11}$$

4.4.3 Visualization

The conceptual relations withdrawn in the preceding section do not make it possible to appreciate in an easy way the similarity between the concepts. It is thus preferable to have a comprehensive view of these semantic relations for better assimilating them. The generated

visual graph comprises a set of nodes and a set of undirected arcs that respectively represent the semantic concepts and semantic relationships 4.10.

In order to emphasize the notion of context, we propose to divide the contextual graph into sub-graphs of which each one represents the conceptual relationships between the concepts in a given context.

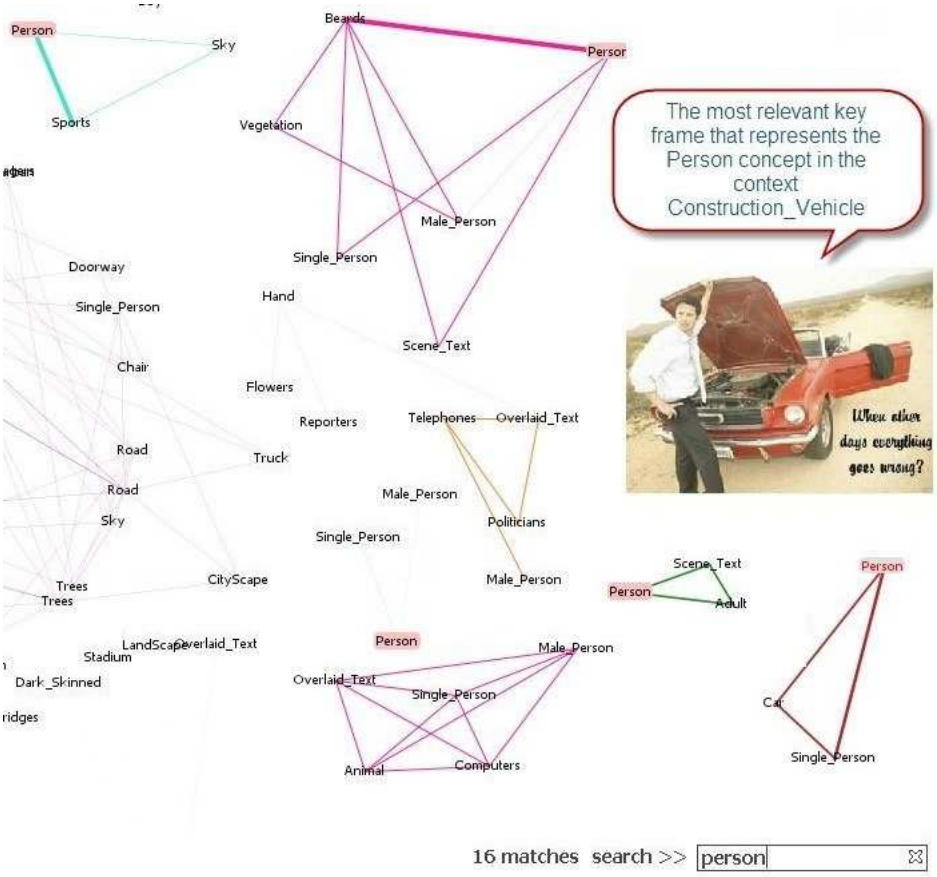


Figure 4.10: Visualization of Conceptual Relationships

4.4.4 Discussion

The proposed collaborative annotation tool delivers an annotated dataset where each image/shot is tagged by a set of semantic concepts/contexts. Based on such annotation outputs, we conducted then an earlier statistical study on concepts co-occurrence. We obtained then interesting relationships that could be considered as valuable knowledge. Indeed, the relationship between contexts and concepts cannot be revealed directly through annotated images/shots. We believe that such valuable knowledge is crucial to enhance multimedia content indexing: given a defined context in an image and inter-relationships between concepts

within that context, the detection of a concept may lead to deduce the existence of other concepts. The importance of this knowledge to improve the indexing performance is detailed in the next chapter.

4.5 Conclusion

In the present chapter, we presented our first proposition for a contextual knowledge based framework for enhancing a semantic interpretation about a multimedia content. The proposed framework deals with rich semantic structure in order to model many information in relation to semantic concepts and their interrelationships.

Our framework effectiveness and performance were proved on a multimedia benchmark (TrecVid 2010). The effectiveness of our knowledge based framework, in terms of precision and recall, is proved on diverse concepts.

Flickr images. In the next chapter, we investigate more work on the automation of the abduction engine, and modeling a generic ontology structure in order to handle various information from various fields.

Chapter 5

Fuzzy Context-Based Ontology Generation Framework for Reasoning in Multimedia Content

In this chapter, we discuss the second contribution C_2: a scalable and generic contextual ontology based approach for reasoning with video interpretations. Our approach is based on a new fuzzy knowledge management: from extracting and populating valuable knowledge, to fuzzy reasoning and evolving. The scalability and the generic aspects are also discussed. An evaluation of the effectiveness of the produced framework is then presented.

The rest of this chapter is structured as follows. In Section 5.1, we present the motivations of our proposal, we review some existing approaches and we emphasize their limitations. In Section 5.2, we introduce the proposed fuzzy ontology based approach for managing and reasoning with semantic interpretation. Section 5.3 presents the evaluation of the proposed framework through the assessment within the *ImageClef 2012* dataset. Finally, the chapter is concluded in Section 5.5.

5.1 Context and Motivations

The use of ontologies in multimedia retrieval alleviates semantic barriers, and promising results proved this trend. However, the multimedia community faces newer issues.

At first, most of the ontology content used by multimedia retrieval systems are populated through manually gathered knowledge. These knowledge are defined by experts a particular domains (like medicine [Rozilawati binti & Masaki 2011], Athletics [Paliouras et al. 2011a], ...). Nevertheless, such a manual knowledge definition is a high cost process [Song et al. 2009], and an automated knowledge discovery from a data source should be more addressed.

Also, in literature, diverse knowledge structures were proposed for handling advanced interrelationships between semantic concepts and contexts [Bannour & Hudelot 2014]. Nevertheless, such ontology conceptualization is rather closed to particular multimedia content domains. Indeed, proposed ontologies conceptualization is unable to cover different multimedia contexts. Such a capability makes ontology-based approaches ready to handle generic knowledge, and then to support an automated ontology population.

Furthermore, the multimedia community is considering a *semantic context* as the key importance in multimedia retrieval approaches [Mylonas et al. 2009, Nguyen 2010, Elleuch et al. 2011, Perpetual Coutinho et al. 2012]. Many definitions were proposed for the term *semantic context*. Generally, they define it as a particular event, or as surrounding objects within a shot, ... Accordingly, a formal definition for the term *semantic context* is inquired in order to promote automated knowledge extraction and ontology population.

The aforementioned problems elicit a challenging task toward an efficient semantic analysis of large-scale multimedia contents. We believe that the use of ontologies within multimedia retrieval approaches should take into consideration a huge amount of knowledge to handle and reason with. This leads to focus on more automated method for both extracting knowledge and populating the ontologies. Such research direction may allow open challenges and opportunities in multimedia retrieval.

In the last decade, a several research works provided various video annotation tools [Dasiopoulou et al. 2011, Ksentini et al. 2012] (like *VIA*, *VideoAnnEx*, *Ontolog*, *Advene*, *Elan*, *Anvil*, ...). And with the emergence of multimedia benchmarks, (like *TrecVid* [Over et al. 2013] and *ImageClef* [Thomee & Popescu 2012]), large-scale multimedia datasets were annotated. We consider that such valuable sources (the annotated data sets) should be used not only for training multimedia semantic concept detectors, but also to gather valuable knowledge that could be used to populate the ontologies content, then to reason with, and finally, to improve semantic interpretation capabilities for multimedia retrieval.

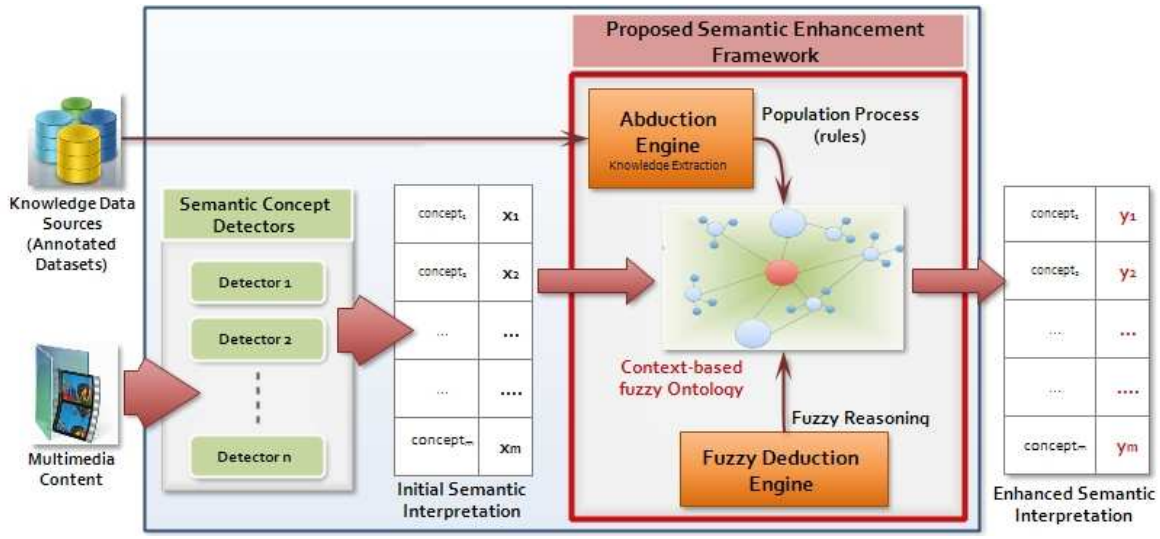


Figure 5.1: Proposed fuzzy context-based ontology framework for semantic Interpretation

In order to explore further this research direction, we aim to contribute in the multimedia community by providing a context-based ontology automated generation framework for improving multimedia content retrieval efficiency. We propose mainly a machine-driven method for extracting valuable knowledge from video annotated data sets, and also a novel and generic ontology engineering for handling and reasoning with knowledge.

5.2 The Proposed Fuzzy Context-Based Ontology Framework

This section introduces our framework for an automated construction of a generic fuzzy ontology in order to handle an initial semantic interpretation about a multimedia content as input and to generate, then, an enhanced one as output (see figure 5.1). The present section provides key ideas behind the automated generation of fuzzy context-based ontology. Thus, we first display the proposed ontology knowledge structure. Then we explain how to populate that ontology through an automatic knowledge extraction method. Then, we point out how to enhance a semantic interpretation through a deduction engine. Finally, the ontology evolving task and the framework scalability are discussed.

5.2.1 Ontology Structure

The objective of our proposed structure for modeling fuzzy knowledge in our ontology is to handle all possible relationships that can exist between semantic concepts within a defined context.

Classical ontology description languages are not relevant to handle fuzzy knowledge. *Fuzzy Description Logics* have been introduced by various approaches to handle uncertainty and vagueness. In our work, we used the $f-\mathcal{SHIN}$ *Description Logics* [Stoilos et al. 2005a], and we used the *"tableau"* algorithm [Horrocks & Sattler 2005] for reasoning in $f-\mathcal{SHIN}$. Fuzzy knowledge axioms are grouped in three parts: fuzzy $ABox$ $\mathcal{A}$ for individuals, fuzzy $TBox$ $\mathcal{T}$ for concepts, and fuzzy $RBox$ $\mathcal{R}$ for roles.

Due to the large amount of multimedia content to handle, it is important to adopt a modular modeling approach. Such ontology modeling has gained widespread attention. Therefore, we propose to define a fuzzy ontology per a defined context.

Let $\mathcal{K}^f$ be a set of fuzzy ontologies defined as follows:

$$\begin{aligned} \mathcal{K}^f \quad = \quad & \{\mathcal{K}^f_{t_1}, \mathcal{K}^f_{t_2}, \ldots, \mathcal{K}^f_{t_n}\} \text{ a set of } n \text{ fuzzy ontologies} \\ & \text{where } \mathcal{K}^f_{t_k} \text{ is a fuzzy ontology for the context } t_k. \end{aligned} \tag{5.1}$$

Definition 2. *A fuzzy ontology $\mathcal{K}^f_{t_k}$ of the context t_k can be defined as follows:*

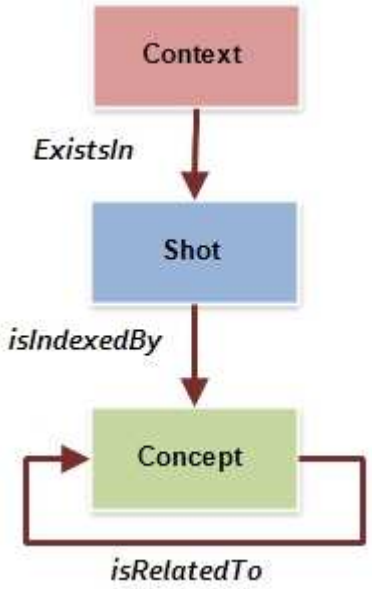
$\mathcal{K}^f_{t_k} = \langle \mathcal{T}, \mathcal{R}, \mathcal{A} \rangle$, *where*

$$
\begin{aligned}
\mathcal{T} &= \{\mathsf{Shot} \sqsubseteq \top, \\
&\quad \mathsf{Concept} \sqsubseteq \top, \\
&\quad \mathsf{Context} \sqsubseteq \top, \\
&\quad \mathsf{Context} \sqsubseteq \leq 1\mathsf{ExistsIn}.\mathsf{Shot}, \\
&\quad \mathsf{Shot} \sqsubseteq \exists \mathsf{isIndexedBy}.\mathsf{Concept}, \\
&\quad \mathsf{Concept} \sqsubseteq \exists \mathsf{isRelatedTo}.\mathsf{Concept}\}, \\
\mathcal{A} &= \{(\langle \mathsf{Context}, \mathsf{Shot} \rangle : \mathsf{ExistsIn}) \geq p_1 \\
&\quad (\langle \mathsf{Shot}, \mathsf{Concept} \rangle : \mathsf{isIndexedBy}) \geq p_2 \\
&\quad (\langle \mathsf{Concept}, \mathsf{Concept} \rangle : \mathsf{isRelatedTo}) \geq p_3\}, \\
\mathcal{R} &= \{\mathsf{Trans}(\mathsf{isRelatedTo}) \\
&\quad \mathsf{Disjoint}(\mathsf{isRelatedTo}, \mathsf{isRelatedTo}^-)\}
\end{aligned}
\tag{5.2}
$$

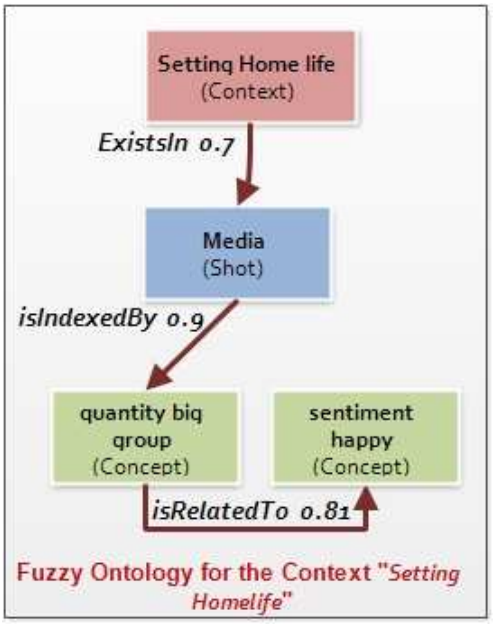
Shot, Context and Concept are defined as ontology concepts. The $ABox$ $\mathcal{A}$ illustrates possible relationships between concepts, contexts and multimedia content shots defined in the $TBox$ $\mathcal{T}$. Then, isRelatedTo, isIndexedBy and ExistsIn are defined as three roles (figure 5.2a).

The role $\mathsf{isRelatedTo}(\mathsf{Concept}, \mathsf{Concept})$ in the $\mathcal{K}^f_{t_k}$ ontology depicts that there is a generic relationship of a fuzzy weight p_3 between two concepts within the context t_k. The two roles $\mathsf{ExistsIn}(\mathsf{Context}, \mathsf{Shot})$ and $\mathsf{isIndexedBy}(\mathsf{Shot}, \mathsf{Concept})$ translate a semantic interpretation for a given multimedia content shot. While the role $\mathsf{ExistsIn}(\mathsf{Context}, \mathsf{Shot})$ depicts that the context Context figures in the shot Shot with a fuzzy weight p_1, $\mathsf{isIndexedBy}(\mathsf{Shot}, \mathsf{Concept})$ role depicts that the concept Concept exists in the shot Shot with a fuzzy weight p_2.

For instance, figure 5.2b illustrates the ontology structure and individuals for the ontology $\mathcal{K}^f_{Setting\ Home\ life}$ which is specific for the context Setting Home life. A shot Media is indexed by this context with a fuzzy degree of 0.7. Then, the role $\mathsf{isIndexedBy}$ (Media, QuantityBigGroup) depicts the fact that the shot Media is indexed by the concept QuantityBigGroup with a fuzzy degree of 0.9. And finally, the role $\mathsf{isRelatedTo}(\mathsf{QuantityBigGroup}, \mathsf{SentimentHappy})$ defines a specific relationship between concepts QuantityBigGroup and SentimentHappy with a fuzzy degree of 0.6 within the ontology $\mathcal{K}^f_{Setting\ Home\ life}$. Thus, if a shot is relevant to the con-



(a) Conceptual representation of a fuzzy ontology for a given context



(b) Example of the $\mathcal{K}^{f}_{Setting\ Home\ life}$ ontology content for the context *"Setting Home life"*

Figure 5.2: Conceptual representation and an example of a fuzzy ontology

cept QuantityBigGroup, it could be relevant too to the concept SentimentHappy if the context Setting Home life exists in this shot.

As described in figure 5.1, the proposed framework aims to handle a semantic interpretation as input and to generate an enhanced one as output. Mathematically, and as input, we have a set of shots where each one is tagged by some semantic concepts. And as output, the framework delivers the same set of images, but with an improved concept tagging. For the ontology side, this set of shots and their related concepts are translated to *Abox* $\mathcal{A}$ within the fuzzy context-based ontology set.

Definition 3. *Let Input and Output be sets of quadruplet* $(t_k, c, shot_i, (\alpha_1, \alpha_2))$. *Each quadruplet depicts that the shot* $shot_i$ *is tagged by the concept* c *by a fuzzy weight* α_2 *within the context* t_k, *and tagged by the context* t_k *by a fuzzy weight* α_1. *These quadruplets are correlated with the* isIndexedBy *and* ExistsIn *roles* ($\langle (t_k, s) : \text{ExistsIn} \geqslant (\alpha_1) \rangle$ *and* $\langle (shot_i, c) : \text{isIndexedBy} \geqslant (\alpha_2) \rangle$).

5.2.2 Abduction Engine and Ontology Population

The population process aims to extract new knowledge through an abduction engine and to populate the ontology with new instances defined in a specific context, as well as with their properties and relations. In what follows, we itemize the proposed abduction engine, then we show how to populate the fuzzy context-based ontology with new extracted knowledge.

The Abduction Engine

In this section, we are particularly interested in the isRelatedTo role (since the other two roles instances are directly populated from an initial semantic interpretation).

The key idea behind our method for the abduction engine is to explore similarities between semantic concepts within an annotated multimedia content. Indeed, many annotated multimedia datasets (particularly image and video content) are available and accessible: *ImageCLEF Flickr Photo Annotation and Retrieval* [Thomee & Popescu 2012] and TrecVid [Over et al. 2013] training datasets, *Flickr API* and *Picasa API* for image/video crawling, ... are valuable data sources that can be explored in order to extract knowledge to be inserted into an ontology and to be used then to enhance a semantic interpretation.

We would like to remind that a *semantic context* is an abstract meaning that cannot be well defined because it makes sense only in particular situations [Elleuch et al. 2011, Ksentini et al. 2012]. For example, the concept *"airplane_flying"* is considered as a context since it specifies a particular relationship between two other concepts: *"sky"'* and *"airplane"*.

In our case, we define a context as follows:

Definition 4. *A context is defined as a concept that can stipulate a specific relationship between other concepts.*

Thanks to such a definition, we can give a more abstract meaning of a context rather than a spatial or temporal information [Brilhault 2009], and also provide an automated way to discover contexts within a concept set. Then, for a given concept c, we look for some eventual similarities between other concepts within shots annotated with the concept c. If such relationships are found, the concept c is considered as a context.

