A Framework for Temporal Content Modeling of Video Data Using an Ontological Infrastructure

Behrang QasemiZadeh
Iran University of Science and Technology
qasemizadeh@comp.iust.ac.ir

Hadi Rastgou Haghi
Iran University of Science and Technology
h_rastgou@magfa.com

Mohammadreza Kangavari
Iran University of Science and Technology
kangavari@iust.ac.ir

Abstract

Semantic video retrieval has emerged in the last decade as one of the most important features of pervasive multimedia systems. In this paper, a novel framework for temporal semantic video modeling based on ontologies is introduced. This proposed framework produces an ontological infrastructure named temporal ontology for semantic representation of video data independent of domain. This ontology later can be used to answer temporal queries.

1. Introduction

Driven by development of high-capacity storage devices and the ubiquity of digital media in global networked environments, image and video retrieval has emerged in the last decade as one of the most important features of pervasive multimedia systems. Although the initial excitement led to a hype-boosted research effort dedicated to achieve this functionality, the outcome was limited to few specialized systems based on the query-by-example retrieval model. It is only recently that the research community has focused on the challenging problem caused by the gap between the information that can be extracted automatically from visual data and the interpretation that the same data has for a user in a given situation: the semantic gap [1][2].

Because machine's understanding of the video data is still an unsolved research problem, text annotations are usually used to describe the content of the video data according to the annotator's understanding and the purpose of that video data [4]. In general, computer vision techniques may aid in answering the question "what is in the video?" but cannot answer questions such as "what is happening in the video?" or "what is the video trying to tell us?". For example, the background information of a video stream cannot be obtained directly from the video but needs to be annotated. [3][4] The importance of capturing video semantic associations lies in the fact that it can greatly improve the effectiveness of video querying by providing knowledge-based query processing.

The works in this field can be classified in two categories. In the first category, there has been increasing research efforts done about the automatic generation of the links between low-level features and high-level concepts, video annotation, to bridge the semantic gap [1], while works in the other category emphasize video modeling for efficient video data indexing, retrieval, and mining. Authors believe that works in both categories are in interaction with each other and best results can be achieved with an intelligent harmony between video content modeling and video annotation.

Former video data models such as Informedia [5], VideoText [6], VideoSTAR [7], whether they use the video annotation layering (stratification) approach [7], [18] or the keyword-based annotation approach [11] to represent video semantics, fail to model semantic relationships among the concepts expressed in the video. In face, in the previous works, there has been no good connection between frameworks and approaches to annotate video contents. Moreover, most of these works lack an explicit description of integration of temporal information in their structure which is crucial when developing an abstract model for video data due to the fact that video data essentially is a temporal document.

On the other hand and simultaneously, ontologies are newly emerged fields in Artificial Intelligence; and nowadays they are commonly used to build knowledge bases. Ontologies are able to operate as repositories to organize information for specific communities. They can be used as a tool for knowledge acquisition, in information retrieval applications, ontologies serve to disambiguate user queries, to elaborate taxonomies of terms or thesaurus in order to enhance the quality of retrieved results.

This paper introduces a general framework for unified temporal video annotation and modeling; ontologies are used to achieve this goal. The proposed framework takes video data as an input and performs automatic temporal video annotation which is organized in a new ontological infrastructure named temporal ontology, to deliver semantics of the video integrated with temporal information to the end users. Later, this temporal ontology can be used to answer users' queries about temporal information containing videos as well as high-level concepts. This paper shows how ontologies are used to represent a framework for automatic video annotation and abstraction.

This paper is organized as follows: the next section describes related works in short. The intended meaning of ontologies is described in section 3. In section 4 issues about time modeling and temporal reasoning are discussed. In section 5, the proposed framework is described. Conclusion and future works are discussed in section 6.

2. Related Works

One of the old challenges in Artificial intelligence and more specific in the field of "Information Retrieval" is the semantic retrieval of information. Although classical information retrieval systems have achieved some success in semantic retrieval of structured data, there are few enhancements for semi or unstructured data like video data. Several systems have been proposed to solve this problem [8] [16] [13] [5] [6] [7] but they have their own special strengths and weaknesses. This refers to the nature of these data.

In [4], Kokkoras et. al. introduce a video data model based on conceptual graphs. The video data model utilizes the conceptual graph knowledge representation formalism to capture the semantic associations among the concepts described in the text annotations of video data; but there is not an explicit description of modeling underlying temporal information of video data. In addition there is no framework for temporal reasoning about the conceptual model which is introduced. Also an approach for knowledge assisted semantic analysis and annotation of video content, based on an ontology infrastructure is presented in [13].

Simultaneously, there are several systems trying to prepare automatic annotation for videos. [1][13][15] But they do not specify how these annotations can be used for semantic reasoning about their underlying videos. Also most of these frameworks are domain specific and they do not show how they can be used for different kind of videos. In addition, they do not specify how they model temporal issues in their prepared annotation.