We use the vector space model based method to compute the similarity between concepts.

$$sim(c,d) = cossim(c,d) = \frac{\vec{V_c}.\vec{V_d}}{|V_c|.|V_d|} \tag{5.3}$$

where for each concept c, a weighted vector V_c can be constructed as

$$V_c = \{v_c^{shot_1}, v_c^{shot_2}, \ldots, v_c^{shot_m}\} \tag{5.4}$$

where $v_c^{shot_i} \in [0,1]$ is the weight of the concept c in the video shot $shot_i$, and

$$v_i = tf_{shot_i,c}.idf_c = tf_{shot_i,c}.log\frac{|S|}{|s_c|} \tag{5.5}$$

where $tf_{shot_i,c}$ is the frequency of the concept c in the video shot $shot_i$, $|S|$ is the total number of shots and $|s_c|$ is the number of shots tagged by the concept c.

Based on what has just been developed, we propose a formal method to discover contexts. Thus, we define the similarity $sim_{t_k}(c,d)$ as a similarity measurement between the two concepts c and d within the context t_k. This similarity will be used as a fuzzy degree for the rule that relates the concepts c to the concept d within the role isRelatedTo for the $\mathcal{K}_{t_k}^f$ ontology.

Furthermore, isRelatedTo is a non-symmetric role (as appears in the *Rbox* $\mathcal{R}$). As an example, when talking about the concept *"car_racing"*, it is obvious that we talk about the concept *"car"*, but the opposite in not often true. Thus $sim(car, car\_racing) \neq sim(car\_racing, car)$. To do that, the similarity function that is used to compute the similarity $sim_{t_k}(c,d)$ in a defined context t_k has to be non-symmetric. Then, the two vectors V_c and V_d are modified as follow to ensure that $sim_{t_k}(c,d) \neq sim_{t_k}(d,c)$:

$$\begin{aligned} V_c &= \{v_c^{shot_1}, v_c^{shot_2}, \ldots, v_c^{shot_m}\} \\ V_d &= \{v_d^{shot_1}, v_d^{shot_2}, \ldots, v_d^{shot_m}\} \\ &\quad \text{where for every considered shot } shot_i \text{ in both } V_c \text{ and } V_d, \\ &\quad \text{we have } v_c^{shot_i} \neq 0 \text{ and } v_{t_k}^{shot_i} \geqslant \gamma \end{aligned} \tag{5.6}$$

we suggest $\gamma \in [0,1]$ as a threshold to judge if a concept is strongly relevant to a video shot or not. Thus, only tagged shots by the concept c and that are strongly tagged by the

concept/context t_k are considered for both V_c and V_d in order to guarantee the non-symmetric aspect of the role isRelatedTo.

Ontology Population

The ontology population passes through two steps:

i. *Knowledge Population*: For each defined context t_k, this step constructs an empty ontology $\mathcal{K}^f_{t_k}$ based on the proposed structure detailed in the previous section. Then, the context t_k is instantiated as an individual from the class Context. Next, for every new similarity $sim_{t_k}(c, d)$ computed between two concepts c and d within the context t_k, two new instantiations from the class Concept are introduced in the ontology for both concepts c and d, and finally a new isRelatedTo role instantiation between c and d is inserted to the $\mathcal{K}^f_{t_k}$ ontology with a fuzzy weight equal to this similarity value.

ii. *Semantic Interpretation Population* In this step, semantic interpretations about shots are introduced to the set of fuzzy ontologies $\mathcal{K}^f$. For a given shot $shot_i$, and for each context t_k, if $shot_i$ is tagged by the context t_k, then:

   (a) the shot $shot_i$ is instantiated as an individual from the class Shot,

   (b) a new instantiation of the role ExistsIn is introduced between $shot_i$ and t_k within the ontology $\mathcal{K}^f_{t_k}$,

   (c) for every concept c that tags $shot_i$, c is introduced to the ontology $\mathcal{K}^f_{t_k}$ as an individual from the class Concept, and a new IsIndexedBy role instantiation is created between $shot_i$ and c.

Once the two steps are performed, a set of fuzzy context-based ontology is built and ready to fire the reasoning engine (the deduction engine) in order to generate an enhanced semantic interpretation.

5.2.3 Ontology Reasoning

Within the scope of reasoning in this paper, we are particularly interested in the outputs that such a system produces. In such a case, and for each ontology $\mathcal{K}^f_{t_k}$ of the context t_k, we generate the corresponding ontology $\mathcal{K}'^f_{t_k}$ as an ontology that handle enhanced semantic

interpretations. This is done by interpreting the *Abox* $\mathcal{A}$ in order to look for updated or additional relationships between Shot and Concept (for the role name isIndexedBy).

We used and adapted the *"tableau"* based reasoning algorithm [Dentler et al. 2011]. The latter adopts a federated reasoning process with our set of fuzzy ontologies. Thus, each fuzzy ontology is associated with a local reasoner. The final and global reasoning result is inferred through all the local reasoner results.

Considering concepts and roles as fuzzy sets, the semantics of concepts and roles are defined by fuzzy interpretations $\mathcal{I} = \langle \triangle^{\mathcal{I}} =, \cdot^{\mathcal{I}} \rangle$, where $\triangle^{\mathcal{I}}$ is a nonempty domain, and $\cdot^{\mathcal{I}}$ is an interpretation function mapping individuals a into $a^{\mathcal{I}} \in \triangle^{\mathcal{I}}$, and mapping concept (roles) names $A(R)$ into membership functions $A^{\mathcal{I}}(R^{\mathcal{I}}) : \triangle^{\mathcal{I}}(\triangle^{\mathcal{I}} * \triangle^{\mathcal{I}}) \rightarrow [0,1]$. An interpretation $\mathcal{I}$ satisfies a knowledge database $\mathcal{K}^f$, if and only if $\mathcal{I}$ satisfies any axiom in $\mathcal{R}$, $\mathcal{T}$ and $\mathcal{A}$. $\mathcal{K}^f$ is satisfiable if and only if it has a fuzzy model.

We define the symbols $\rhd$ and $\lhd$ as a placeholder for the inequalities ($\leq$, $<$, $\geq$ and $>$) and the symbol $\bowtie$ as a placeholder for all types of inequalities.

Definition 5. *let R_A be the set of roles occurring in $\mathcal{A}$, $I_{\mathcal{A}}$ the set of individuals in $\mathcal{A}$ and $\mathcal{X} = \{\leq, <, \geq, >\}$. A fuzzy tableau T for $\mathcal{A}$ w.r.t $\mathcal{R}$ is a quadruple $(S, \mathcal{L}, \mathcal{E}, \mathcal{V})$ in such a way that:*

- *S is a nonempty set of individuals (considered also as nodes),*
- *$\mathcal{L} : S \longrightarrow 2^{sub(A)} \times \mathcal{X} \times [0,1]$ maps each individual in S to membership triples,*
- *$\mathcal{E} : R_A \longrightarrow 2^{S \times S} \times \mathcal{X} \times [0,1]$ maps each role to membership triples,*
- *$\mathcal{V} : I_{\mathcal{A}} \longrightarrow S$ maps individuals occurring in $\mathcal{A}$ to elements in S.*

For all $s, t \in S$, $R \in R_A$ and $C, E \in sub(\mathcal{A})$, T satisfies a particular transformation rule defined in algorithm 1 (in addition to default transformation rules defined in [Stoilos et al. 2005a]).

Property 1 is a consequence of the fact that if there are three axioms in $\mathcal{A}$ where $shot_i$ is indexed by a concept c, and the context t_k exists in the $shot_i$, then the $shot_i$ could be indexed by the concept d taking into account that within the fuzzy ontology of the context t_k, there is a relationship between the two concepts c and d. If this new deduced relationship already exists in the ontology, then we only update its fuzzy weight.

if $\langle (t_k, shot_i) : \mathsf{ExistsIn} \bowtie \alpha_1 \rangle \in \mathcal{A}$ *AND*
$\langle (shot_i, c) : \mathsf{isIndexedBy} \bowtie \alpha_2 \rangle \in \mathcal{A}$ *AND*
$\langle (c, d) : \mathsf{isRelatedTo} \bowtie \alpha_3 \rangle \in \mathcal{A}$ **then**
  if $\langle (shot_i, d) : \mathsf{isIndexedBy} \bowtie \alpha_4 \rangle \notin \mathcal{A}$ **then**
  $\mathcal{A}' := \mathcal{A} \sqcup \{\langle \mathsf{shot_i, d} \rangle : \mathsf{isIndexedBy}) \bowtie \alpha_4\}$
  and $\langle \langle \mathcal{V}(shot_i), \mathcal{V}(d) \rangle \bowtie \alpha_4 \rangle \in \mathcal{E}(\mathsf{isIndexedBy})$
  where $\alpha_4 = \alpha_1 \times \alpha_2 \times \alpha_3$;
  else $\langle \langle \mathcal{V}(shot_i), \mathcal{V}(d) \rangle \bowtie \alpha'_4 \rangle \in \mathcal{E}(\mathsf{isIndexedBy})$
  where $\alpha'_4 = max(\alpha_4, (\alpha_1 \times \alpha_2 \times \alpha_3))$;
end

Algorithm 1: Transformation Rule: Property 1

We used and adapted a *"tableau"* based reasoning algorithm. For a given context-based ontology, the latter model the knowledge in form of a tree in which nodes correspond to individuals (semantic concepts, semantic contexts and shots) and edges correspond to the defined roles (isIndexedBy,isRelatedTo, and ExistsIn). Each node x is labeled with a set of concepts $\mathsf{L}(x)$ that the individual must satisfy, and each edge is labeled with a role name. The reasoning algorithm starts with a single node labeled $\{D\}$ (in our case, a node labeled as a shot), and proceeds by repeatedly applying a set of expansion rules that recursively decompose the concepts in node labels.

The reasoning process should not be endless. This problem can deal with the use of a blocking technique: stopping the expansion when a cycle is detected. The blocking procedure consists in checking the label of each new node y, and if it's a subset of the label of an existing node x, then the expansion of y is stopped: x is said to block y.

Thus, the reasoning process termination is guaranteed: all concepts in node labels are derived from the decomposition of D, so all node labels must be a subset of the sub-concepts of D. A cycle should be avoided within a finite number of expansion steps.

for *each shot* $shot_i \in$ Shot **do**
  for *each concept c connected with the shot* $shot_i$ *by an* isIndexedBy *labeled edge* **do**
    Handle_isRelatedTo_Role($shot_i$,c);
  end
end

Algorithm 2: Reasoning Process for enhancing a semantic interpretation about shots.

The isRelatedTo role is cyclic and transitive. Such a definition makes the termination and the decidability of our reasoning process hard. We adapted the *"tableau"* algorithm is such a way to solve this problem. Apart from the generic blocking technique above detailed, we

Function Handle_isRelatedTo_Role($shot_i$,c)
if *isMarked(c)* **then**
  return;
else
  for *each concept d connected with c by an edge labeled* isRelatedTo **do**
    Mark(c);
    apply Algorithm 1 ;
    Handle_isRelatedTo_Role($shot_i$,d);
  end
end

Algorithm 3: Recursive call for *"Handle_isRelatedTo_Role()"* Function

used a second blocking technique which marks every passage through semantic concepts that are connected with isRelatedTo role labeled edges. We used two functions: $Mark(c)$ which marks a concept c, and $isMark(c)$ which returns whether the concept c is marked or not yet.

Algorithms 2 and 3 detail how we extend a semantic interpretation of a given shot taking into consideration the proposed blocking technique.

The node marking test in the algorithm 3 allows to detect an eventual cycle. Then, a possible endless call to the *"Handle_isRelatedTo_Role()"* recursive function is blocked. The proposed algorithm achieved the decidability of our reasoning process.

5.2.4 Ontology Evolving

Evolving and updating fuzzy knowledge is a task that must be addressed in an ontology life cycle. Such a test consists in updating the ontology content in order to take into account both, eventual erroneous or senseless knowledge, or newer knowledge. In our work, we consider the evolution of ontology individuals content and not the conceptual level. Then, only individuals and inter-relationships in *Abox* $\mathcal{A}$ could be updated.

The evolving task proceeds as follows: Let $\mathcal{A}$ be the $ABox$ of an ontology, and $\mathcal{A}_+$ the updated one. Then the evolving task can be illustrated with the algorithm 4. The latter always considers the new discovered knowledge by updating the fuzzy weight of a isRelatedTo role to the new value defined in $\mathcal{A}_+$.

In what follows, we show how to generate the new $ABox$ $\mathcal{A}_+$ used to update the ontology content. Thus, we consider that the defined fuzzy ontologies are populated through a knowledge discovery technique based on annotation data. We consider, too, that this extracted knowledge may be inaccurate depending on many factors (annotators quality, dataset

if $\langle (c,d) : \mathsf{isRelatedTo} \bowtie \alpha \rangle \in \mathcal{A}$ *AND*
$\langle (c,d) : \mathsf{isRelatedTo} \bowtie \alpha' \rangle \in \mathcal{A}_+$ **then**
  $\mathcal{A} := \mathcal{A} \setminus \{\langle (\mathsf{c},\mathsf{d}) : \mathsf{isRelatedTo} \rangle \bowtie \alpha\}$
  $\mathcal{A} := \mathcal{A} \sqcup \{\langle (\mathsf{c},\mathsf{d}) : \mathsf{isRelatedTo} \rangle \bowtie \alpha'\}$
end

Algorithm 4: The knowledge evolving task

content quality, ...). Then, we propose to correct the ontology contents through inspecting isRelatedTo roles by human experts.

We thus propose a user interface to let the experts give their relevance judgment for a given role within a given contextual ontology. For a given context t_k and an affiliated $\{\langle (\mathsf{c},\mathsf{d}) : \mathsf{isRelatedTo} \rangle \bowtie \alpha\}$ role, this user interface shows to the expert a set of rows where each one details one shot that both the context t_k and the concept c exist. The expert then has to select one from these four options:

- *Strongly Relevant* when the concept d effectively in the showed shot. We consider that the fuzzy weight for this case is higher or equal to 0.7,
- *Weakly Relevant* When the concept d may exist in the showed shot. We consider that the fuzzy weight for this case is higher or equal to 0.3 and lower than 0.7,
- *Not Relevant* When the concept d does not exist in the showed shot. We consider that the fuzzy weight for this case is lower than 0.3,
- *Skip* when the expert cannot judge clearly if the concept d exists or not in the showed shot.

The original value of a fuzzy weight of a given role is taken into consideration and appears when the evolving user interface displays a role to an expert. In fact, every role is displayed with its fuzzy weight, and the expert judges to input his judgment (if he finds that the role is not pertinent) or to skip to analyze another role.

After the evaluation process done by the experts, the knowledge evolving task computes new fuzzy weights for evaluated roles. These weights are computed as follows: Let $N_{experts}$ be the number of experts who evaluated the role r. Let $N_{StrongR}$, N_{WeakR} and N_{NotR} be, respectively, the number of *Strongly Relevant*, *Weakly Relevant* and *Not Relevant*

judgments made by the $N_{experts}$ experts for the role r where:

$$N_{experts} \geqslant N_{StrongR} + N_{WeakR} + N_{NotR}. \tag{5.7}$$

Then, we compute the new fuzzy weight for the evaluated role r as follows:

$$\alpha' = \begin{cases} max(\alpha, \frac{N_{StrongR}}{N_{experts}}) & \text{if } N_{StrongR} \geqslant (N_{WeakR}, N_{NotR}) \\ min(\alpha, \frac{1-N_{WeakR}}{N_{experts}}) & \text{if } N_{WeakR} > (N_{StrongR}, N_{NotR}) \\ min(\alpha, \frac{1-N_{NotR}}{N_{experts}}) & \text{if } N_{NotR} > (N_{StrongR}, N_{WeakR}) \end{cases} \tag{5.8}$$

The equation above is inspired by the majority vote. The new computed role fuzzy weight is based on whether experts are commonly improving or disregarding a role.

For an analyzed role, only one new fuzzy weight will be considered and introduced to the ontology. Thus, no eventual conflicting situation occurs when firing the reasoning engine. This choice is preselected by the initial value of fuzzy role weight.

5.2.5 Approach Scalability

The scalability aspect is very important and concerns mainly two aspects: how our framework performs when a large amount of concepts and interrelationships are defined, and when the number of contexts increases.

Our proposed approach could be considered as scalable: at first, the *Tbox* $\mathcal{T}$ remains unchanged, and only new *Abox* $\mathcal{A}$ instances are being handled within the defined contextual ontologies. Then, the amount of knowledge populated within an ontology depends only on what has been gathered from annotated video datasets in addition to the semantic interpretation of the analyzed video shot. Secondly, since the proposed approach is decomposed into several ontologies, the deduction engine is applied only on *Abox* $\mathcal{A}$ that are related only to each defined contexts. This leads to an optimized and a lightened reasoning process: only partial defined roles in ontologies will be considered and fired.

In the next section, we conduct an experiment to investigate our approach effectiveness and scalability.

5.3 Experimental Study

In the previous chapter, we discussed our participation within the TrecVid 2010. Only the deduction process was taken into consideration since we used a manually predefined set of rules, and we were interested in visual (key frames) semantic analysis.

In order to illustrate the semantic enhancement of concept detection introduced by our proposed ontology-based framework, we have conducted an experiment within a different multimedia evaluation campaign. Hence, we expose the experimental setup and the obtained results of our framework within *ImageClef 2012* [Thomee & Popescu 2012] at the *Photo Annotation and Retrieval* Task. Finally, we discuss the scalability of our proposed framework.

5.3.1 Datasets Description

The *ImageClef 2012* Photo Annotation and Retrieval task provides a dataset built from Flickr social shared photos. This image dataset consists of 25 thousand images: 15 thousand for the learning process and 10 thousand for the test one. In addition to the images, 94 semantic concepts are defined referring to many and various subjects (people, nature, events, ...).

5.3.2 Evaluation metrics

In order to be able to compare different indexing approaches, various system effectiveness metrics have been used. These metrics are commonly based on precision (which is defined as the number of relevant answers as a part of the total number of retrieved ones), and recall (which is defined as the number of relevant answers as part of total relevant ones in the collection). In our experiments, we consider the two evaluation measures: *Interpolated Mean Average Precision* (map_i) and *Interpolated Geometric Mean Average Precision* ($gmap_i$) [Thomee & Popescu 2012] for the *ImageClef 2012* Flickr Photo Annotation and Retrieval task.

The $gmap_i$ is considered as an extension to the map_i measure as it uses the same computing procedure, but it's very useful when focusing on low performing semantic interpretation for particular concepts

5.3.3 Experiments with *ImageClef 2012* dataset

With this experiment, we aim to explore all the 15 thousand annotated images in order to extract valuable fuzzy Knowledge. The latter will be used then within the deduction engine to evaluate the knowledge-based semantic enhancement performances. The obtained results will be compared to a non-knowledge-based concept detection approach detailed in [Ksibi et al. 2012]. Thus, we used the results given by the non-knowledge-based method as input for our framework looking for enhancing the semantic interpretation and for proving the effectiveness of the use of fuzzy context-based ontology and fuzzy reasoning in multimedia retrieval systems.

The abduction process

In the interest of populating our proposed ontology, we start by defining the set of considered contexts. At first, we consider every concept as a context, and then, we look for specific relationships between other concepts within this context. If any relationships are found, the defined concept will be considered as a context, else, the concept will not be considered as a context. By doing so, our proposed framework can provide automatically a set of context to work with. We consider that a context exists in a shot only if this latter is strongly tagged by that context to a degree equal to or higher than 0.7. Then, we adjust the value of the threshold γ at 0.7.

Our experimentation has led to define 90 contexts among the defined 94 concepts in *ImageClef 2012* dataset. Table 5.1 enumerates a partial view of this set of contexts, and shows the first and the last 5 contexts sorted by the number of extracted isRelatedTo roles.