In most of these works, proposed frameworks do not support temporal relations. Even in supporting temporal information in these works, there is an important shortcoming about ontological issues about time. Actually, in most of these works, there is no concern about the explicit description of time and its abstraction in the application which is used.

3. Ontologies as a Content Theory

Theories in Artificial Intelligence (AI) fall into two broad categories: mechanism theories and content theories. Ontologies are content theories about the sorts of objects, properties of objects, and relations between objects that are possible in a specified domain of knowledge. They provide potential terms for describing our knowledge about a domain of interest [18].

Ontological analysis clarifies the structure of knowledge. Given a domain, its ontology forms the heart of any system of knowledge representation for that domain with providing the domain vocabulary. Thus, the first step in devising an effective knowledge representation system, and vocabulary, is to perform an effective ontological analysis of the field, or domain. Weak analyses lead to incoherent knowledge bases. In our study, then domain expert performs ontological analysis.

As mentioned in [18] the mechanisms are proposed as the secret of making intelligent machines but they cannot do much without a good content theory of the domain on which it is to work. Moreover, once a good content theory is available, many different mechanisms might be used equally well to implement effective systems, all using essentially the same content. In this paper we introduce ontologies as a content theory to model video data and its specific properties such as temporality.

There are several definitions for ontologies in the field of AI namely [19], [12], [20] and … but in most of them following aspects have been of major concern:

1. Ontologies are used to describe a specific domain.
2. The terms and relations are clearly defined in that domain.
3. There is a mechanism to organize the terms, (commonly a hierarchical structure is used as well as IS−A or HAS−A relationships).
4. There is an agreement between users of an ontology in such a way the meaning of the terms is used consistently.
5. Ontologies encode an implicit knowledge (semantic relations) in their structure

3.1. Why Ontologies?

Video data models are useless without specific requirements. These requirements are [14]:

1. The ability of representing video content in its structure
2. The capability of being used with a video database
3. The ability of representing temporal and structural information of video data in its structure

Comparing the mentioned requirements with the specifications that ontologies deliver by definition, the ontologies are selected as outstanding alternatives for video data modeling.

Clearly defined relations and existence of mechanisms to organize the terms in ontological structures, guarantee the capturing of semantic associations between terms that can be of a great help for video content modeling. The agreements between users of ontology guarantee the integrity of the model. In addition, semantic relations between terms, help us improve the results of query as well as data mining.

4. Time and Time Ontology

The discussion about time is rooted in philosophy. In the field of information systems, different representations of time have been used. Temporal Logic satisfies the interest of computer scientists both for its capability of expressing the needs of temporal description in knowledge engineering applications and for providing a rigorous formalism for specification and verification of concurrent and Real-Time software. Discussions and studies about this topic results in Common fundamental ontological issues which are listed below [26] [24] [10]:

1. Primitive Time Entities
2. Time Topology
3. Temporal Relationships
4. Boundedness
5. Time Structure
6. Temporal Metrics

The problems computer engineers face, do not ask for metaphysical answers on the very nature of time but they need a set of pragmatic guidelines which could assist designers and programmers in the realization of architectures and applications. Temporal Reasoning is the major concern in AI applications.

Temporal Reasoning consists of formalizing the notion of time and providing means to represent and reason about the temporal aspects of knowledge Hence a Temporal Reasoning framework should provide:

1. An extension of the language for representing the temporal aspects of the knowledge
2. A Temporal Reasoning System. A method for reasoning about the assertions which are formed using the extended language which allows one to determine the truth of any temporal logical assertion.

It is important to consider that fundamental ontological issues about time directly affect the temporal reasoning. For this reason, each temporal reasoner must indicate these fundamental issues either in an implicit or explicit way.

Again, here ontologies can be used as content theories to provide unified fundamental ontological issues about time. Time ontology aims to develop a representative ontology of time that expresses temporal concepts and properties. Time axiomatization can be declared by using time ontologies. This specifies what sort of object is going to be taken as the primitive to represent time, and imposes a set of constraints on temporal relations.

DAML-TIME [11] [23] and SOUPA time [25] are well known examples of time ontologies. In [22] sub-ontology of time is introduced. The purpose of this entry sub-ontology of time is to provide quick access to the essential vocabulary in OWL for the basic temporal concepts and relations. It is believed that it should be able to help describe most of the temporal properties of real world services, since they usually only require basic topological relations, and information about durations, dates and times.

5. Proposed Framework

Figure 1 shows the block diagram of the proposed framework. System comprises a domain knowledge, temporal ontology constructor, video annotator, video analysis tools as well as query analyzer. Interfaces are prepared to facilitate working with system.

The User defines domain knowledge including time ontology and Domain ontology through interfaces. In addition, it is possible to import the domain knowledge from other exist predefined ontologies. As a matter of fact, elements of domain ontology could be related to one of the primitive elements which are described in time ontology.

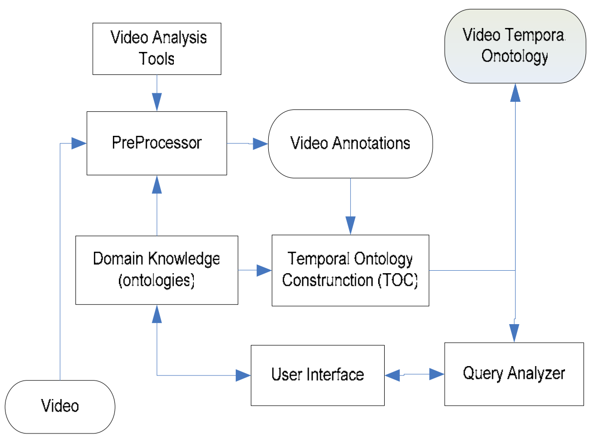

Figure 1. Block diagram of the proposed system

The task of Video Annotator is to provide video annotations in terms of domain knowledge elements as well as temporal information. For this reason, video annotator uses some rules; the user forms these rules. Interface provides a set of elements for the user to define rules. These elements are domain knowledge primitives, logical and arithmetic operations as well as video analysis functions. The user defines a pipeline from the provided elements to provide proper annotation for a video. Also it performs temporal information for each output annotation according to the specified primitives in domain knowledge and more specific in time ontology. Figure 2 shows the video annotator sub system and relations with other modules.

TOC module provided Temporal Ontology from the video annotations. In literature, usually there is no difference between the terms time ontology and Temporal Ontology but in this paper, we distinguish these terms. In the proposed framework, Temporal Ontology is defined as follows: an ontology which is used to represent the body of knowledge of a certain example (here a certain video) in a domain which is augmented with temporal primitives that are defined in a time ontology. As an example consider an ontology which is described using RDF. RDF describes the ontology in terms of triples. Here the temporal ontology will be described using Temporal RDF which is introduced in [21]. In [21], there is no explicit assertion about time. We made this assertion using a time ontology. In this way, the framework of temporal reasoning will be fully described. We have to take the differences between time ontology and temporal ontology into account. In the proposed framework time ontology represents a fact base about time while, Temporal Ontology has been used to represent truth statement which are contributed to domain ontologies for a specific example.

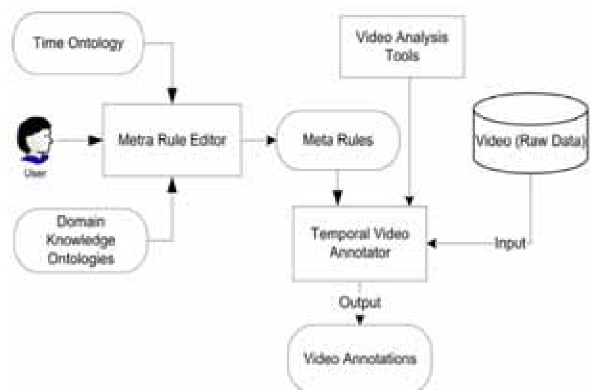

Figure 2. proposed annotation architecture

To demonstrate the advantages of this framework, we have added a simple query analyzer. The user can query about events, concepts or anything which is described in the domain ontology in addition to temporal information. For example, he/she can ask about "All events which occurred before the event B". Note that temporal relations are defined in time ontology and temporal reasoning is based on these detentions; as a result, it is possible that before relation have different intended meanings in different applications. Query analyzer infers about temporal information using the primitive relations which are defined in time ontology. In addition, the user interface provides a time-line representation of videos.

6. Conclusion and Future Works

Meaning is not a datum that is present in the image or video and that can be computed and decoded prior to the query process. It is rather a complex instantiation of static and dynamic elements emerging from relations within the system: database record itself, temporal context, user's circumstances, etc. [2]. During the last few years there has been increasing research effort done about automatic generation of the links between low-level features and high-level concepts. In this paper we have tried to cover these problems.

A new framework for automatic video abstraction independent of video data context was proposed. The proposed system provides a unified framework for automatic video annotation according to the defined domain knowledge and it constructs a temporal ontology from these annotations as an abstract model of an input video which can be used to represent meaning of videos using temporal context and relation between entities in domain knowledge.

Suggesting a unified platform to define domain knowledge including temporal information, promising coordination between the annotation process and modeling, temporal semantic representation and providing a platform for temporal reasoning are the advantages of this proposed framework.

In this version, the video analysis tools are restricted to shot detection, Text extraction, tracking and OCR. As a future work, we will try to provide more video analysis tools for annotation processes. Also in the current version, the proposed system do not support semantic web languages e.g. OWL. We will adopt the system with these languages to benefit from the future semantic web in our system.