Taking some roles as examples, we focus on the contexts *age_adult*, *transport_car*, *setting_homelife* and *age_teenager*:

$$
\begin{array}{llll}
R1 & - & \mathsf{isRelatedTo}_{age\_adult} & (age\_teenager, sentiment\_scary) \geq 0.12 \\
R2 & - & \mathsf{isRelatedTo}_{transport\_car} & (sentiment\_happy, timeofday\_day) \geq 0.95 \\
R3 & - & \mathsf{isRelatedTo}_{setting\_homelife} & (quantity\_biggroup, sentiment\_happy) \geq 0.81 \\
R4 & - & \mathsf{isRelatedTo}_{age\_teenager} & (sentiment\_funny, quantity\_two) \geq 0.48
\end{array}
$$

Table 5.1: Distribution of extracted isRelatedTo roles according to their fuzzy weights and the defined context

	Detected Contexts	Number of extracted fuzzy rules according to their fuzzy weight intervals			
		$[0, 0.3[$	$[0.3, 0.5[$	$[0.5, 0.7[$	$[0.7, 1.0]$
1	sentiment_euphoric	783	605	510	534
2	sentiment_happy	2020	864	519	444
3	setting_sportsrecreation	1264	752	483	456
4	age_child	1045	649	493	441
5	relation_familyfriends	1150	903	398	522
86	fauna_cat	115	105	91	203
87	fauna_insect	107	63	43	192
88	fauna_rodent	43	53	43	139
89	fauna_amphibianreptile	24	29	43	127
90	fauna_spider	3	20	16	48
	All contexts	91803	50210	27435	33667
	Total	**203115** isRelatedTo **roles**			

The role $R1$ depicts that in the case of a shot tagged by the concept *age_teenager* and strongly tagged by the concept *age_adult* (here considered as a context), then the concept *sentiment_scary* could exist in the same shot. Naturally, this case is not always true, and this is why the fuzzy weight of this role is very weak (0.12). Then, the roles $R2$ and $R3$ depict strong relationships between the two concepts *sentiment_happy* and *timeofday_day* within the context *transport_car*, and between the two concepts *quantity_ biggroup* and *sentiment_happy* within the context *setting_homelife*. These two relationships are considered as strong roles (where the fuzzy weights are high).

Finally, and after extracting knowledge from the learning dataset and generating isRelatedTo roles, these roles are introduced into the corresponding fuzzy context based ontology (as detailed in section 5.2.2).

We continue to experiment deduction engine as a fuzzy reasoning process based on these roles to assess the semantic enrichment in the next section.

5.3.4 Enhancement Evaluation

To evaluate our proposed framework, we have compared our results with an image annotation method detailed in [Ksibi et al. 2012] within the *Flickr Photo Annotation Task*. The latter is based on constructing a semantic context network to depict intrinsic contextual information

between concepts in order to enhance automatic photo annotation performances. In our work, we are interested in improving this method. Thus, we used the results given by this method as input to our framework aiming at enhancing the semantic interpretation.

Considering constructed fuzzy context-based ontology set in the previous section, we introduce the *Input* set to these ontologies. Then, the deduction engine is fired. Finally, the *Output* set is extracted and translated into a semantic interpretation as an enhanced one. For the deduction engine, only isRelatedTo roles with a fuzzy weight higher or equal to $0,7$ are addressed. We think that the other roles figure weak fuzzy weights and shouldn't be used in the deduction process.

After extracting the *Output* set, the system is ready for evaluation. So, we used the evaluation tool supplied by *ImageCLEF* (*imageclef2012CR*). We carried out unit evaluations over ontologies (each ontology is evaluated separately). Finally, we applied an overall assessment showing the performance of our framework (all ontologies combined).

Table 5.2 displays enhancement performances delivered by our proposed framework over the non-knowledge-based approach detailed in [Ksibi et al. 2012]. Table 5.3 shows enhancement performances delivered by each defined fuzzy context-based ontology.

In order to make table 5.2 and table 5.3 interpretation easier, three styles were used: bold values depict enhancing performances compared to the corresponding evaluation of the system detailed in [Ksibi et al. 2012], underlined values depict a decreasing performance, and finally, the non-styled values depict same performance as reference.

Table 5.2: ImageClef 2012: Overall performance evaluation

	$\boldsymbol{map}_i$		$\boldsymbol{gmap}_i$	
	Value	*Enhancement*	*Value*	*Enhancement*
Our proposed framework	**0,1324**	5.08%	**0, 0807**	3.46%
System in [Ksibi et al. 2012]	0.126	-	0.078	

Computed performances displayed in table 5.3 and table 5.2 use the mean average measurement $\boldsymbol{map}_i$ and $\boldsymbol{gmap}_i$. The $\boldsymbol{gmap}_i$ one is very useful in case we focus on low performing semantic interpretation where scores are close to 0.0.

Analyzing $\boldsymbol{map}_i$ and $\boldsymbol{gmap}_i$ values in table 5.2, our proposed framework delivers a semantic interpretation enhancement at (respectively) 5.08 % and 3.46 %.

Table 5.3: ImageClef 2012: Unit performances evaluation

#	*Context*	*Evaluation*		#	*Context*	*Evaluation*	
		$\mathbf{map}_i$	$\mathbf{gmap}_i$			$\mathbf{map}_i$	$\mathbf{gmap}_i$
1	age_adult	0.126	0.078	46	scape_forestpark	**0.1277**	**0.0792**
2	age_baby	**0.1275**	**0.0787**	47	scape_graffiti	**0.1265**	**0.0784**
3	age_child	0.126	0.078	48	scape_mountainhill	**0.1294**	**0.0801**
4	age_elderly	0.126	0.078	49	scape_rural	**0.1279**	**0.0792**
5	age_teenager	**0.1279**	**0.0793**	50	sentiment_active	**0.1284**	**0.079**
6	celestial_moon	**0.1273**	**0.0787**	51	sentiment_calm	**0.1261**	**0.0781**
7	celestial_stars	**0.1277**	**0.0789**	52	sentiment_euphoric	**0.1281**	**0.0794**
8	celestial_sun	**0.1269**	**0.0786**	53	sentiment_funny	**0.1272**	**0.0788**
9	combustion_fireworks	**0.1268**	**0.0785**	54	sentiment_happy	**0.1274**	**0.0789**
10	combustion_flames	**0.1283**	**0.0789**	55	sentiment_inactive	**0.1277**	**0.0788**
11	combustion_smoke	**0.1283**	**0.079**	56	sentiment_melancholic	**0.1282**	**0.0791**
12	fauna_amphibianreptile	**0.1271**	**0.0785**	57	sentiment_scary	**0.1281**	**0.0788**
13	fauna_bird	**0.1268**	**0.0784**	58	sentiment_unpleasant	**0.1269**	**0.0787**
14	fauna_cat	**0.1282**	**0.0794**	59	setting_citylife	**0.1276**	**0.0789**
15	fauna_dog	**0.1275**	**0.0788**	60	setting_fooddrink	**0.1265**	**0.0783**
16	fauna_fish	**0.1275**	**0.0788**	61	setting_homelife	**0.127**	**0.0786**
17	fauna_horse	**0.1294**	**0.0794**	62	setting_partylife	**0.1272**	**0.0787**
18	fauna_insect	**0.1274**	**0.0786**	63	setting_sportsrecreation	**0.1265**	**0.0783**
19	fauna_rodent	**0.1268**	**0.0783**	64	style_circularwarp	**0.127**	**0.0786**
20	fauna_spider	**0.1262**	**0.0781**	65	style_gray	**0.1271**	**0.0787**
21	flora_flower	**0.1279**	**0.0792**	66	style_overlay	**0.1276**	**0.079**
22	flora_grass	**0.1275**	**0.0789**	67	style_pictureinpicture	**0.1272**	**0.079**
23	flora_plant	**0.1283**	**0.0792**	68	timeofday_night	**0.1278**	**0.0789**
24	flora_tree	**0.1273**	**0.0786**	69	timeofday_sunrisesunset	**0.1278**	**0.0789**
25	gender_female	**0.1286**	**0.0796**	70	transport_air	**0.1262**	**0.0783**
26	gender_male	**0.1275**	**0.0796**	71	transport_car	**0.1291**	**0.0796**
27	lighting_lenseffect	**0.128**	**0.0789**	72	transport_cycle	**0.1277**	**0.0789**
28	lighting_reflection	**0.1292**	**0.0794**	73	transport_rail	**0.1283**	**0.0794**
29	lighting_shadow	**0.1291**	**0.0792**	74	transport_truckbus	**0.128**	**0.079**
30	lighting_silhouette	**0.1269**	**0.0786**	75	transport_water	**0.127**	**0.0786**
31	quality_artifacts	**0.1286**	**0.0795**	76	view_closeupmacro	**0.1268**	**0.0785**
32	quality_completeblur	**0.1286**	**0.0793**	77	view_indoor	**0.1283**	**0.08**
33	quality_motionblur	**0.1277**	**0.0791**	78	view_portrait	**0.1279**	**0.0793**
34	quality_partialblur	**0.1273**	**0.0786**	79	water_lake	**0.1284**	**0.0791**
35	quantity_biggroup	**0.1273**	**0.0792**	80	water_other	**0.1277**	**0.079**
36	quantity_one	**0.1303**	**0.0805**	81	water_riverstream	**0.1275**	**0.0789**
37	quantity_smallgroup	**0.1276**	**0.0794**	82	water_seaocean	**0.1277**	**0.0788**
38	quantity_three	**0.1271**	**0.0789**	83	water_underwater	**0.1266**	**0.0784**
39	quantity_two	**0.1284**	**0.0797**	84	weather_clearsky	**0.127**	**0.0785**
40	relation_coworkers	**0.1275**	**0.0795**	85	weather_cloudysky	**0.1272**	**0.0787**
41	relation_familyfriends	**0.1277**	**0.0793**	86	weather_fogmist	0.1256	0.0776
42	relation_strangers	**0.1282**	**0.0795**	87	weather_lightning	**0.1277**	**0.0794**
43	scape_city	0.1254	0.078	88	weather_overcastsky	**0.127**	**0.0785**
44	scape_coast	**0.1269**	**0.0785**	89	weather_rainbow	**0.1274**	**0.0787**
45	scape_desert	**0.1273**	**0.0787**	90	weather_snowice	**0.1263**	**0.0791**

While the assessment in table 5.2 was done on the whole fuzzy context-based ontology set, we discuss the performance of the proposed framework in the following by considering each ontology separately (the deduction engine is applied for a single context).

In table 5.3, we observe that almost fuzzy context-based ontologies show a semantic enhancement. In term of $\boldsymbol{map}_i$ measurement, our framework performs enhancement up to $3,41\%$ for the context $quantity\_one$, $2,70\%$ for contexts $scape\_mountainhill$ and $fauna\_horse$, $2,54\%$ for the context $lighting\_reflection$, $2,46\%$ for the context $lighting\_shadow$, ... No

enhancement for contexts *age_adult*, *age_child* and *age_elderly*. And a performance decrease for contexts *weather_fogmist* and *scape_city* (respectively $-0,32\%$ and $-0,48\%$).

In term of $\boldsymbol{gmap}_i$ measurement, our framework performs also enhancement up to 3.21% for the context *quantity_one*, 2.69% for the context *scape_mountainhill*, 2.56% for the context *view_indoor*, 2.18% for the context *quantity_two*, 2.05% for the context *gender_female*, ... No enhancement for contexts *age_adult*, *age_child*, *age_elderly* and *scape_city*. And a performance decrease for the context *weather_fogmist* (-0.51%).

When looking to such results, we can conclude that although the promising obtained results through our preliminary experiment, it still some issues to take into consideration. Indeed, the enhancement rate (the maximum obtained was $+5.08\%$) is precious but still weak and do not yet succeed to meet a considerable semantic enrichment expectations. Effectively, obtained result performances are strongly dependent on the knowledge appropriateness and rightness (stored in the constructed fuzzy context-based ontology set). We believe that the knowledge discovery efficiency is strongly related to the learning dataset quality (particularly images selection and annotation process), and the effectiveness of the co-occurrence computing between semantic concepts. By this way, we point out two challenging tasks that we intend to tackle in our future work.

At first, we consider that the *cosine* measure used for computing semantic concepts similarities (co-occurrence) is widely used and approved in information retrieval community. However, other fuzzy functions, and in particular fuzzy similarities measures [Baccour et al. 2011; 2013; 2014] should be attempt in order to enhance more the abduction engine and to deliver more realistic and appropriate semantic concept similarities (rather than the probabilistic *cosine* measure).

Secondly, we consider also that the ontology content of the proposed fuzzy context-based ontologies has to evolve continuously throughout its life cycle in order to be able to answer different change requirements. Indeed, the ontology evolution aims at growing the background knowledge in order to better enrich its semantic capabilities and to validate concepts and their semantic relationships [Gargouri & Jaziri 2010]. Thus, our proposed framework should support content validation through other data sources. Then, erroneous and inappropriate knowledge gathered by the abduction engine will be reviewed and refined through an ontology evolution process.

Table 5.4: ImageClef 2012: Evolving performance evaluation

	map_n		map_i		$gmap_n$		$gmap_i$	
Evolved Ontologies	value	%	value	%	value	%	value	%
View_portrait	0,122	**0.00**	0,1278	**-0.08**	0,0746	**0.27**	0,0793	**0.25**
quantity_one	0,1218	**0.00**	0,1304	**0.08**	0,0748	**0.40**	0,0808	**0.37**

5.3.5 The Ontology Evolving Evaluation

Despite the promising results, the fuzzy ontologies content can be improved through expert's validation. In this section, we discuss how could this validation process improve the accuracy of obtained semantic enhancement. Thus, we applied an evolving process for two contextualized ontologies of the contexts *View_portrait* and *quantity_one*. Two experts inspected and validated about 800 isRelatedTo roles related to these ontologies. Table 5.4 shows obtained results from the evolved ontologies against the original ones.

With a small performance decrease of the context *View_portrait* (-0.08%), the evolving task has slightly increase the performances of the two contextualized ontologies in term of ***gmap***. Indeed, the evolving task gains accurate results for the context *quantity_one*.

With a preliminary experimentation of the evolving task, we can conclude that the ontology evolving could enhance the fuzzy ontologies content, and consequently the semantic interpretation.

5.3.6 Proposed Framework Scalability

The scalability of our proposed framework concerns two aspects: how it performs when a large amount of concepts and interrelationships are defined, and when the number of contexts increases.

Theoretically, our framework can be considered as scalable when looking at the proposed ontology structure. Indeed, the *Tbox* $\mathcal{T}$ remains unchanged, and both abduction and deduction engine are handling only *Abox* $\mathcal{A}$ instances. Then, the amount of knowledge stored in an ontology depends only on what has been gathered from annotated multimedia datasets in addition to the semantic interpretation of the analyzed shots. On the other hand, we proposed a modular ontology architecture (each ontology represents knowledge about one context). Thus, the deduction engine is applied only on *Abox* $\mathcal{A}$ that are related to a defined

context. This leads to an optimized and a lightened reasoning process: only partial defined roles in ontologies will be considered and fired.

To develop our framework, we used the database management system PostgreSql (v. 9.1) and its procedural language PL-pgSql. And, we applied our experiments on a modern desktop machine with a dual-core processor and 4 GB of RAM.

For experiments within *ImageCLEF2012's*, the abduction engine required two CPU cores where their use was balancing between 40% and 80%. This engine tooks $01h : 19min : 14s$ to extract isRelatedTo roles from *ImageCLEF2012's* training dataset for the 90 fuzzy context-based ontologies, at the rate of less than one minute for each ontology population process.

For the deduction engine, one CPU core was used constantly at 100%. Then, we conduct an experiment to investigate this engine scalability. Thus, we applied two deduction operations: the first used 61731 isRelatedTo roles (with fuzzy weight between $0,5$ and 1) and the second used 33952 isRelatedTo roles (with fuzzy weight between $0,7$ and 1). We noted that the performance of the deduction engine gave the same execution rate with the use of the two amounts of isRelatedTo roles:

- for 61731 isRelatedTo roles, the deduction engine took $07h : 31min : 52s$, at the rate of 2,28 fired isRelatedTo roles per a second,
- for 33952 isRelatedTo roles, the deduction engine took $03h : 38min : 44s$, at the rate of 2,36 fired isRelatedTo roles per a second.

This can be interpreted as follows: with doubling isRelatedTo roles, the system kept the same computing rate and load to reasoning with knowledge. Thus, we can conclude that our approach can be considered as scalable.

5.4 Discussion

Despite the promising obtained results by our preliminary evaluations, it still some potential works to be achieved for future improvement. At first, defining ontologies content was based on computing similarities between concepts granted by a large annotated images dataset. More advanced fuzzy similarity functions could be addressed in order to handle real fuzzy knowledge to be inserted in proposed ontologies. Then, ontology evolution task could enhance

significantly ontology semantic capabilities, and then a semantic interpretation. Thus, we believe that ontology evolution is a challenging task that we could follow in our future work.

5.5 Conclusion

In the actual chapter, we presented a fuzzy context-based ontology generation framework for enhancing a semantic interpretation about a multimedia content. The proposed framework deals with an abduction engine for extracting valuable knowledge (contexts, concepts and their relationships) from available data sources, and a deduction engine for inferring new knowledge from extracted ones. Our framework effectiveness and performance were proved on a multimedia benchmark (*ImageCLEF2012's Flickr Photo Annotation and Retrieval*).

In the following chapter, we continue to discuss the scalability of our proposed knowledge-based approach to enhance semantic interpretations. In fact, and in our second contribution, the discussed scalability concerns the large amount of multimedia content to be analyzed and indexed. Our third contribution focuses on the proposed approach scalability but with a large amount of semantic concept to index with.

Chapter 6

Scalable Fuzzy Ontology based Framework for Hierarchical Image Annotation

In this chapter, we propose a scalable ontology-based approach for automatically constructing hierarchical semantic concept detectors for visual contexts, which constitutes our third contribution: C_3. Having discussed fuzzy knowledge management in the previous chapter, we propose now to further explore such valuable knowledge in order to introduce our proposed ontology driven hierarchical semantic structure in order to efficiently train and construct scalable semantic concept detectors.

The rest of this chapter is structured as follows. In Section 6.1, we present the motivations of our proposal. In Section 6.2, we introduce our proposed scalable framework to construct semantic concept detectors. Section 6.3, introduces technical overview for an SVM based concept detector, then discusses the proposed framework scalability and efficiency through the participation within the *ImageClef2015*. Finally, the chapter is concluded in Section 6.4.

6.1 Context and Motivations

As the third contribution C_3, our aim is to construct an automated image annotation framework that focuses on the scalability aspect through reducing semantic concept detection cost and complexity.

Automatic photo annotation is considered as a classification problem that consists in assigning a set of semantic concepts to a semantic content of a given image [C. G. M. Snoek & Worring 2009, D. Zhang et al. 2012, Piras & Giacinto 2014].

Image collections are increasing staggeringly. Thus, retrieving from large-scale image datasets is a challenging task [Villegas et al. 2013, Villegas & Paredes 2014, Gilbert et al. 2015, Villegas et al. 2015]. The access to such considerable contents has forced the multimedia retrieval community to look for advanced approaches and techniques in order to make the availability of automated and efficient semantic annotation for such contents [F. Wang 2011, D. Zhang et al. 2012, Benavent et al. 2013, Sahbi 2013, Reshma et al. 2014].

Some previous works focused on the use of a knowledge based approach. For instance, in [Reshma et al. 2014] an ontology was generated and used both: (1) in training phase to select images that should be used for optimizing classifiers, and (2) in testing phase for deducing new annotations through concept inter-relationships. Seeking to contribute towards this direction, in the previous chapter, we presented a fuzzy ontology based framework for enhancing a multimedia content indexing accuracy. Key dimensions of this inquiry constitute the three main issues addressed by the existing ontologies, namely a generic ontology structure aspect, an automated knowledge extraction process for populating an ontology content, and a machine-driven context detection for a multimedia content. What was accomplished in this study is a novel ontology management method which is intended to a machine-driven knowledge database construction. The experiment that we conducted on the *ImageClef2012* dataset displayed semantic improvements over a classical image annotation framework used in large-scale multimedia contents.

Our approach relies on a visual analysis of the image content. As visual features, we used a *k-means* [Sculley 2010] algorithm to classify training local feature extract by SURF algorithm [Bay et al. 2008]. For scalability, our approach intends to show that we can go further in such aspect by reducing computing cost. In fact, we propose an ontology based approach that alleviates the computing cost for labeling a given image by candidate semantic concepts. By reading some recent papers ([Mller et al. 2010, Villegas et al. 2013, Villegas & Paredes 2014, Cappellato et al. 2015, Villegas et al. 2015] to cite a few), it is obvious that there is a serious focus made on scalability through reducing the candidate concept list to be analyzed within an image content. Mainly, these works rely on dividing candidate concepts

into: (1) initial concepts that can be detected directly through analyzing an image content, and (2) extended concepts that can be detected through reasoning with the initial ones.

In our second contribution C_2 ([Zarka et al. 2016]), we focused on a fuzzy framework for enhancing a semantic interpretation through reasoning with a given initial concept set. Thus, our third contribution C_3 consists in developing a fuzzy ontology to guide the annotation process through reducing the number of concepts to be detected.

6.2 A Scalable Ontology driven Framework for Hierarchical Concept Detection

6.2.1 Framework overview

In this section, we propose a scalable image annotation framework based on hierarchical annotators. We investigate research works on semantic hierarchies for hierarchical image annotation. Our framework relies on constructing and managing a fuzzy ontology that handle a semantic hierarchy. Such a hierarchy is used then to train more accurate image annotators (see figure 6.1).

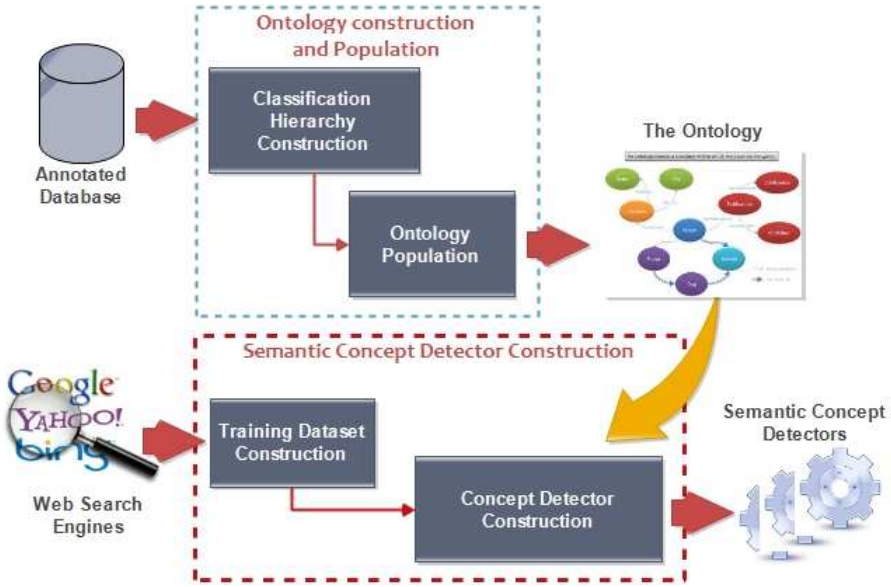


Figure 6.1: Ontology based semantic annotator hierarchy for image annotation

Image annotation is considered as a multi-class classification problem. Many approaches were proposed to handle the annotation scalability aspect (large number of concepts to annotate with) through combining semantic hierarchical structures with classification techniques

(like SVM : Support Vector Machine) [Cevikalp 2010, L.-J. Li et al. 2010, Bannour & Hudelot 2012, McNamara et al. 2015]. Mainly, two different approaches were proposed for constructing the semantic hierarchy. The first one is qualified as top-down method: the semantic hierarchy is built through recursive class set clustering [Cevikalp 2010]. The second one is qualified as bottom-up method: the hierarchy is defined by agglomerative partitioning of the classes [L.-J. Li et al. 2010]. Furthermore, two different approaches were proposed also for hierarchical image classification: the first is the *Binary Hierarchical Decision Trees* (BHDTS)[Cevikalp 2010], and the second is the *Decision Directed Acyclic Graphs* (DDAGS) [Gao & Koller 2011].

Let $C = \{c_1, c_2, \ldots, c_N\}$ be a set of N semantic concept. The DDAGS approach trains $N * (N-1)/2$ binary classifiers and uses a DAG to decide if an image *image* belongs or not to a semantic concept class $c_i \in C$. At each given node at a distance d from the tree root, d semantic concept classes are eliminated, and $N - d$ decision nodes remain to be evaluated. The BHDTS approach handles the semantic hierarchy as a binary tree: concept classes are clustered hierarchically into two subsets. This clustering step is iterated until a single concept class set is reached. For every clustering step, an SVM classifier is trained in order to decide if an image *image* could be annotated by the first or the opposite semantic concept class. A total of $log_2(N)$ SVM classifiers are trained and used for analyzing a test image. Despite the fact that these two approaches enable accurate classifiers, they handle semantic hierarchy as binary structures which requests a considerable structure to handle with large amount of concept classes.

Considering that BHDT approach aims to optimize the SVM classifiers accuracy through reducing the unnecessary comparisons [Cevikalp 2010], we are motivated to use such an approach in our scalable image annotation framework.

In our proposed framework, we aim to define a new method for constructing a hierarchical classifiers for scalable image annotation. At first, an annotated image dataset is analyzed to construct the hierarchy tree for concept classes. Then, and for every level of the defined tree structure, an SVM is trained for predicting if a test image *image* belongs to the first concept class set or the second one. By starting by the first level (root node), the hierarchy is walked until reaching leafs nodes through computing classifier votes (see figure 6.2).

Our method is inspired from fuzzy decision tree based method [X. Wang et al. 2015, Bujnowski et al. 2015] to extract uncertain knowledge in a classification problem. Fuzzy set

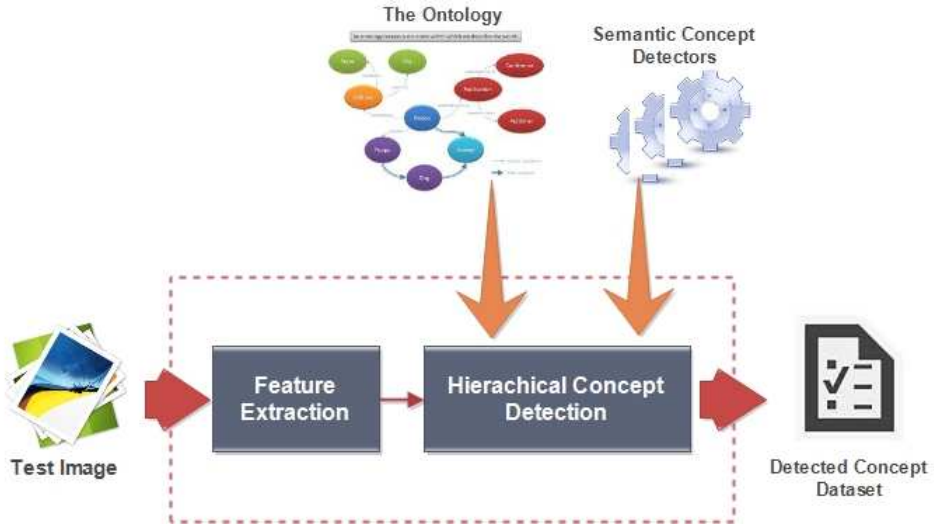


Figure 6.2: Ontology based Hierarchical image classification

theory is used to model the tree structure. Thus, our proposed approach is based on a fuzzy ontology that handles such a decision tree. In what follows, we discuss the structure of our fuzzy ontology, we show how we populate its content, and how to infer available knowledge in order to use the hierarchical classifiers to annotate a test image accurately.

6.2.2 Ontology Structure

The ontology structure is based on three conceptual classes: the semantic concept Concept, the hierarchical node Node, and the test image Image. We define also a set of relationships between these conceptual classes (see table 6.1).

Table 6.1: Semantic Relationships between conceptual classes

Relationships	Definition	Meaning
isIndexedBy	$(\langle \text{Image}, \text{Concept} \rangle : \text{isIndexedBy}) \geq p_1$	The image Image is annotated by the concept Concept by a fuzzy weight p_1
votesFor	$(\langle \text{Node}, \text{Image} \rangle : \text{votesFor}) \geq p_2$	The SVM for the node Node votes for the image image by a fuzzy weight p_2
existsIn	$(\langle \text{Concept}, \text{Node} \rangle : \text{existsIn}) \geq p_3$	The image image exists in the node Node by a fuzzy weight p_3
isChildOf	$(\langle \text{Node}, \text{Node} \rangle : \text{isChildOf})$	The first node Node has a semantic concept subset of the second node Node

The relationship isChildOf depicts that a node $node_1 \in$ Node is a child of another node $node_2 \in$ Node. This relationship is used then for modeling the semantic hierarchy for concept classes.

The relationship existsIn enumerates for each node $node \in$ Node the contained set of concept classes. A concept $concept \in$ Concept can exist in many nodes, but for separate levels.

The relationship votesFor is used when an image *image* is being annotated and the hierarchy is walked from the root node to the leafs. A node $node \in$ Node votes for an image $image \in$ Image by a fuzzy weight p_2 when a SVM classification on that image predicts that the image *image* could be annotated by the set of semantic concepts that exists in the node *node*.

Finally, the relationship isIndexedBy depicts that an image $image \in$ Image is annotated by the concept $concept \in$ Concept by a fuzzy weight equal to p_1.

The proposed ontology structure is used to enable handling the hierarchical classifiers, to trace the hierarchy walk for classifying a given test image, and then to model the set of semantic concepts that annotate that image (see figure 6.3). In what follows, we expose the population process for our ontology, then, we discuss the reasoning process used to guide and assist the hierarchical annotation.

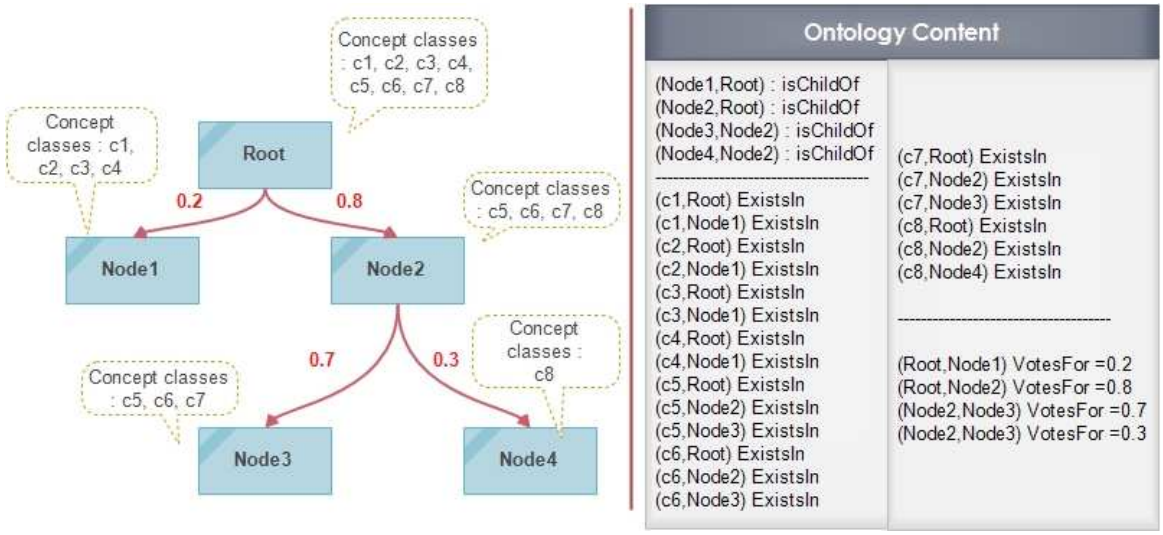


Figure 6.3: Ontological Hierarchy content for image annotation

6.2.3 Ontology population

Given a defined set of semantic concepts, we start by clustering it through analyzing annotated image dataset provided by the *ImageCLEF 2015 Scalable Concept Image Annotation* task.

At first, we apply a binary clustering for the whole concept set, and we define two new nodes in the ontology $node_1$ and $node_2$. We use a *k-means* clustering algorithm with $k = 2$. Then, each concept is instantiated within the ontology, and for every concept *concept* that belongs to the node *node*, a new relationship existsIn is instantiated between *concept* and *node*. This process is recursively called on $node_1$ and $node_2$ until a sub-node contains only one semantic concept class, or the clustering process seems unable to cluster a given semantic concept classes. At each iteration, the new defined nodes are populated within the ontology through instantiating the isChildOf relationships.

6.2.4 Hierarchical classificators construction

Once the hierarchical structure is defined through the above mentioned recursive binary clustering, an SVM based classifier is trained for all the nodes that belong to the same level. As training images, we select some development images for every concept that belongs to a node. In section 6.3.3, we detail the development image dataset used for the training task.

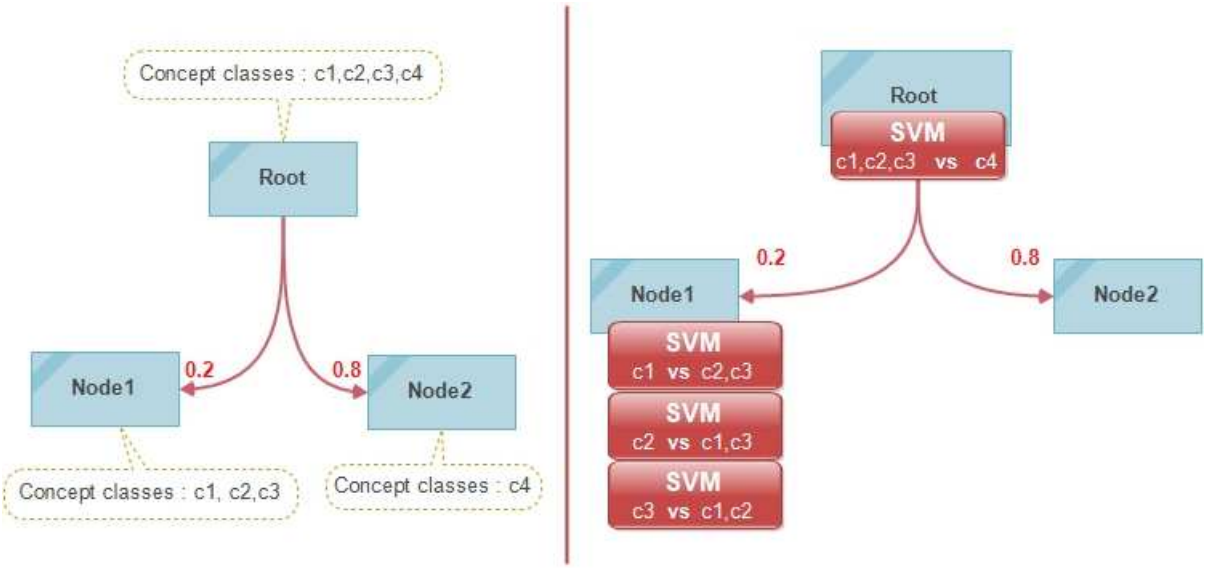


Figure 6.4: Hierarchical SVM classificator construction

At a given level, two possible nodes are figuring (see figure 6.4). By exploring existsIn relationships, we construct a training image dataset. For the first node (see node *Root* in figure 6.4), and for each concept that belongs to that node, a subset of images that are annotated by this concept are selected to be training images for corresponding node.

For a leaf node (see node *Node1* in figure 6.4), we proceed as follow: let $C_m = \{c_1, c_2, \ldots, c_k\}$ be a set of k concepts that belongs to the node $node_m$. We construct then k classifiers. Each classifier is related to a given concept and trained against the other concepts. Then, and for a classifier f of a concept $c_f \in C_m$, we train an SVM classifier based on two image sets: the first set is based on images that are annotated by the concept c_f, and the other set is based on images that are annotated by the other concepts $(C_m \setminus c_f)$.

For a leaf node that contains only one concept class (see node *Node2* in figure 6.4), no SVM classifier will be constructed. And an image annotation for this concept will be computed through the leaf node classification vote.

6.2.5 Reasoning

We start reasoning from the root node (top node) of the constructed fuzzy tree (see figure 6.3). For a given node, we compute the values of the membership functions (μ) for the child nodes through firing the corresponding SVM classifiers. The classification results (the vote) are populated into the fuzzy ontology through instantiating the votesFor relationship.

In order to improve reasoning accuracy and to minimize the decision tree walk (which will also minimize the number of SVM classifiers to be fired), we define a *Fuzziness control threshold* $\theta_r = 0.1$: given two sub-nodes $node_1$ and $node_2$, firing the SVM classifier at this level provides two membership function values μ_1 for $node_1$ and μ_2 for $node_2$. Then, we compute $\theta_r = |\mu_1 - \mu_2|$.

if $\theta_r \leq 0.1$, then we could not be sure if the SVM classifier is discriminative to judge if the content of a test image belongs to the first or to the second node. We proceed so to walk both sub-nodes ($node_1$ and $node_2$). For the opposite case ($\theta_r > 0.1$), the reasoner walks only the node that has the greater membership function value (μ).

Given the example in figure 6.3, the SVM classifier of the node *root* computed $\mu_1 = 0.2$ for the node $Node1$, and $\mu_2 = 0.8$ for the node $Node2$. Then, the reasoning algorithm stops

walking the node $Node1$ and proceeds to walk the $Node2$ since $\theta_r = 0.8 - 0.2 = 0.6$ and $0.1 \leq 0.6$.

A leaf node can contain a set of concept classes, or only one concept class. In the first case, and for every contained concept class, an SVM classifier is fired for that concept against the other contained concept classes. The classification result is populated in the ontology through the instantiation of the relationship isIndexedBy between the concept class and the test image. The fuzzy weight for the new relationship is computed as an average of μ values computed from the root node to the leaf one. In case of a single concept class, a new isIndexedBy relationship is instantiated within the ontology between that concept and the test image. The fuzzy weight is computed as in the first case.

Our proposed fuzzy decision tree reasoner assists the annotation of a given test image through firing recursive trained SVM classifiers in order to optimize the number of concept to be detected. Such an optimization should reduce also the computing cost of a given test image annotation process.

In the next section, we expose how we construct an SVM classifier for each node in the constructed fuzzy hierarchical semantic structure of concept classes.

6.3 Experiments and Results

In this section, we discuss the obtained experimental results from our participation in the *Imageclef 2015* (within the *Scalable Concept Image Annotation* task). In such an experiment, we look particularly in the assessment of our approach scalability, rather than the semantic enhancement.

In the rest of this section, we start by presenting the used dataset and metrics for the evaluation, we, then, show how we build hierarchical concept detectors in accordance with what was presented in the previous section. And finally, we discuss the obtained results.

6.3.1 Datasets Description

The image dataset provided by the *ImageClef 2015* evaluation campaign is constructed through querying popular search engines (mainly GOOGLE, BING and YAHOO!). A total of 500 000 images were gathered [Gilbert et al., 2015].

The test dataset contains all the defined 500 000 images. And the development dataset contains 5 520 annotated images taken from the test dataset. A total of 251 semantic concepts where used to annotated the content of the development set of images.

For each image in the dataset, a full text description of the image content is provided through extracting text content from the web-page where the image is located.

6.3.2 Evaluation metrics

In *imageclef 2015 Scalable Concept Image Annotation* task, the annotation accuracy is evaluated using the MAP metric.

Another metric is used: the PASCAL VOC [Everingham et al. 2015]. But the latter evaluates not only the annotation accuracy, but also the semantic concept localization. Since our approach desn't handle yet such an ability, we consider only the MAP metric.

6.3.3 Svm Classifier Construction

In our participation within *ImageCLEF 2015 Scalable Concept Image Annotation* task, we aimed basically to evaluate the scalability aspect of our preliminary automatic annotation framework. For semantic concept detector/annotator, we have not really defined an original approach, but we implemented state-of-the-art bags of quantized local features and linear classifiers learned by support vector machines. In fact, and as pointed in Piras & Giacinto [2014], bag-of-features and codebook approach has gained a great attention by image classification and annotation community as it showed notable semantic accuracy [Jurie & Triggs 2005, Van Gemert et al. 2008, Mylonas et al. 2009, Ngo et al. 2010, Grana et al. 2013, Hidaka et al. 2013, Kanehira et al. 2014, Xu et al. 2014, Elleuch et al. 2015]. In what follow, we expose how we construct SVM classifiers for semantic concept detection and annotation.

Construct a learning dataset

Image annotation has always been heavily dependent on good development datasets. First, datasets were mainly hand-collected. However, and recently, several researches attempt to automate such a laborious task. Re-ranking images gathered from popular Image search engines (GOOGLE, YAHOO!, BING, ...) can construct automatically an image learning dataset [Fergus et al. 2004; 2005, Schroff et al. 2011].

As a development dataset, we have not used one provided by the *ImageCLEF 2015 Scalable Concept Image Annotation task*. In fact, not all the concepts were annotated. We relied then on FLICKR image search engine to obtain image set and construct a learning dataset. We used so the information provided with concept list to query the search engine and we gathered first 100 result images for each given concept.

At the outset, it seems to be curious to use an external data source as a development dataset. Our aim is to explore available on-line data-sources (like search engines) to train non annotated semantic concepts.