7. References

[1] J. Calic, and E. Izquierdo. A Rule-Based Video Annotation System, Andres Dorado, IEEE Transactions on Circuits and Systems for Video Technology, VOL. 14, NO. 5, MAY 2004.

[2] B. QasemiZadeh. A Framework for Temporal Ontology Construction from Annotated Multimedia Data, MSc. Thesis, Iran University of Science and Technology, 2006.

[3] H. Jiang, D. Montesi and A.K. Elmagarmid. Integrated Video and Text for Content-Based Access to Video Databases, Multimedia Tools, 1999.

[4] F. Kokkoras, H. Jiang, I. Vlahavas, A.K. Elmagarmid, E.N. Houstis, W.G. Aref. Smart VideoText: a Video Data Model Based on Conceptual Graphs, Multimedia Systems, Springer-Verlag 2002.

[5] Informedia Digital Video Library Project Web Site. Carnegie Mellon University, Pittsburgh, PA, USA. URL: http://www.informedia.cs.cmu.edu.

[6] H. Jiang, D. Montesi and A.K. Elmagarmid. VideoText database systems. In Proceedings of the 4th IEEE International Conference on Multimedia Computing and Systems, pp. 334–351, 1997.

[7] R. Hjelsvold, "VideoSTAR – A database for video information sharing", Dr. Ing. Thesis, Norwegian Institute of Technology, November 1995.

[8] S. Dasiopoulou, V. K. Papastathis, V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. An Ontology Framework for Knowledge-Assisted Semantic Video Analysis and Annotation, Proc. 4th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot 2004) at the third International Semantic Web Conference, 2004.

[9] H. C. Shih and C. L Huang. A Semantic Network Modeling for Understanding Baseball Video, ICASSP 2003.

[10] James F. Allen: Towards a general theory of action and time. Artificial Intelligence, 23:123-154, 1984.

[11] J. R. Hobbs and F. Pan. An Ontology of Time for the Semantic Web, ACM Transactions on Asian Language Information Processing, Vol. 3, No. 1, pp. 66-85 2004.

[12] I. Mani, J. Pustejovsky, and B. Sundheim. Introduction to the Special Issue on Temporal Information Processing, ACM Trans. Asian Lang. Inf. Process. 3(1): 1-10, 2004.

[13] S. Dasiopoulou, V. K. Papastathis, V. Mezaris, I. Kompatsiaris and M. G. Strintzis. An Ontology Framework for Knowledge-Assisted Semantic Video Analysis and Annotation, Egov Open Source Conference, 2003.

[14] S Hassas, and M S Hacid. Video Data, Kogan Page, ISBN: 1903996228, 2003.

[15] Naoko Nitta, Semantic Content Analysis of Broadcasted Sports Videos with Intermodal Collaboration, PhD Thesis, Osaka University, 2003.

[16] H.C. Shih and C. L. Huang,: Statistical Understanding of Broadcasted Baseball Video Using Multi-Level Semantic Network, IEEE Trans. on Broadcasting, , 2005.

[17] Andres Dorado, Janko Calic, and Ebroul Izquierdo,A Rule-Based Video Annotation System, IEEE Transactions On Circuits And Systems For Video Technology, VOL. 14, NO. 5, MAY 2004

[18] B. Chandrasekaran, JR Josephson, and VR Benjamins. What are ontologies and why do we need them? IEEE Intelligent Systems, 14(1):20-26, Jan/Feb 1999.

[19] B. Swartout, R. Patil, K. Knight, T. Russ. Toward Distributed Use of Large-Scale Ontologies. In Proceedings of the Tenth Knowledge Acquisition for Knowledge−Based Systems Workshop, 1996.

[20] L. Khan, L. Wang: Automatic Ontology Derivation Using Clustering for Image Classification. Multimedia Information Systems 2002: 56-65

[21] C. Gutiérrez, C. A. Hurtado, A. A. Vaisman. Temporal RDF. ESWC, 93-107, 2005.

[22] F. Pan and J. R. Hobbs. Time in OWL-S, In Proceedings of AAAI Spring Symposium Series on Semantic Web Services, 2004.

[23] http://www.cs.rochester.edu/~ferguson/daml/, DAML-Time Homepage.

[24] F. A. Schreiber. Is Time a Real Time? An Overview of Time Ontology in Informatics, in: W. A. Halang, A. D. Stoyenko: Real Time Computing, Springer 1994.

[25] H. Chen, F. Perich, T. W. Finin, A. Joshi. SOUPA: Standard Ontology for Ubiquitous and Pervasive Applications, 258-267, MobiQuitous 2004.

[26] L. Vila. A Survey of Temporal Reasoning in Artificial Intelligence, AI Communications 7(1), 4--128, 1994.