Local Feature Extraction

Our framework extracts feature from an input image through a robust local feature extractor. We followed a basic and state-of-the-art framework for such purpose (as described in [Piras & Giacinto 2014]). Leading extractors for such a purpose include *Scale Invariant Feature Transform* (SIFT) and *Speeded Up Robust Features* (SURF). Local feature descriptors handle a pixel within an image by analyzing its neighborhood pixels. Many different descriptors and interest-point detectors were proposed and discussed in the literature. While the SIFT descriptor [Lowe 2004] is considered as the most widely used descriptor, SURF [Bay et al. 2008] is known as robust local feature extraction to various image perturbations.

Our framework extracts local features and descriptors using SURF. Such a choice is argued by SURF concise descriptor length (64 floating point values). The SURF implementation that we used is provided by OPENCV [Bradski & Kaehler 2008].

For query image analysis, local features are extracted and mapped into nearest computed cluster centroids. The query image is then handled by a vector that represents defined visual bag-of-words.

Classification of local Features and Constructing the bag-of-words model

After extracting local features, a bag-of-words model is used to represent these descriptors. The latters are extracted from training images and are grouped into N clusters of visual words using *k-means*. Each defined descriptor is classified into its cluster centroid by computing the *Euclidean distance* metric. For our runs, we choose a value of $N = 100$. This value is argued by a balance between high bias (under-fitting) and high variance (over-fitting).

In order to alleviate the computing cost of *k-means* clustering, we used *Mini Batch k-means* Sculley [2010] as an alternative to the *k-means* algorithm for clustering massive datasets. *Mini Batch k-means* reduces the computational cost by handling fixed size subsample instead of all the data in the database. This strategy reduces the amount of distance to be computed at each clustering iteration.

Learning Algorithm

The learning algorithm consists in training one-vs.-one linear SVM to operate in the bag of SURF feature space. Training images are classified through a histogram vector conctructed in the *k-means* based clustering. We used a linear kernel for our SVM based learning algorithm in view of its simplicity and computational efficiency in training and classification: $K(x, y) = x^T y + c$.

Basically, SVM are binary classifier. For a given detector, an image is annotated by one of two distinct groups. A one-vs.-one scheme is used in which each SVM trained for each combination of individual classes. The SVM implementation used in our runs is given by SCIKIT-LEARN library [Pedregosa et al. 2011].

Decision

As an SVM decision function, a class membership probability estimation fits the decision values.

SCIKIT-LEARN library uses a PLATT SCALING in order to calibrate the SVM classifier to produce, in addition to class predictions, probabilities. When the SVM is trained, an optimization process is called to optimize parameter vectors A and B such that: $P(y|X) = 1/(1+exp(A*f(X)+B))$ where $f(X)$ is the signed distance of a sample from the hyperplane.

6.3.4 Experiments with *ImageClef 2015*

Our goal behind our participation in *ImageClef 2012* is to assess the proposed ontology based hierarchical concept detection scalability. However, and unlike the scalability assessment presented in the previous chapter (where we tested the scalability of the system to handle a large-scale image dataset), we aim to test if our approach could be scalable when the number of semantic concepts increases. Our idea behind the ontology based hierarchical concept

detection consists in filtering the concepts to be detected: not all the concept detectors will be called in the detection process.

Image Semantic Annotation Evaluation

For the image annotation, we only annotated only 300 000 images of the 500 000 images provided in the test dataset. Furthermore, we used just low resolution images (thunbnail) instead of the full resolution images. In fact, the large amount of images to be annotated has forced us to reduce the quality of processed images in the aim to accelerate the annotation computational cost. Yet, we haven't proceeded to annotate the full image dataset (only 300.000 images).

The tables 6.2 and 6.3 display the obtained results in terms of semantic annotation accuracy. We think that a full size image dataset annotation with more tweaked SVM classifiers should give better results. Furthermore, fuzzy ontology based semantic enhancement (described in the previous chapter [Zarka et al. 2016]) should also enhance our framework annotation accuracy.

Table 6.2: IMAGECLEF 2015: $MAP\_0\_Overlap$ Runs evaluation

	MAP	
Best run	0, 795403	(/SMIVA/21.run)
Worst run	0, 0305398	
Average	0, 31046	
Our best run	0, 0366072	(position 85/89)

Table 6.3: IMAGECLEF 2015: $MAP\_0.5\_Overlap$ Runs evaluation

	MAP	
Best run	0, 659507	(/SMIVA/21.run)
Worst run	0, 000231898	(
Average	0, 18673	
Our best run	0, 0161687	(position 75/89)

As illustrated in tables 6.2 and 6.3, the system SMIVA [Pravin et al. 2015] showed great annotation performance. This system annotates images not only in the basis of its visual content, but also through analyzing textual information in the web page where the analyzed image is localized. For our evaluation, we only used visual analysis.

Scalability Evaluation

We would remind that our objective is to alleviate the computation cost of the image annotation through a proposed framework that is based on an ontology reasoning driven hierarchical concept detector.

For the training process, we trained the SVM classifiers in about 100 hours (we used 100 learning images per a concept). This task was executed on a modern machine (Intel $i5$ processor with 16 GB RAM memory).

The annotation task was done on 10 machines (each one has one core CPU and 1 GB of RAM). The annotation of 300 000 images elapsed about 1 633 hours (without taking into consideration the VPS parallel computing).

Our framework annotates a test image with an average of 19.615 seconds (the maximum record was 597.250 seconds and the minimum one was 0.066 second). And for a given test image, an average of 52 SVM classifiers were fired (the maximum was 175 and the minimum was 6). Our framework has reduced the number of SVM classifiers to be fired in order to annotate a given test image.

We can conclude that our proposed framework reduced the number of requested semantic concept detector to annotate an image: from 251 concept detectors, to an average of 52 ones. Then, the time required to annotate an image is decreased by 80%

6.4 Conclusion

In this chapter note, we described our annotation framework for scalable ontology driven semantic image annotation. We discussed our ontology based framework for reducing the number of concepts to be detected for a given image. We developed a state-of-the art bag-of-words based concept detector (that uses SURF feature extractor and *k-means* classification). Then, concept detectors are selected through reasoning with a fuzzy ontology content. Thus, not all the concept detectors are used for a given image.

The main discussed contribution was to propose a scalable framework for multimedia content analysis. We focused then on the use of an ontology in order to construct fast semantic concept detectors. Although the promising obtained results, we believe that the implementation of these constructed detectors within *Hadoop/MapReduce* and even *GPU*

based runtime-environment (like CUDA, OPENCL [Mahmoudi & Manneback 2015, Osipyan et al. 2015, Dantas et al. 2015, Peters & Savakis 2015] could enable more efficiency for multimedia content analysis.

Part III

Conclusions and Future Research Directions

Chapter 7

Conclusions and Perspectives

The general framework of this dissertation is video information retrieval. The main tackled challenge is how to enable a user to easily access and interpret video contents. Particularly, our thesis works are focused on a contextual concept-based video indexing that represents a motivating solution to such a challenge. The major inherent difficulty in this task is the semantic gap: what separates the signal representation from semantics (concept, contexts and relationships).

In multimedia indexing and retrieval literature, the multimedia community primarily focused on concept model building through a supervised learning technique that analyzes the extracted low-level features: representative features vectors are extracted from a set of representative manually annotated data (commonly images), then used for a supervised learning algorithm in order to approach the semantic interpretation. Nevertheless, such an approach partially solved the semantic gap problem: it is difficult to detect efficiently a variety of concepts, and still unable to detect implicit objects (semantic concepts that do not figure in the content, or that cannot be easily detected). On the one hand, such issues are induced by the large variety of semantics that could be handled, on the other hand, the inability of these approaches taking into consideration semantic concept relationships. Thus, recent research works are focusing on using ontologies for multimedia retrieval in order to allow semantic interpretation and reasoning over extracted descriptions. However, much remains to be done in order to achieve less human aid ontology modeling approaches.

7.1 Summary of Contributions

Our thesis work proposed a generic and scalable fuzzy knowledge based framework to enhance concept-based multimedia indexing. In this context, our contributions aim to improve the accuracy, the scalability and the generalization capability of our semantic video indexing system: RegimVid. These novelties are enumerated as follows: (1) a new knowledge based model for multimedia indexing. The objective is to develop a framework able to handle various information about a multimedia content, then to operate with this information in order to infer new information/knowledge through a reasoning process. Such a novel model has to define and highlight pertinent components for an efficient knowledge based indexing process. (2) an ontology based framework to handle fuzzy knowledge and reason with semantic interpretations in order to enhance and enrich them. This objective aims to define a semantic structure to model required knowledge, then to specify an automated ontology population from available annotated image/video datasets, and finally to handle ontology content evolving to further improve semantics capabilities through analyzing and revising inaccurate and irrelevant knowledge. (3) an approach for an automatic multimedia indexing by the use of an ontology-based semantic hierarchy handled at both learning and annotation steps.

While recent works focused on the use of semantic hierarchies to improve concept detector accuracy, this objective means the use of such hierarchies to reduce detector complexity and then, to handle efficiently large-scale multimedia datasets.

7.2 Future Research Directions

Many approaches were proposed in this dissertation in order to deal with video indexing issues. Indeed, the semantic gap problem remains an open problem and could not be solved in the near future. Therefore, research works on video indexing are witnessing an incremental improvement, and many supplementary efforts are still needed.

In the following, we discuss some potential future directions that could be explored further over the described achievements throughout this dissertation.

Fuzzy Similarities Defining ontologies content was based on computing similarities between concepts granted by a large annotated images dataset. Although we used the

statistical *cosine* similarity function, more advanced fuzzy similarity functions [Bacour et al. 2013; 2014] could be addressed in order to handle real fuzzy knowledge to be inserted in proposed ontologies.

Video genre We proposed a knowledge structure that enhances the accuracy of a semantic interpretation. Nevertheless, when analyzing a video annotated dataset, all videos are handled as if they were addressing the same subject. Furthermore, the videos can be classified according to their genres. Thus, the relationships between concepts/contexts can differ from one genre to another, and subsequently, from one video to another. In literature, the video classification is well-addressed ([J. Wu & Worring 2012, Huang & Wang 2012, Chattopadhyay & Maurya 2013, Muneesawang et al. 2014] to cite a few). We are convinced that addressing the video genre as a semantic information can further refine the extracted fuzzy relationships, and consequently, the video semantic interpretation enhancement. Thus, we consider that it could be an interesting task that should be followed for a future work.

Big data video analysis Many research works considers deep neural networks as a powerful framework for big data video repositories [Girshick et al. 2014, Jiang 2015, Sainath et al. 2015, Tong et al. 2015]. In [Q. Wu et al. 2015, Druzhkov & Kustikova 2016], a comprehensive survey on deep learning and neural networks based approaches for video analysis and indexing is discussed. The effectiveness of such approaches resides in exploring parallel processing units (such as CUDA based machines [Osipyan et al. 2015]) to run the neural networks pipelines. However, and as discussed in section 3.4, we are focusing more in our thesis work on a knowledge-based approach rather than a parallel one.

As a third contribution (discussed in chapter 6), we showed that a knowledge-based approach could reduce the computing cost of semantic concept detectors. But, we believe that this contribution should be extended to deep neural networks in order to achieve robust and real time multimedia content analysis.

Domain Adaptation and Ontology evolving Ontology evolving is an important step in an ontology construction. In fact, this step ensures not only the accuracy of the managed knowledge, but leads also to a continuous adaptation of an ontology to a variety

of application domains. The latter is based on the use of particular approaches for constructing a discriminative model within a shift between observed data (training) and analyzed one (test). Such a situation is widely discussed particularly when dealing with big data where information is diverse and heterogeneous, and defined as *domain adaptation* [Pan & Yang 2010]. Many approaches were discussed in literature for the domain adaptation mainly by bridging the learned (observed) and target (analyzed) domains by learning domain invariant features : the classifier learned from observed domain can then be applied to the target domain [Long et al. 2016].

Recent studies are focusing particularly on deep neural networks to handle invariant features for domain adaptation [Donahue et al. 2014, Yosinski et al. 2014, S. Kumar et al. 2016]. Indeed, the domain adaptation is embedded in the pipeline of deep feature learning in order to extract domain invariant representation [Tzeng et al. 2014, Long & Wang 2015].

In multimedia analysis, the domain adaptation leads to alleviate manual construction of labeled data, and to enhance the ability to handle extended domains [Saenko et al. 2010, Gopalan et al. 2011, Gong et al. 2012, Duan et al. 2012, Hoffman et al. 2014, G. Liu et al. 2016]. Some recent works addressed the domain adaptation with more focus on deep neural networks [Ganin & Lempitsky 2015, Tzeng et al. 2015].

Thus, the domain adaptation is an interesting research direction that should be tackled when dealing with ontology population and evolving. In the present dissertation, we opted to use classical approaches to extract and evolve the content of an ontology. Thereby, we believe that such a research direction will be highly considered as a future work.

Bibliography

Ai, L.-f., Yu, J.-q., He, Y.-f., & Guan, T. (2013). High-dimensional indexing technologies for large scale content-based image retrieval: a review. *Journal of Zhejiang University SCIENCE*, *14*(7), 505–520.

Alimi, A., Hassine, R., & Selmi, M. (2000). Beta fuzzy logic systems: approximation properties in the siso case. *International Journal of Applied Mathematics and Computer Science*, *10*(4), 857–875.

Amit, G., Caspi, Y., Vitale, R., & Pinhas, A. (2006, July). Scalability of multimedia applications on next-generation processors. In *Multimedia and Expo, 2006 IEEE international conference on* (p. 17-20).

Antani, S., Kasturi, R., & Jain, R. (2002). A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern Recognition*, *35*(4), 945 - 965.

Aouiti, C., Alimi, A. M., Karray, F., & Maalej, A. (2003). Evolutionary approach for the beta function based fuzzy systems. In *The 12th IEEE international conference on fuzzy systems, FUZZ-IEEE 2003, St. Louis, Missouri, USA, 25-28 may 2003* (pp. 179–184).

Arndt, R., Troncy, R., Staab, S., Hardman, L., & Vacura, M. (2007). Comm: Designing a well-founded multimedia ontology for the web. In K. Aberer et al. (Eds.), *The semantic web* (Vol. 4825, p. 30-43). Springer Berlin Heidelberg.

Atif, J., Hudelot, C., & Bloch, I. (2014, May). Explanatory reasoning for image understanding using formal concept analysis and description logics. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, *44*(5), 552-570.

Atrey, P., Hossain, M., El Saddik, A., & Kankanhalli, M. (2010). Multimodal fusion for multimedia analysis: a survey. *Multimedia Systems (Springer)*, *16*(6), 345-379.

Ayache, S., Quenot, G., & Gensel, J. (2007). Image and video indexing using networks of operators. *Hindawi - Journal of Image Video Process*(4), 1–13.

Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., & Patel-Schneider, P. F. (Eds.). (2003). *The description logic handbook: Theory, implementation, and applications.* New York, NY, USA: Cambridge University Press.

Baader, F., & Peñaloza, R. (2011). On the undecidability of fuzzy description logics with GCIs and product t-norm. In C. Tinelli & V. Sofronie-Stokkermans (Eds.), *Frontiers of combining systems* (Vol. 6989, p. 55-70). Springer Berlin Heidelberg.

Baccour, L., Alimi, A., & John, R. (2011, June). Relationship between intuitionistic fuzzy similarity measures. In *Fuzzy Systems (Fuzz), 2011 IEEE International Conference on* (p. 971-975).

Baccour, L., Alimi, A. M., & John, R. I. (2013). Similarity measures for intuitionistic fuzzy sets: State of the art. *Journal of Intelligent and Fuzzy Systems*, *24*(1), 37–49.

Baccour, L., Alimi, A. M., & John, R. I. (2014). Some notes on fuzzy similarity measures and application to classification of shapes, recognition of arabic sentences and mosaic. *IAENG International Journal of Computer Science*, *41*(2), 81–90.

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern information retrieval: The concepts and technology behind search.* Addison Wesley.

Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval* (Vol. 463). ACM press New York.

Bahmanyar, R., Murillo Montes de Oca, A., & Datcu, M. (2015, Oct). The semantic gap: An exploration of user and computer perspectives in earth observation images. *Geoscience and Remote Sensing Letters, IEEE*, *12*(10), 2046-2050.

Bannour, H., & Hudelot, C. (2011, June). Towards ontologies for image interpretation and annotation. In *Content-based multimedia indexing (CBMI), 2011 9th international workshop on* (p. 211-216).

Bannour, H., & Hudelot, C. (2012). Hierarchical image annotation using semantic hierarchies. In *Proceedings of the 21st acm international conference on information and knowledge management* (pp. 2431–2434). New York, NY, USA: ACM.

Bannour, H., & Hudelot, C. (2014). Building and using fuzzy multimedia ontologies for semantic image annotation. *Springer - Multimedia Tools and Applications*, *72*(3), 2107–2141.

Baroffio, L., Cesana, M., Redondi, A., Tubaro, S., & Tagliasacchi, M. (2013). Coding video sequences of visual features. In *Image processing (ICIP), 2013 20th IEEE international conference on* (pp. 1895–1899).

Bay, H., Ess, A., Tuytelaars, T., & Gool, L. V. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding*, *110*(3), 346 - 359. (Similarity Matching in Computer Vision and Multimedia)

Benavent, X., Castellanos, A., de Ves, E., Hernández-Aranda, D., Granados, R., & García-Serrano, A. (2013). A multimedia ir-based system for the photo annotation task at imageclef2013. In *Working notes for CLEF 2013 conference , Valencia, Spain, september 23-26, 2013.*

Bengio, Y. (2009). Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, *2*(1), 1–127.

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., et al. (2007). Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, *19*, 153.

Berners-Lee, T., Hendler, J., Lassila, O., et al. (2001). The semantic web. *Scientific american*, *284*(5), 28–37.

Blasch, E., Kadar, I., Salerno, J. S., Kokar, M. M., Das, S., Powell, G. M., ... Ruspini, E. H. (2006). Issues and challenges in situation assessment (level 2 fusion). *Journal of advances in information fusion*, *1*(2), 122–139.

Bloch, I. (2005). Fuzzy spatial relationships for image processing and interpretation: a review. *Image and Vision Computing*, *23*(2), 89 - 110. (Discrete Geometry for Computer Imagery)

Bosko, B. (1990). Fuzziness vs. probability. *International Journal of General Systems*, *17*(2-3), 211-240.

Bradski, D. G. R., & Kaehler, A. (2008). *Learning Opencv, 1st edition* (First ed.). O'Reilly Media, Inc.

Brilhault, A. (2009). *Indexation et recherche par le contenu de documents vidéos* (Tech. Rep.). Joseph Fourier University.

Brunelli, R., Mich, O., & Modena, C. M. (1999). A survey on the automatic indexing of video data. *Journal of visual communication and image representation*, *10*(2), 78–112.

Buckland, M. K., & Gey, F. C. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science (JASIS)*, *45*(1), 12–19.

Bujnowski, P., Szmidt, E., & Kacprzyk, J. (2015). Intuitionistic fuzzy decision tree: A new classifier. In P. Angelov et al. (Eds.), *Intelligent systems'2014* (Vol. 322, p. 779-790). Springer International Publishing.

Cao, Y., Wang, C., Zhang, L., & Zhang, L. (2011). Edgel index for large-scale sketch-based image search. In *Computer vision and pattern recognition (CVPR), 2011 IEEE conference on* (pp. 761–768).

Cao, Y., Wang, H., Wang, C., Li, Z., Zhang, L., & Zhang, L. (2010). Mindfinder: interactive sketch-based image search on millions of images. In *Proceedings of the international conference on multimedia* (pp. 1605–1608).

Cappellato, L., Ferro, N., Jones, G., & San Juan, E. (Eds.). (2015). *Clef 2015 labs and workshops, notebook papers* (No. 994).

Caputo, B., Müller, H., Martinez-Gomez, J., Villegas, M., Acar, B., Patricia, N., ... Morell, V. (2014). Imageclef 2014: Overview and analysis of the results. In E. Kanoulas et al. (Eds.), *Information access evaluation. multilinguality, multimodality, and interaction* (Vol. 8685, p. 192-211). Springer International Publishing.

Cerami, M., & Straccia, U. (2013). On the (un) decidability of fuzzy description logics under łukasiewicz t-norm. *Information Sciences*, *227*, 1–21.

Cevikalp, H. (2010). New clustering algorithms for the support vector machine based hierarchical classification. *Pattern Recognition Letters*, *31*(11), 1285 - 1291.

Chaira, T., & Ray, A. (2005). Fuzzy measures for color image retrieval. *Fuzzy Sets and Systems*, *150*(3), 545 - 560.

Chang, N.-S., & Fu, K.-S. (1980, November). Query-by-pictorial-example. *IEEE Trans. Softw. Eng.*, *6*(6), 519–524.

Chattopadhyay, C., & Maurya, A. (2013). Genre-specific modeling of visual features for efficient content based video shot classification and retrieval. *International Journal of Multimedia Information Retrieval*, *2*(4), 289-297.

Chen, P.-I., Lin, S.-J., & Chu, Y.-C. (2011). Using google latent semantic distance to extract the most relevant information. *Expert Systems with Applications*, *38*(6), 7349 - 7358.

Chen, W.-N., & Hang, H.-M. (2008). H. 264/avc motion estimation implmentation on compute unified device architecture (cuda). In *Multimedia and expo, 2008 IEEE international conference on* (pp. 697–700).

Chen, Y., & Wang, J. (2002, Sep). A region-based fuzzy feature matching approach to content-based image retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *24*(9), 1252-1267.

Cheng, Y., & Xiong, Y. (2012). Research on model of ontology-based semantic information retrieval. In D. Jin & S. Lin (Eds.), *Advances in multimedia, software engineering and computing vol.1* (Vol. 128, p. 271-276). Springer Berlin Heidelberg.

Cioara, T., Anghel, I., Salomie, I., & Dinsoreanu, M. (2009). A context - based semantically enhanced information retrieval model. In *Intelligent computer communication and processing, 2009. ICCP 2009. IEEE 5th international conference on* (p. 245-250).

Ciregan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *Computer vision and pattern recognition (cvpr), 2012 ieee conference on* (pp. 3642–3649).

Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search engines: Information retrieval in practice*. Addison-Wesley Reading.

Cross, V. (1994). Fuzzy information retrieval. *Journal of Intelligent Information Systems*, *3*(1), 29-56.

Dantas, D. O., Danilo Passos Leal, H., & Sousa, D. O. B. (2015). Fast 2d and 3d image processing with opencl. In *Image processing (ICIP), 2015 IEEE international conference on* (pp. 4858–4862).

Darwish, S. M., & Ali, R. A. (2015). Observations on using type-2 fuzzy logic for reducing semantic gap in content–based image retrieval system. *International Journal of Computer Theory and Engineering*, *7*(1), 1.

Dasiopoulou, S., Giannakidou, E., Litos, G., Malasioti, P., & Kompatsiaris, Y. (2011). A survey of semantic image and video annotation tools. In G. Paliouras, C. Spyropoulos, & G. Tsatsaronis (Eds.), *Knowledge-driven multimedia information extraction and ontology evolution* (Vol. 6050, p. 196-239). Springer Berlin Heidelberg.

Dasiopoulou, S., & Kompatsiaris, I. (2010). Trends and issues in description logics frameworks for image interpretation. In S. Konstantopoulos, S. Perantonis, V. Karkaletsis, C. Spyropoulos, & G. Vouros (Eds.), *Artificial intelligence: Theories, models and applications* (Vol. 6040, p. 61-70). Springer Berlin Heidelberg.

Dasiopoulou, S., Kompatsiaris, I., & Strintzis, M. G. (2008). Using fuzzy dls to enhance semantic image analysis. In *Semantic multimedia* (pp. 31–46). Springer.

Dasiopoulou, S., Kompatsiaris, I., & Strintzis, M. G. (2009). Applying fuzzy dls in the extraction of image semantics. In *Journal on data semantics xiv* (pp. 105–132). Springer.

Dasiopoulou, S., Mezaris, V., Kompatsiaris, I., Papastathis, V.-K., & Strintzis, M. (2005, Oct). Knowledge-assisted semantic video object detection. *Circuits and Systems for Video Technology, IEEE Transactions on*, *15*(10), 1210-1224.

Dasiopoulou, S., Tzouvaras, V., Kompatsiaris, I., & Strintzis, M. (2009). Capturing mpeg-7 semantics. In M.-A. Sicilia & M. Lytras (Eds.), *Metadata and semantics* (p. 113-122). Springer US.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, *40*(2), 5.

Datta, R., Li, J., & Wang, J. Z. (2005). Content-based image retrieval: Approaches and trends of the new age. In *Proceedings of the 7th ACM SIGMM international workshop on multimedia information retrieval* (pp. 253–262). New York, NY, USA: ACM.

Davies, J., Sure, Y., Vrandecic, D., Pinto, S., Tempich, C., & Sure, Y. (2005). The DILIGENT knowledge processes. *Journal of Knowledge Management*, *9*(5), 85–96.

Dean, J., & Ghemawat, S. (2008). Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, *51*(1), 107–113.

Dean, J., & Ghemawat, S. (2010). Mapreduce: a flexible data processing tool. *Communications of the ACM*, *53*(1), 72–77.

Deb, S. (2004). *Multimedia systems and content-based image retrieval*. Idea Group Publishing.

De Mantaras, R. L., Godo, L., Plaza, E., & Sierra, C. (2015). A survey of contributions to fuzzy logic and its applications to artificial intelligence at the IIIA. In *Enric trillas: A passion for fuzzy sets* (pp. 67–78). Springer.

Deng, J., Berg, A., Li, K., & Fei-Fei, L. (2010). What does classifying more than 10,000 image categories tell us? In K. Daniilidis, P. Maragos, & N. Paragios (Eds.), *Computer vision – ECCV 2010* (Vol. 6315, p. 71-84). Springer Berlin Heidelberg.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on* (pp. 248–255).

Dentler, K., Cornet, R., ten Teije, A., & de Keizer, N. (2011). Comparison of reasoners for large ontologies in the OWL 2 EL profile. *Semantic Web*, *2*(2), 71–87.

de Ves, E., Ayala, G., Benavent, X., Domingo, J., & Dura, E. (2015). Modeling user preferences in content-based image retrieval: A novel attempt to bridge the semantic gap. *Neurocomputing*, *168*, 829 - 845.

Dingli, A. (2011). *Knowledge annotation: Making implicit knowledge explicit*. Springer Berlin Heidelberg.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In *Icml* (pp. 647–655).

Dou, D., Wang, H., & Liu, H. (2015, Feb). Semantic data mining: A survey of ontology-based approaches. In *Semantic computing (ICSC), 2015 IEEE international conference on* (p. 244-251).

Douze, M., Jégou, H., Schmid, C., & Pérez, P. (2010). Compact video description for copy detection with precise temporal alignment. In *Computer vision–ECCV 2010* (pp. 522–535). Springer.

Druzhkov, P. N., & Kustikova, V. D. (2016). A survey of deep learning methods and software tools for image classification and object detection. *Pattern Recognition and Image Analysis*, *26*(1), 9–15.

Duan, L., Xu, D., Tsang, I. W.-H., & Luo, J. (2012). Visual event recognition in videos by learning from web data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(9), 1667–1680.

Dumitrescu, A., & Santini, S. (2009). Context based semantics for multimodal retrieval. In *Is&t/spie electronic imaging* (pp. 72550C–72550C).

Egozi, O., Markovitch, S., & Gabrilovich, E. (2011). Concept-based information retrieval using explicit semantic analysis. *ACM Transactions on Information Systems (TOIS)*, *29*(2), 8:1–8:34.

Elgesem, D., & Nordbotten, J. C. (2007). The role of context in image interpretation. In *Proceedings of the CIR'07 workshop on context-based information retrieval in conjunction with context-07, roskilde, denmark, 20 august 2007* (Vol. 326). CEUR-WS.org.

Elleuch, N. (2015). *A generic visual video indexing framework: Semantic concepts detection based on contexts reasonning (svi_regimvid system)* (Unpublished doctoral dissertation). National School of Engineering of Sfax.

Elleuch, N., Ben Ammar, A., & Alimi, A. M. (2015). A generic framework for semantic video indexing based on visual concepts/contexts detection. *Multimedia Tools Applications*, *74*(4), 1397–1421.

Elleuch, N., Ben Animar, A., & Alimi, A. (2010, Sept). A generic system for semantic video indexing by visual concept. In *I/V communications and mobile network (ISVC), 2010 5th international symposium on* (p. 1-4).

Elleuch, N., Zarka, M., Ammar, A. B., & Alimi, A. M. (2011). A fuzzy ontology: Based framework for reasoning in visual video content analysis and indexing. In *Proceedings of the eleventh international workshop on multimedia data mining* (pp. 1:1–1:8). New York, NY, USA: ACM.

Elleuch, N., Zarka, M., Feki, I., Ben Ammar, A., & Alimi, A. M. (2010). REGIMVID at TRECVID2010: semantic indexing. In *TRECVID 2010 workshop participants notebook papers, Gaithersburg, MD, USA, november 2010.*

Esteban, J., Starr, A., Willetts, R., Hannah, P., & Bryanston-Cross, P. (2005). A review of data fusion models and architectures: towards engineering guidelines. *Neural Computing & Applications*, *14*(4), 273–281.

Everingham, M., Eslami, S., Van Gool, L., Williams, C., Winn, J., & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, *111*(1), 98-136.

Fan, J., Gao, Y., & Luo, H. (2008). Integrating concept ontology and multitask learning to achieve more effective classifier training for multilevel image annotation. *Image Processing, IEEE Transactions on*, *17*(3), 407–426.

Faria, C., & Girardi, R. (2011). An information extraction process for semi-automatic ontology population. In E. Corchado, V. Snasel, J. Sedano, A. Hassanien, J. Calvo, & D. Slezak (Eds.), *Soft computing models in industrial and environmental applications, 6th international conference SOCO 2011* (Vol. 87, p. 319-328). Springer Berlin Heidelberg.

Fauzi, F., & Belkhatir, M. (2014). Image understanding and the web: a state-of-the-art review. *Journal of Intelligent Information Systems*, 1-36.

Fei-Fei, L., & Perona, P. (2005). A bayesian hierarchical model for learning natural scene categories. In *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on* (Vol. 2, pp. 524–531).

Feki, G., Ksibi, A., Ammar, A. B., & Amar, C. B. (2012). Regimvid at imageclef2012: Improving diversity in personal photo ranking using fuzzy logic. In *CLEF 2012 evaluation labs and workshop, online working notes, Rome, Italy, september 17-20, 2012.*

Feki, I. (2013). *Content-bbsed video retrieval based on audio concepts indexing* (Unpublished doctoral dissertation). National School of Engineering of Sfax.

Feki, I., Ammar, A. B., & Alimi, A. M. (2011). Environmental sound extraction and incremental learning approach for real time concepts identification. In *IEEE symposium on computational intelligence for multimedia, signal and vision processing* (p. 33-38).

Fel. (2006). *LSCOM Lexicon Definitions and Annotations Version 1.0* (Tech. Rep.). Columbia University. ADVENT Technical Report.

Fellbaum, C. (2010). Wordnet. In R. Poli, M. Healy, & A. Kameas (Eds.), *Theory and applications of ontology: Computer applications* (p. 231-243). Springer Netherlands.

Feng, D., Siu, W., & Zhang, H. (2013). *Multimedia information retrieval and management: Technological fundamentals and applications.* Springer Berlin Heidelberg.

Feng, L., & Bhanu, B. (2012, Nov). Utilizing co-occurrence patterns for semantic concept detection in images. In *Pattern recognition (ICPR), 2012 21st international conference on* (p. 2918-2921).

Feng, L., & Bhanu, B. (2016, April). Semantic concept co-occurrence patterns for image annotation and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(4), 785-799.

Fergus, R., Fei-Fei, L., Perona, P., & Zisserman, A. (2005, Oct). Learning object categories from google's image search. In *Computer vision, 2005. ICCV 2005. tenth IEEE international conference on* (Vol. 2, p. 1816-1823 Vol. 2).

Fergus, R., Perona, P., & Zisserman, A. (2004). A visual category filter for google images. In T. Pajdla & J. Matas (Eds.), *Computer vision - ECCV 2004* (Vol. 3021, p. 242-256). Springer Berlin Heidelberg.

Fernández-López, M. (1999). Overview of methodologies for building ontologies. In *Proceedings of the ijcai-99 workshop on ontologies and problem solving methods (KRR5) Stockholm, Sweden, august 2, 1999.*

Fuhr, N. (1992). Probabilistic models in information retrieval. *The Computer Journal*, *35*(3), 243–255.

Furht, B. (2010). Cloud computing fundamentals. In B. Furht & A. Escalante (Eds.), *Handbook of cloud computing* (pp. 3–19). Boston, MA: Springer US.

Förstner, W. (1994). A framework for low level feature extraction. In J.-O. Eklundh (Ed.), *Computer vision — ECCV '94* (Vol. 801, p. 383-394). Springer Berlin Heidelberg.

Gaines, B. R. (1978). Fuzzy and probability uncertainty logics. *Information and Control*, *38*(2), 154 - 169.

Gani, A., Siddiqa, A., Shamshirband, S., & Hanum, F. (2015). A survey on indexing techniques for big data: taxonomy and performance evaluation. *Knowledge and Information Systems*, 1-44.

Ganin, Y., & Lempitsky, V. S. (2015). Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32nd international conference on machine learning, ICML 2015, lille, france, 6-11 july 2015* (pp. 1180–1189).

Gao, T., & Koller, D. (2011). Discriminative learning of relaxed hierarchy for large-scale visual recognition. In *Proceedings of the 2011 international conference on computer vision* (pp. 2072–2079). Washington, DC, USA: IEEE Computer Society.

Garcı, À., Armengol, E., Esteva, F., et al. (2010). Fuzzy description logics and t-norm based fuzzy logics. *International Journal of Approximate Reasoning*, *51*(6), 632–655.

García, R., & Celma, O. (2005, November). Semantic Integration and Retrieval of Multimedia Metadata. In S. Handschuh, T. Declerck, & M.-R. Koivunen (Eds.), *5th international workshop on knowledge markup and semantic annotation (semannot 2005).* Galway, Ireland: CEUR Workshop Proceedings.

Gargouri, F., & Jaziri, W. (2010). *Ontology theory, management and design: Advanced tools and models.* Information Science Reference.

Garnaud, E., Smeaton, A. F., & Koskela, M. (2006). Evaluation of a video annotation tool based on the LSCOM ontology. In *Poster and demo proceedings of the 1st international conference on semantic and digital media technologies, Athens, Greece, december 6-8, 2006.*

Gilbert, A., Piras, L., Wang, J., Yan, F., Dellandrea, E., Gaizauskas, R., ... Mikolajczyk, K. (2015, September 8-11). Overview of the imageclef 2015 scalable image annotation, localization and sentence generation task. In *Clef2015 working notes* (Vol. 1391). Toulouse, France: CEUR-WS.org.

Ginsca, A. L., Popescu, A., Ionescu, B., Armagan, A., & Kanellos, I. (2014). Toward an estimation of user tagging credibility for social image retrieval. In *Proceedings of the 22nd acm international conference on multimedia* (pp. 1021–1024). New York, NY, USA: ACM.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 580–587).

Gong, B., Shi, Y., Sha, F., & Grauman, K. (2012). Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE conference on computer vision and pattern recognition, providence, ri, usa, june 16-21, 2012* (pp. 2066–2073).

Gopalan, R., Li, R., & Chellappa, R. (2011). Domain adaptation for object recognition: An unsupervised approach. In *Proceedings of the 2011 international conference on computer vision* (pp. 999–1006). Washington, DC, USA: IEEE Computer Society.

Grana, C., Serra, G., Manfredi, M., Cucchiara, R., Martoglia, R., & Mandreoli, F. (2013). UNIMORE at imageclef 2013: Scalable concept image annotation. In *Working notes for CLEF 2013 conference , Valencia, Spain, september 23-26, 2013.*

Griffin, G., & Perona, P. (2008). Learning and using taxonomies for fast visual categorization. In *Computer vision and pattern recognition, 2008. CVPR 2008. IEEE conference on* (pp. 1–8).

Grossman, D., & Frieder, O. (2012). *Information retrieval: Algorithms and heuristics*. Springer Netherlands.

Guerrero, J. L., García, J., & Molina, J. M. (2009). Multi-agent data fusion architecture proposal for obtaining an integrated navigated solution on uav's. In *Distributed computing, artificial intelligence, bioinformatics, soft computing, and ambient assisted living* (pp. 13–20). Springer.

Guðmundsson, G. Þ., Amsaleg, L., & Jónsson, B. Þ. (2012, June). Distributed High-Dimensional Index Creation using Hadoop, HDFS and C++. In *CBMI - 10th Workshop on Content-Based Multimedia Indexing.* Annecy, France.

Hamadi, A., Mulhem, P., & Quénot, G. (2014). Extended conceptual feedback for semantic multimedia indexing. *Multimedia Tools and Applications*, *74*(4), 1225–1248.

Hare, J. S., Lewis, P. H., Enser, P. G. B., & Sandom, C. J. (2006). Mind the gap: Another look at the problem of the semantic gap in image retrieval. In E. Y. Chang, A. Hanjalic, & N. Sebe (Eds.), *Multimedia content analysis, management and retrieval 2006* (Vol. SPIE V, pp. 607309–1). SPIE and IS&T. (Event Dates: 17-19 January)

Hartz, J., & Neumann, B. (2007, Dec). Learning a knowledge base of ontological concepts for high-level scene interpretation. In *Machine learning and applications, 2007. ICMLA 2007. sixth international conference on* (p. 436-443).

Hauptmann, A., Yan, R., & Lin, W.-H. (2007). How many high-level concepts will fill the semantic gap in news video retrieval? In *Proceedings of the 6th ACM international conference on image and video retrieval* (pp. 627–634). New York, NY, USA: ACM.

Hauptmann, A., Yan, R., Lin, W.-H., Christel, M., & Wactlar, H. (2007, Aug). Can high-level concepts fill the semantic gap in video retrieval? a case study with broadcast news. *Multimedia, IEEE Transactions on*, *9*(5), 958-966.

Haykin, S., & Network, N. (2004). A comprehensive foundation. *Neural Networks*, *2*(2004).

Hidaka, M., Gunji, N., & Harada, T. (2013). MIL at imageclef 2013: Scalable system for image annotation. In *Working notes for CLEF 2013 conference, Valencia, Spain, september 23-26, 2013.*

Hinton, G., & Salakhutdinov, R. (2011). Discovering binary codes for documents by learning deep generative models. *Topics in Cognitive Science*, *3*(1), 74–91.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, *18*(7), 1527–1554.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.

Hoffman, J., Guadarrama, S., Tzeng, E., Hu, R., Donahue, J., Girshick, R. B., ... Saenko, K. (2014). LSDA: large scale detection through adaptation. In *Advances in neural information processing systems 27: Annual conference on neural information processing systems 2014, december 8-13 2014, montreal, quebec, canada* (pp. 3536–3544).

Hole, A. W., & Ramteke, P. L. (2015). Design and implementation of content based image retrieval using data mining and image processing techniques. *database*, *3*(3).

Hori, T., & Aizawa, K. (2003). Context-based video retrieval system for the life-log applications. In *Proceedings of the 5th ACM SIGMM international workshop on multimedia information retrieval* (pp. 31–38). New York, NY, USA: ACM.

Horrocks, I., Patel-Schneider, P. F., & van Harmelen, F. (2003). From $\mathcal{SHIQ}$ and RDF to OWL: the making of a web ontology language. *Web Semantics: Science, Services and Agents on the World Wide Web*, *1*(1), 7 - 26.

Horrocks, I., & Sattler, U. (2005). A tableaux decision procedure for $\mathcal{SHOIQ}$. In *Proceedings of the 19th international joint conference on artificial intelligence* (pp. 448–453). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Huang, Y.-F., & Wang, S.-H. (2012). Movie genre classification using svm with audio and video features. In R. Huang, A. Ghorbani, G. Pasi, T. Yamaguchi, N. Yen, & B. Jin (Eds.), *Active media technology* (Vol. 7669, p. 1-10). Springer Berlin Heidelberg.

Hudelot, C., Atif, J., & Bloch, I. (2008). Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, *159*(15), 1929 - 1951. (From Knowledge Representation to Information Processing and Management Selected papers from the French Fuzzy Days (LFA 2006))

Hudelot, C., Atif, J., & Bloch, I. (2010). Integrating bipolar fuzzy mathematical morphology in description logics for spatial reasoning. In *19th european conference on artificial intelligence* (pp. 497–502).

Hunter, J. (2001, August). Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In *Proceedings of the 1st International Semantic Web Working Symposium.* Stanford, USA.

Huurnink, B., Snoek, C. G. M., de Rijke, M., & Smeulders, A. W. M. (2012, August). Content-based analysis improves audiovisual archive retrieval. *IEEE Transactions on Multimedia*, *14*(4), 1166–1178.

Jaeger, H. (2016). Artificial intelligence: Deep neural reasoning. *Nature*, *538*(7626), 467–468.

Jaimes, A., Christel, M., Gilles, S., Sarukkai, R., & Ma, W.-Y. (2005). Multimedia information retrieval: What is it, and why isn't anyone using it? In *Proceedings of the 7th ACM SIGMM international workshop on multimedia information retrieval* (pp. 3–8). New York, NY, USA: ACM.

Jiang, Y. G. (2015, April). Categorizing big video data on the web: Challenges and opportunities. In *Multimedia big data (bigmm), 2015 ieee international conference on* (p. 13-15).

Jiang, Y.-G., Ngo, C.-W., & Yang, J. (2007). Towards optimal bag-of-features for object categorization and semantic video retrieval. In *Proceedings of the 6th acm international conference on image and video retrieval* (pp. 494–501). New York, NY, USA: ACM.

Jiang, Y.-G., Wang, J., Chang, S.-F., & Ngo, C.-W. (2009, 29). Domain adaptive semantic diffusion for large scale context-based video annotation. In *Computer vision, 2009 IEEE 12th international conference on.*

Jiang, Y.-G., Yang, J., Ngo, C.-W., & Hauptmann, A. (2010, Jan). Representations of keypoint-based semantic concept detection: A comprehensive study. *Multimedia, IEEE Transactions on*, *12*(1), 42-53.

Jin, C., & Jin, S.-W. (2016). Image distance metric learning based on neighborhood sets for automatic image annotation. *Journal of Visual Communication and Image Representation*, *34*, 167–175.

Juneja, K., Verma, A., Goel, S., & Goel, S. (2015, Feb). A survey on recent image indexing and retrieval techniques for low-level feature extraction in CBIR systems. In *Computational intelligence communication technology (CICT), 2015 IEEE international conference on* (p. 67-72).

Jurie, F., & Triggs, B. (2005, Oct). Creating efficient codebooks for visual recognition. In *Computer vision, 2005. iccv 2005. tenth IEEE international conference on* (Vol. 1, p. 604-610 Vol. 1).

Kallel, I., & Alimi, A. (2006). Magad-bfs: A learning method for beta fuzzy systems based on a multi-agent genetic algorithm. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, *10*, 757–772.

Kanehira, A., Hidaka, M., Makuta, Y., Tsuchiya, Y., Mano, T., & Harada, T. (2014). MIL at imageclef 2014: Scalable system for image annotation. In *Working notes for CLEF 2014 conference, Sheffield, UK, september 15-18, 2014.* (pp. 372–379).

Kang, H.-B. (2003). Affective content detection using hmms. In *Proceedings of the eleventh acm international conference on multimedia* (pp. 259–262). New York, NY, USA: ACM.

Kannan, P., Bala, P. S., & Aghila, G. (2012). A comparative study of multimedia retrieval using ontology for semantic web. In *Advances in engineering, science and management (icaesm), 2012 international conference on* (pp. 400–405).

Kennedy, L., & Hauptmann, A. (2006). *Lscom lexicon definitions and annotations (version 1.0)* (Tech. Rep.). Columbia University.

Kompatsiaris, Y., & Hobson, P. (2008). *Semantic multimedia and ontologies: Theory and applications*. Springer London.

Kraft, D., Colvin, E., Bordogna, G., & Pasi, G. (2015). Fuzzy information retrieval systems: A historical perspective. In D. E. Tamir, N. D. Rishe, & A. Kandel (Eds.), *Fifty years of fuzzy logic and its applications* (Vol. 326, p. 267-296). Springer International Publishing.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Ksentini, N., Zarka, M., Ammar, A. B., & Alimi, A. M. (2012). Toward an assisted context based collaborative annotation. In *10th international workshop on content-based multimedia indexing, CBMI 2012, Annecy, France, june 27-29, 2012* (pp. 71–76).

Ksibi, A., Ammar, B., Ammar, A. B., Amar, C. B., & Alimi, A. M. (2013). Regimrobvid: Objects and scenes detection for robot vision 2013. In *Working notes for CLEF 2013 conference , Valencia, Spain, september 23-26, 2013.*

Ksibi, A., Dammak, M., Ben Ammar, A., Mejdoub, M., & Ben Amar, C. (2012, Oct). Flickr-based semantic context to refine automatic photo annotation. In *Image processing theory, tools and applications (IPTA), 2012 3rd international conference on* (p. 377-382).

Kumar, P. (2015, April). High performance object detection on big video data using gpus. In *Multimedia big data (bigmm), 2015 ieee international conference on* (p. 383-388). doi: 10.1109/BigMM.2015.65

Kumar, S., Gao, X., & Welch, I. (2016). Learning under data shift for domain adaptation: A model-based co-clustering transfer learning solution. In *Pacific rim knowledge acquisition workshop* (pp. 43–54).

Kumar, S., Rowley, H., & Makadia, A. (2014, juil.). *Content-based image ranking* (No. US 8781231 B1). Google Patents. (US Patent 8,781,231)

Landset, S., Khoshgoftaar, T. M., Richter, A. N., & Hasanin, T. (2015). A survey of open source tools for machine learning with big data in the hadoop ecosystem. *Journal of Big Data*, *2*(1), 1–36.

Lazaridis, M., Axenopoulos, A., Rafailidis, D., & Daras, P. (2013). Multimedia search and retrieval using multimodal annotation propagation and indexing techniques. *Signal Processing: Image Communication*, *28*(4), 351 - 367. (Special Issue: VS&AR)

Leite, M., & Ricarte, I. (2008). A framework for information retrieval based on fuzzy relations and multiple ontologies. In H. Geffner, R. Prada, I. Machado Alexandre, & N. David (Eds.), *Advances in artificial intelligence – IBERAMIA 2008* (Vol. 5290, p. 292-301). Springer Berlin Heidelberg.

Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006, February). Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.*, *2*(1), 1–19.

Li, L.-J., Wang, C., Lim, Y., Blei, D., & Fei-Fei, L. (2010, June). Building and using a semantivisual image hierarchy. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on* (p. 3336-3343).

Li, Z., Gong, D., Li, X., & Tao, D. (2015, Sept). Learning compact feature descriptor and adaptive matching framework for face recognition. *Image Processing, IEEE Transactions on*, *24*(9), 2736-2745.

Liggins II, M., Hall, D., & Llinas, J. (2008). *Handbook of multisensor data fusion: theory and practice*. CRC press.

Liu, G., Yan, Y., Subramanian, R., Song, J., Lu, G., & Sebe, N. (2016). Active domain adaptation with noisy labels for multimedia analysis. *World Wide Web*, *19*(2), 199–215.

Liu, Y., Zhang, D., Lu, G., & Ma, W.-Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, *40*(1), 262 - 282.

Long, M., & Wang, J. (2015). Learning transferable features with deep adaptation networks. *CoRR, abs/1502.02791*, *1*, 2.

Long, M., Wang, J., & Jordan, M. I. (2016). Unsupervised domain adaptation with residual transfer networks. *CoRR*, *abs/1602.04433*.

Lopresti, M., Miranda, N., Piccoli, M. F., & Reyes, N. S. (2012). Efficient similarity search on multimedia databases. In *Xviii congreso argentino de ciencias de la computación.*

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91-110.

Lucien, W. (1999). Some terms of reference in data fusion. *IEEE Transactions on Geosciences and Remote Sensing*, *37*(3), 1190–1193.

Luo, Y. M., & Duraiswami, R. (2008). Canny edge detection on nvidia cuda. In *Computer vision and pattern recognition workshops, 2008. CVPRW'08. IEEE computer society conference on* (pp. 1–8).

Ma, Z., Zhang, F., Wang, H., & Yan, L. (2013). An overview of fuzzy description logics for the semantic web. *The Knowledge Engineering Review*, *28*(01), 1–34.

Magalhães, J., & Pereira, F. (2004). Using mpeg standards for multimedia customization. *Signal Processing: Image Communication*, *19*(5), 437–456.

Mahmoudi, S. A., & Manneback, P. (2015). Multi-cpu/multi-gpu based framework for multimedia processing. In *Computer science and its applications* (pp. 54–65). Springer.

Mamdani, E., & Assilian, S. (1975). an experment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Machine Studies*, *7*, 1–13.

Manning, C. D., Raghavan, P., Schütze, H., et al. (2008). *Introduction to information retrieval* (Vol. 1). Cambridge university press Cambridge.

Marcia, J. B. (2012). *Understanding information retrieval systems: Management, types, and standards.* CRC Press.

Marques, O., & Furht, B. (2012). *Content-based image and video retrieval.* Springer US.

Martinet, J., Chiaramella, Y., & Mulhem, P. (2011). A relational vector space model using an advanced weighting scheme for image retrieval. *Information processing & management*, *47*(3), 391–414.

McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, *23*(0), 35 - 59.

Meghini, C., Sebastiani, F., & Straccia, U. (2001). A model of multimedia information retrieval. *Journal of the ACM (JACM)*, *48*(5), 909–970.

Memar, S., Affendey, L. S., Mustapha, N., Doraisamy, S. C., & Ektefa, M. (2013). An integrated semantic-based approach in concept based video retrieval. *Multimedia Tools and Applications*, *64*(1), 77–95.

Meng, X., Bradley, J., Yuvaz, B., Sparks, E., Venkataraman, S., Liu, D., ... others (2016). Mllib: Machine learning in apache spark. *JMLR*, *17*(34), 1–7.

Miyamoto, S. (2012). *Fuzzy sets in information retrieval and cluster analysis* (Vol. 4). Springer Science & Business Media.

Mizzaro, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science*, *48*(9), 810–832.

Mller, H., Clough, P., Deselaers, T., & Caputo, B. (2010). *Imageclef: Experimental evaluation in visual information retrieval* (1st ed.). Springer Publishing Company, Incorporated.

Mohamed, H., & Marchand-Maillet, S. (2012). Distributed media indexing based on mpi and mapreduce. *Multimedia Tools and Applications*, *69*(2), 513–537.

Mourão, A., & Magalhães, J. a. (2015). Scalable multimodal search with distributed indexing by sparse hashing. In *Proceedings of the 5th ACM on international conference on multimedia retrieval* (pp. 283–290). New York, NY, USA: ACM.

Mukesh, R., Penchala, S., & Ingale, A. (2013). Ontology based zone indexing using information retrieval systems. In S. Unnikrishnan, S. Surve, & D. Bhoir (Eds.), *Advances in computing, communication, and control* (Vol. 361, p. 181-186). Springer Berlin Heidelberg.

Muneesawang, P., Zhang, N., & Guan, L. (2014). Scalable video genre classification and event detection. In *Multimedia database retrieval* (p. 247-278). Springer International Publishing.

Mustafa, J., Khan, S., & Latif, K. (2008). Ontology based semantic information retrieval. In *Intelligent systems, 2008. IS'08 4th international IEEE conference* (Vol. 3, p. 22-14-22-19).

Mylonas, P., Athanasiadis, T., Wallace, M., Avrithis, Y., & Kollias, S. (2008). Semantic representation of multimedia content: Knowledge representation and semantic indexing. *Multimedia Tools and Applications*, *39*(3), 293–327.

Mylonas, P., & Avrithis, Y. (2005). Context modelling for multimedia analysis. In *Proc. of 5th international and interdisciplinary conference on modeling and using context (context 05), Paris, France.*

Mylonas, P., Spyrou, E., Avrithis, Y., & Kollias, S. (2009). Using visual context and region semantics for high-level concept detection. *Multimedia, IEEE Transactions on*, *11*(2), 229–243.

Möller, R., & Neumann, B. (2008). Ontology-based reasoning techniques for multimedia interpretation and retrieval. In Y. Kompatsiaris & P. Hobson (Eds.), *Semantic multimedia and ontologies* (p. 55-98). Springer London.

Naphade, M., Smith, J., Tesic, J., Chang, S.-F., Hsu, W., Kennedy, L., ... Curtis, J. (2006, July). Large-scale concept ontology for multimedia. *MultiMedia, IEEE*, *13*(3), 86-91.

Naphide, H., & Huang, T. (2001, Mar). A probabilistic framework for semantic video indexing, filtering, and retrieval. *Multimedia, IEEE Transactions on*, *3*(1), 141-151.

Neumann, B., & Möller, R. (2008). On scene interpretation with description logics. *Image and Vision Computing*, *26*(1), 82–101.

Ngo, C., Zhu, S., Tan, H., Zhao, W., & Wei, X. (2010). VIREO at TRECVID 2010: Semantic indexing, known-item search, and content-based copy detection. In *TRECVID 2010 workshop participants notebook papers, Gaithersburg, MD, USA, November 2010.*

Nguyen, C.-T. (2010). Bridging semantic gaps in information retrieval: Context-based approaches. In *VLDB doctoral workshop, Singapore 2010.*

Nixon, M., Nixon, M., & Aguado, A. (2012). *Feature extraction & image processing for computer vision.* Academic Press.

Noy, N. F., & Mcguinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology* (Tech. Rep. No. KSL-01-05). Stanford Knowledge Systems Laboratory.

Oberle, D., Ankolekar, A., Hitzler, P., Cimiano, P., Sintek, M., Kiesel, M., ... Zhou, J. (2007). DOLCE Ergo SUMO: On foundational and domain models in the smartweb integrated ontology (swinto). *Web Semant.*, *5*(3), 156–174.

Oh, C., Yi, S., & Yi, Y. (2015). Real-time face detection in full hd images exploiting both embedded cpu and gpu. In *Multimedia and expo (ICME), 2015 IEEE international conference on* (pp. 1–6).

Ordonez, V., Han, X., Kuznetsova, P., Kulkarni, G., Mitchell, M., Yamaguchi, K., ... Berg, T. L. (2015). Large scale retrieval and generation of image descriptions. *International Journal of Computer Vision*, 1–14.

Osipyan, H., Kruliš, M., & Marchand-Maillet, S. (2015). A survey of cuda-based multidimensional scaling on gpu architecture. In *Oasics-openaccess series in informatics* (Vol. 49).

Over, P., et al. (Eds.). (2010). *TRECVID 2010 workshop participants notebook papers, Gaithersburg, MD, USA, November 2010.* National Institute of Standards and Technology (NIST).

Over, P., Awad, G., Michel, M., Fiscus, J., Sanders, G., Kraaij, W., ... Quéenot, G. (2013). Trecvid 2013 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In *Proceedings of trecvid 2013.*

Over, P., Awad, G., Michel, M., Fiscus, J., Sanders, G., Kraaij, W., ... Quéenot, G. (2014). Trecvid 2014 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In *Proceedings of trecvid 2014.*

Over, P., Awad, G. M., Fiscus, J., Antonishek, B., Michel, M., Smeaton, A. F., ... Quénot, G. (2011). Trecvid 2010–an overview of the goals, tasks, data, evaluation mechanisms, and metrics.

Paliouras, G., Spyropoulos, C. D., & Tsatsaronis, G. (2011a). Bootstrapping ontology evolution with multimedia information extraction. In *Knowledge-driven multimedia information extraction and ontology evolution* (pp. 1–17). Springer.

Paliouras, G., Spyropoulos, C. D., & Tsatsaronis, G. (Eds.). (2011b). *Knowledge-driven multimedia information extraction and ontology evolution - bridging the semantic gap* (Vol. 6050). Springer.

Pan, S. J., & Yang, Q. (2010, Oct). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345-1359.

Parsons, K., A., M., Marcus Butavicius, M., S., D., & L., F. (2009). *The use of a context-based information retrieval technique.* Command, Control, Communications and Intelligence Division., DSTO Edinburgh, S. Aust.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peraldi, S. E., Kaya, A., Melzer, S., Möller, R., & Wessel, M. (2007). Multimedia interpretation as abduction. In *Proc. DL-2007: International workshop on description logics.*

Perpetual Coutinho, F., Asnani, K., & Amos Caeiro, D. (2012). Context Based Information Retrieval. *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)*, *1*(7).

Petasis, G., Karkaletsis, V., Paliouras, G., Krithara, A., & Zavitsanos, E. (2011). Ontology population and enrichment: State of the art. In *Knowledge-driven multimedia information extraction and ontology evolution* (p. 134-166).

Peters, E., & Savakis, A. (2015). Svm with opencl: High performance implementation of support vector machines on heterogeneous systems. In *Image processing (ICIP), 2015 IEEE international conference on* (pp. 4322–4326).

Petridis, K., Kompatsiaris, I., Strintzis, M. G., Bloehdorn, S., Handschuh, S., Staab, S., ... Avrithis, Y. S. (2004). Knowledge representation for semantic multimedia content analysis and reasoning. In *Proceedings of the european workshop on the integration of knowledge, semantics and digital media technology (EWIMT).*

Piras, L., & Giacinto, G. (2014). Open issues on codebook generation in image classification tasks. In *Machine learning and data mining in pattern recognition - 10th international conference, MLDM 2014, st. Petersburg, Russia, july 21-24, 2014. proceedings* (pp. 328–342).

Pravin, K., Xiangyu, W., & Alex Yong-Sang, C. (2015). Automatic image annotation using weakly labelled web data. In *Working notes for CLEF 2015 conference , Toulouse, France, september 8–11, 2015.*

Puri, A., & Chen, T. (Eds.). (2000). *Multimedia systems, standards, and networks.* New York: Marcel Dekker.

Rawat, S., Schulam, P. F., Burger, S., Ding, D., Wang, Y., & Metze, F. (2013). Robust audio-codebooks for large-scale event detection in consumer videos. In *INTERSPEECH 2013, 14th annual conference of the international speech communication association, lyon, france, august 25-29, 2013* (pp. 2929–2933).

Reimer, U., Maier, E., Streit, S., Diggelmann, T., & Hoffleisch, M. (2012). Learning a lightweight ontology for semantic retrieval in patient-centered information systems. *IGI Global - Dynamic Models for Knowledge-Driven Organizations.*

Reshma, I. A., Ullah, M. Z., & Aono, M. (2014). KDEVIR at imageclef 2014 scalable concept image annotation task: Ontology based automatic image annotation. In *Working notes for CLEF 2014 conference, Sheffield, UK, september 15-18, 2014.* (pp. 386–397).

Rodríguez-García, M., Valencia-García, R., & García-Sánchez, F. (2012). An ontology evolution-based framework for semantic information retrieval. In P. Herrero, H. Panetto, R. Meersman, & T. Dillon (Eds.), *On the move to meaningful internet systems: Otm 2012 workshops* (Vol. 7567, p. 163-172). Springer Berlin Heidelberg.

Rozilawati binti, D., & Masaki, A. (2011). Ontology based approach for classifying biomedical text abstracts. *International Journal of Data Engineering*, *2*(1).

Rueger, S. (2010). Multimedia information retrieval. In *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval* (pp. 906–906). New York, NY, USA: ACM.

Rui, Y., Huang, T. S., & Chang, S.-F. (1999). Image retrieval: Current techniques, promising directions, and open issues. *Journal of visual communication and image representation*, *10*(1), 39–62.

Saenko, K., Kulis, B., Fritz, M., & Darrell, T. (2010). Adapting visual category models to new domains. In K. Daniilidis, P. Maragos, & N. Paragios (Eds.), (pp. 213–226). Berlin, Heidelberg: Springer Berlin Heidelberg.

Sahbi, H. (2013). CNRS - TELECOM paristech at imageclef 2013 scalable concept image annotation task: Winning annotations with context dependent svms. In *Working notes for CLEF 2013 conference , Valencia, Spain, september 23-26, 2013.*

Sainath, T. N., Kingsbury, B., Saon, G., Soltau, H., rahman Mohamed, A., Dahl, G., & Ramabhadran, B. (2015). Deep convolutional neural networks for large-scale speech tasks. *Neural Networks*, *64*, 39 - 48. (Special Issue on "Deep Learning of Representations")

Salakhutdinov, R., & Hinton, G. (2009). Semantic hashing. *International Journal of Approximate Reasoning*, *50*(7), 969–978.

Salembier, P., & Sikora, T. (2002). *Introduction to mpeg-7: Multimedia content description interface* (B. Manjunath, Ed.). New York, NY, USA: John Wiley & Sons, Inc.

Sanjaa, B., & Tsoozol, P. (2007, Oct). Fuzzy and probability. In *Strategic technology, 2007. IFOST 2007. international forum on* (p. 141-143).

Schoeffmann, K., Benois-Pineau, J., Merialdo, B., & Szirányi, T. (2015). Guest editorial: Content-based multimedia indexing. *Multimedia Tools and Applications*, *74*(4), 1137-1142.

Schoeffmann, K., Mérialdo, B., Hauptmann, A., Ngo, C., Andreopoulos, Y., & Breiteneder, C. (2012). *Advances in multimedia modeling: 18th international conference, MMM 2012, Klagenfurt, Austria, january 4-6, 2012, proceedings.* Springer.

Schroff, F., Criminisi, A., & Zisserman, A. (2011). Harvesting image databases from the web. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(4), 754-766.

Sculley, D. (2010). Web-scale k-means clustering. In *Proceedings of the 19th international conference on world wide web* (pp. 1177–1178). New York, NY, USA: ACM.

Shanahan, M. (2005). Perception as abduction: Turning sensor data into meaningful representation. *Cognitive science*, *29*(1), 103–134.

Shen, J., & Cheng, Z. (2011). Personalized video similarity measure. *Multimedia Systems*, *17*(5), 421-433.

Sheu, P., Yu, H., Ramamoorthy, C., Joshi, A., & Zadeh, L. (2010). Visual ontology construction and concept detection for multimedia indexing and retrieval. In *Semantic computing* (p. 155-180). Wiley-IEEE Press.

Sigurbjörnsson, B., & van Zwol, R. (2008). Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th international conference on world wide web* (pp. 327–336). New York, NY, USA: ACM.

Simou, N., Athanasiadis, T., Stoilos, G., & Kollias, S. (2008). Image indexing and retrieval using expressive fuzzy description logics. *Signal, Image and Video Processing, Springer*, *2*, 321–335.

Simou, N., & Kollias, S. (2007). Fire : A fuzzy reasoning engine for impecise knowledge. *K-Space PhD Students Workshop, Berlin, Germany.*

Simou, N., Tzouvaras, V., Avrithis, Y., Stamou, G., & Kollias, S. (2005). A visual descriptor ontology for multimedia reasoning. In *In proc. of workshop on image analysis for multimedia interactive services (WIAMIS'05), Montreux, Switzerland, april 13-15* (pp. 13–15).

Singh, H., Gupta, M. M., Meitzler, T., Hou, Z., Garg, K. K., Solo, A. M. G., & Zadeh, L. A. (2013). Real-life applications of fuzzy logic. *Adv. Fuzzy Systems*, *2013*, 581879:1–581879:3.

Smeaton, A. F., Over, P., & Kraaij, W. (2006). Evaluation campaigns and trecvid. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval* (pp. 321–330). New York, NY, USA: ACM Press.

Smeaton, A. F., Over, P., & Kraaij, W. (2009). High-Level Feature Detection from Video in TRECVid: a 5-Year Retrospective of Achievements. In A. Divakaran (Ed.), *Multimedia content analysis, theory and applications* (pp. 151–174). Berlin: Springer Verlag.

Smeaton, A. F., Wilkins, P., Worring, M., de Rooij, O., Chua, T.-S., & Luan, H. (2008). Content-based video retrieval: Three example systems from TRECVid. *International Journal of Imaging Systems and Technology*, *18*(2–3), 195–201.

Smeulders, A., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000, Dec). Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*(12), 1349-1380.

Smith, J. R., Naphade, M., & Natsev, A. (2003). Multimedia semantic indexing using model vectors. In *Proceedings of the 2003 international conference on multimedia and expo - volume 1* (pp. 445–448). Washington, DC, USA: IEEE Computer Society.

Snoek, C. G., & Smeulders, A. W. (2010). Visual-concept search solved? *IEEE Computer*(6), 76–78.

Snoek, C. G., & Worring, M. (2008). Concept-based video retrieval. *Foundations and Trends in Information Retrieval*, *2*(4), 215–322.

Snoek, C. G. M., Member, S., Worring, M., Geusebroek, J.-m., Koelma, D. C., Seinstra, F. J., & Smeulders, A. W. M. (2006). The semantic pathfinder: Using an authoring metaphor for generic multimedia indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 1678 – 1689.

Snoek, C. G. M., & Worring, M. (2005, January). Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications*, *25*(1), 5–35.

Snoek, C. G. M., & Worring, M. (2009). Concept-based video retrieval. *Foundations and Trends in Information Retrieval*, *2*(4), 215–322.

Song, H.-J., Park, S.-B., & Park, S.-Y. (2009, June). An automatic ontology population with a machine learning technique from semi-structured documents. In *Information and automation, 2009. ICIA '09. international conference on* (p. 534-539).

Spyrou, E., & Avrithis, Y. (2008). Detection of high-level concepts in multimedia. In B. Furht (Ed.), *Encyclopedia of multimedia* (p. 151-158). Springer US.

Staab, S., & Studer, R. (2009). *Handbook on ontologies* (2nd ed.). Springer Publishing Company, Incorporated.

Stamou, G., & Kollias, S. (2005). *Multimedia content and the semantic web: Standards, methods and tools.* Wiley.

Stoilos, G., Stamou, G., & Pan, J. Z. (2006). Handling imprecise knowledge with fuzzy description logic. In *Proc. 2006 internat. workshop on description logics (dl 2006)* (pp. 119–126).

Stoilos, G., Stamou, G., Tzouvaras, V., Pan, J. Z., & Horrocks, I. (2005a). A fuzzy description logic for multimedia knowledge representation. In *Proc. of the international workshop on multimedia and the semantic web* (pp. 12–19).

Stoilos, G., Stamou, G. B., Pan, J. Z., Tzouvaras, V., & Horrocks, I. (2007). Reasoning with very expressive fuzzy description logics. *J. Artif. Intell. Res.(JAIR)*, *30*, 273–320.

Stoilos, G., Stamou, G. B., Tzouvaras, V., Pan, J. Z., & Horrocks, I. (2005b). The fuzzy description logic f-shin. In *International semantic web conference, ISWC 2005, Galway, Ireland, workshop 3: Uncertainty reasoning for the semantic web, 7 november 2005* (pp. 67–76).

Straccia, U. (1998). A fuzzy description logic. In *Proceedings of the fifteenth national conference on artificial intelligence and tenth innovative applications of artificial intelligence conference, AAAI 98, IAAI 98, july 26-30, 1998, Madison, Wisconsin, USA*. (pp. 594–599).

Straccia, U. (2006). A fuzzy description logic for the semantic web. *Capturing Intelligence*, *1*, 73–90.

Sure, Y., Studer, R., Akkermans, H., Iosif, V., Krohn, U., Lau, T., ... Studer, C. R. (1999). *On-to-knowledge methodology - employed and evaluated version.*

Tahani, V. (1976). A fuzzy model of document retrieval systems. *Information Processing & Management*, *12*(3), 177 - 187.

Terkaj, W., & Urgo, M. (2014, July). Ontology-based modeling of production systems for design and performance evaluation. In *Industrial informatics (INDIN), 2014 12th IEEE international conference on* (p. 748-753).

Thomee, B., & Popescu, A. (2012). Overview of the imageclef 2012 flickr photo annotation and retrieval task. In *Clef (online working notes/labs/workshop)* (Vol. 12).

Tong, W., Song, L., Yang, X., Qu, H., & Xie, R. (2015, June). Cnn-based shot boundary detection and video annotation. In *2015 ieee international symposium on broadband multimedia systems and broadcasting* (p. 1-5). doi: 10.1109/BMSB.2015.7177222

Torralba, A., Murphy, K. P., & Freeman, W. T. (2010, March). Using the forest to see the trees: Exploiting context for visual object detection and localization. *Commun. ACM*, *53*(3), 107–114.

Tousch, A.-M., Herbin, S., & Audibert, J.-Y. (2008). Semantic lattices for multiple annotation of images. In *Proceedings of the 1st acm international conference on multimedia information retrieval* (pp. 342–349).

Tsinaraki, C., Polydoros, P., & Christodoulakis, S. (2007, Feb). Interoperability support between MPEG-7/21 and owl in DS-MIRF. *Knowledge and Data Engineering, IEEE Transactions on*, *19*(2), 219-232.

Tunga, S., Jayadevappa, D., & Gururaj, C. (2015). A comparative study of content based image retrieval trends and approaches. *International Journal of Image Processing (IJIP)*, *9*(3), 127.

Tzeng, E., Hoffman, J., Darrell, T., & Saenko, K. (2015). Simultaneous deep transfer across domains and tasks. In *2015 IEEE international conference on computer vision, ICCV 2015, santiago, chile, december 7-13, 2015* (pp. 4068–4076).

Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., & Darrell, T. (2014). Deep domain confusion: Maximizing for domain invariance. *CoRR*, *abs/1412.3474*.

Vallet, D., Castells, P., Fernandez, M., Mylonas, P., & Avrithis, Y. (2007). Personalized content retrieval in context using ontological knowledge. *Circuits and Systems for Video Technology, IEEE Transactions on*, *17*(3), 336-346.

Van Gemert, J., Geusebroek, J.-M., Veenman, C., & Smeulders, A. (2008). Kernel codebooks for scene categorization. In D. Forsyth, P. Torr, & A. Zisserman (Eds.), *Computer vision – ECCV 2008* (Vol. 5304, p. 696-709). Springer Berlin Heidelberg.

Villegas, M., Müller, H., Gilbert, A., Piras, L., Wang, J., Mikolajczyk, K., ... del Mar Roldán García, M. (2015). General Overview of ImageCLEF at the CLEF 2015 Labs. Springer International Publishing.

Villegas, M., & Paredes, R. (2014). Overview of the imageclef 2014 scalable concept image annotation task. In *Clef 2014 evaluation labs and workshop, online working notes.*

Villegas, M., Paredes, R., & Thomee, B. (2013). Overview of the imageclef 2013 scalable concept image annotation subtask. In *Clef 2013 evaluation labs and workshop, online working notes, Valencia, Spain.*

Volkmer, T., Thom, J., & Tahaghoghi, S. (2007, Aug). Modeling human judgment of digital imagery for multimedia retrieval. *Multimedia, IEEE Transactions on*, *9*(5), 967-974.

Voorhees, E. M. (2000). Variations in relevance judgments and the measurement of retrieval effectiveness. *Information processing & management*, *36*(5), 697–716.

Vrochidis, S., Moumtzidou, A., King, P., Dimou, A., Mezaris, V., & Kompatsiaris, I. (2010). Verge: A video interactive retrieval engine. In *International workshop on content-based multimedia indexing (CBMI)* (pp. 1–6).

Wallace, M., Avrithis, Y., Stamou, G., & Kollias, S. (2005). Knowledge-based multimedia content indexing and retrieval. In *Multimedia content and the semantic web* (pp. 299–338). John Wiley & Sons, Ltd.

Waltz, E., Llinas, J., et al. (1990). *Multisensor data fusion* (Vol. 685). Artech house Norwood, MA.

Wang, C., Zhang, L., & Zhang, H.-J. (2008). Learning to reduce the semantic gap in web image retrieval and annotation. In *Proceedings of the 31st annual international ACM SIGIR conference on research and development in information retrieval* (pp. 355–362). New York, NY, USA: ACM.

Wang, F. (2011). A survey on automatic image annotation and trends of the new age. *Procedia Engineering*, *23*(0), 434 - 438. ({PEEA} 2011)

Wang, J., Jiang, Y.-G., & Chang, S.-F. (2009, June). Label diagnosis through self tuning for web image search. In *Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on* (p. 1390-1397).

Wang, S., Pan, P., Long, G., Chen, W., Li, X., & Sheng, Q. Z. (2015). Compact representation for large-scale unconstrained video analysis. *World Wide Web*, 1–16.

Wang, X., Liu, X., Pedrycz, W., & Zhang, L. (2015). Fuzzy rule based decision trees. *Pattern Recognition*, *48*(1), 50 - 59.

Wang, Z., Liu, G., Qian, X., & Guo, D. (2010). An approach to the compact and efficient visual codebook based on sift descriptor. In G. Qiu, K. Lam, H. Kiya, X.-Y. Xue, C.-C. Kuo, & M. Lew (Eds.), *Advances in multimedia information processing - PCM 2010* (Vol. 6297, p. 461-469). Springer Berlin Heidelberg.

White, T. (2012). *Hadoop: The definitive guide*. " O'Reilly Media, Inc.".

Worring, M., Snoek, C. G. M., Huurnink, B., Van Gemert, J. C., Koelma, D. C., & de Rooij, O. (2006). The mediamill large.lexicon concept suggestion engine. In *Proceedings of the 14th annual acm international conference on multimedia* (pp. 785–786). New York, NY, USA: ACM.

Wu, J., & Worring, M. (2012, April). Efficient genre-specific semantic video indexing. *Multimedia, IEEE Transactions on*, *14*(2), 291-302.

Wu, Q., Zhang, H., Liu, S., & Cao, X. (2015, April). Multimedia analysis with deep learning. In *Multimedia big data (bigmm), 2015 ieee international conference on* (p. 20-23). doi: 10.1109/BigMM.2015.27

Xu, X., Shimada, A., & Taniguchi, R. (2014). MLIA at imageclfe 2014 scalable concept image annotation challenge. In *Working notes for CLEF 2014 conference, Sheffield, UK, september 15-18, 2014*. (pp. 411–420).

Yanagawa, A., Chang, S.-F., Kennedy, L., & Hsu, W. (2007, March). *Columbia University's Baseline Detectors for 374 LSCOM Semantic Visual Concepts* (Tech. Rep.). Columbia University.

Yang, J., Jiang, Y.-G., Hauptmann, A. G., & Ngo, C.-W. (2007). Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the international workshop on workshop on multimedia information retrieval* (pp. 197–206). New York, NY, USA: ACM.

Yao, B., Yang, X., Lin, L., Lee, M. W., & Zhu, S.-C. (2010, Aug). I2t: Image parsing to text description. *Proceedings of the IEEE*, *98*(8), 1485-1508.

Yasmin, M., Mohsin, S., & Sharif, M. (2014). Intelligent image retrieval techniques: A survey. *Journal of Applied Research and Technology*, *12*(1), 87 - 103.

Yilmaz, E., & Aslam, J. A. (2008, July). Estimating average precision when judgments are incomplete. *Knowl. Inf. Syst.*, *16*(2), 173–211.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems* (pp. 3320–3328).

You, Y. (2010). *Audio coding: Theory and applications*. Springer.

Zablith, F., Antoniou, G., d'Aquin, M., Flouris, G., Kondylakis, H., Motta, E., ... Sabou, M. (2015, 1). Ontology evolution: a process-centric survey. *The Knowledge Engineering Review*, *30*, 45–75.

Zadeh, L. A. (1979). Approximate reasoning based on fuzzy logic. In *Proceedings of the sixth international joint conference on artificial intelligence, IJCAI 79, Tokyo, Japan, august 20-23, 1979, 2 volumes* (pp. 1004–1010).

Zadeh, L. A. (2008). Fuzzy logic. *Scholarpedia*, *3*(3), 1766.

Zadeh, L. A. (2014). Fuzzy set theory and probability theory: What is the relationship? In M. Lovric (Ed.), *International encyclopedia of statistical science* (p. 563-566). Springer Berlin Heidelberg.

Zadeh, L. A. (2015). The information principle. *Information Sciences*, *294*, 540 - 549. (Innovative Applications of Artificial Neural Networks in Engineering)

Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... Stoica, I. (2016, October). Apache spark: A unified engine for big data processing. *Commun. ACM*, *59*(11), 56–65.

Zarka, M., Ammar, A. B., & Alimi, A. M. (2016). Fuzzy reasoning framework to improve semantic video interpretation. *Multimedia Tools Appl.*, *75*(10), 5719–5750.

Zarka, M., Ben Ammar, A., & Alimi, A. M. (2011, April). Multimodal fuzzy fusion system for semantic video indexing. In *IEEE symposium on computational intelligence for multimedia, signal and vision processing, CIMSIVP 2011, paris, france* (pp. 60–66). IEEE.

Zarka, M., Ben Ammar, A., & Alimi, A. M. (2015). Regimvid at imageclef 2015 scalable concept image annotation task: Ontology based hierarchical image annotation. In *Working notes for CLEF 2015 conference , Toulouse, France, september 8–11, 2015.*

Zhang, D., Islam, M. M., & Lu, G. (2012, January). A review on automatic image annotation techniques. *Pattern Recogn.*, *45*(1), 346–362.

Zhang, L., Kalashnikov, D., Mehrotra, S., & Vaisenberg, R. (2014). Context-based person identification framework for smart video surveillance. *Machine Vision and Applications*, *25*(7), 1711-1725.

Zheng, Y., Ying, S., & Wang, Y. (2013). Event recognition based on co-occurrence concept analysis. In D. Ji & G. Xiao (Eds.), *Chinese lexical semantics* (Vol. 7717, p. 102-109). Springer Berlin Heidelberg.

Zhou, N., & Fan, J. (2014, April). Jointly learning visually correlated dictionaries for large-scale visual recognition applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *36*(4), 715-730.

Zhu, J. (2010). Cloud computing technologies and applications. In B. Furht & A. Escalante (Eds.), *Handbook of cloud computing* (pp. 21–45). Boston, MA: Springer US.

Zhu, W., Luo, C., Wang, J., & Li, S. (2011, May). Multimedia cloud computing. *IEEE Signal Processing Magazine*, *28*(3), 59-69.

Republic of TUNISIA
Ministry of Higher Education and Scientific Research

University of Sfax
National Engineering School of Sfax

Graduate School Sciences and Technologies

PHD Thesis
Computer System Engineering

Serial N°: .....

Fuzzy Reasoning for Multimedia Semantic Interpretation: Fuzzy Ontology Based Model for Video Indexing

Mohamed ZARKA

الخلاصة : في هذه الأطروحة، نقدم نظامٍ عام من أجل تحسين قدرات آليات الفهرسة للوثائق المتعددة الوسائطٍ. هذا النظام المقترح يقوم على ثلاث مستويات من التحليل. أولا،نقوم باستخراج المعرفة عن طريق تحليل قواعد فيديو مفهرسة يدويا. نقوم بعدها بتكوين وتملأت أنطولوجيا. ثم، نستعمل هذه الانطولوجيا وخاصية التفكير المتاحة في تحسين قدرة الآلة على التعرف على الأشياء في محتوى فيديو معين. تجربة هذا النظام المٍقترح في المسابقات العلمية *TRACVID2010*, *ImageClef2012* *ImageClef2015* أظهرت درجة جيدة من فعالية الفهرسة مقارنة مع طرق أخرى غير قائمة على المعرفة.

المفاتيح : الفهرسة الدلالي، الأونتولوجيا الضبابي، المنطق الضبابي.

Résumé : Les systèmes d'indexation multimédia à base de détecteur de Concept sont incapables de produire une interprétation sémantique satisfaisante. L'efficacité de ces systèmes peut être améliorée par l'utilisation des approches fondées sur les connaissances. Dans cette thèse, nous avons proposé un modèle automatisé et efficient pour générer une ontologie utilisée pour améliorer les interprétations sémantiques des contenus multimédias en traitant des informations conceptuelles et contextuelles. Tout d'abord, les connaissances sont extraites à travers un moteur d'abduction appliqué sur un ensemble de données d'apprentissage. Ensuite, une ontologie floue est construite pour gérer des relations floues entre contextes et concepts sémantiques. Enfin, un moteur de déduction est appliqué pour enrichir les interprétations sémantiques à propos d'un contenu vidéo. Les expérimentations réalisées dans les compagnes d'évaluation *TRECVID2010, ImageClef2012 et ImageClef2015* ont révélé une amélioration considérable dans la détection des concepts sémantiques.
Mots clés : Indexation de la Vidéo, Interprétation Sémantique, Ontologie Floue, Raisonnement Flou, Détection Hiérarchique des Concepts.

Abstract: Concept detector based multimedia indexing systems are unable to generate a satisfying semantic interpretation. The effectiveness of these systems can be improved by the use of knowledge-based approaches. In this thesis work, we proposed an automated and scalable framework to generate an ontology used to enhance semantic interpretations about multimedia contents by dealing with contextual information about concepts. First, a semantic knowledge is extracted via an abduction engine applied on a learning dataset. Second, a fuzzy ontology is constructed to handle fuzzy relationships among contexts and their semantic concepts. Third, a deduction engine is applied to deliver richer results for semantic indexing system. Experiments on *TrecVid2010, ImageClef2012* and *ImageClef2015* benchmarks have been performed to evaluate the performance of this approach. The obtained results revealed consistent improvement in semantic concepts detection when a context space is used, and a good degree of indexing effectiveness.
Key-words: Video Indexing, Semantic Interpretation, Fuzzy Ontology, Fuzzy Reasoning, Hierarchical Concept Detector.