Semantic-based Scene Retrieval Using Ontologies for Video Server

Min Young Jung and Sung Han Park
Department of Computer Science & Engineering Hanyang University
1271, Sa-3 Dong Sangnokgu, Ansan, KyungKi-Do, Korea
Tel: +82-31-400-4109, Fax: +82-31-406-4673
E-mail:{jmini, shpark}@hanyang.ac.kr

Abstract: To ensure access to rapidly growing video collection, video indexing is becoming more and more important. In this paper, video ontology system for retrieving a video data based on a scene unit is proposed. The proposed system creates a semantic scene as a basic unit of video retrieval, and limits a domain of retrieval through a subject of that scene. The content of semantic scene is defined using the relationship between object and event included in the key frame of shots. The semantic gap between the low level feature and the high level feature is solved through the scene ontology to ensure the semantic-based retrieval.

Keywords: Video representation, Semantic Search, Ontologies, Scene search

1. Introduction

As the rapid development of the computer communication and the growth of multimedia information, multimedia data management becomes important. The contents-based research has been steadily done after the development of the QBIC system to efficiently search multimedia data [1]. However the early contents-based research such as QBIC system is focused on an automatic information extraction. The meta-data are generated in advance to accelerate a speed of retrieving the low level feature.

For efficient indexing and retrieving of the contents-based video, accordingly to this requirement, a standardization of multimedia feature expression was required. MPEG-7(Moving Picture Experts Group layer-7) was established. The MPEG-7 unifies the information expression and storage method for the retrieval [2].

The MPEG-7 standard provides MDS(Multimedia Description Scheme) that describes low level feature and high level feature. However, the relationship between the low level information and the high level information hasn't been described. In addition, the structural approach about overall contents contained in the scene can not be discussed since the scene is indexed with one keyword or one contents. To solve such problems, fixing of the semantic gap between the low level feature and the high level feature is required. For this purpose, multimedia retrieval system employing a background knowledge structured by ontology or thesaurus has been attempted.

Many researches have been done to construct ontology using the hierarchical tree structure of terms. Structured ontology can reason not only the meaning of certain terms but also the actual meaning in the relations of vocabularies. Also, it can infer another meaning of terms. The representative ontology which has such the function is MESH or Gene Ontology. These are mainly used in the limited domain. In addition, non-domain specific knowledge structures such as WordNet[3] and Cyc[4] is existed.

Ontologies have been utilized for the video retrieval. Hoogs and Stein et al. [5]-[9] describe the low level feature and high level feature using extended WordNet. However, this Hoogs and Stein's method is not suitable for the structure of the video retrieval, because the existing ontology WordNet is used so that the structure is too expensive and complicated. Because the indexing structure describes only the keyframe image, it retrieves the image not the video.

This paper is intended to propose the semantic indexing structure for scenes using the ontology. For this purpose, it is necessary to segment the video into the semantic scene. In Chapter 2, the video ontology system for indexing semantic scene is implemented. In Chapter 3 a simulation to compare the performance of implemented method and previous method is presented.

2. Video Ontology System

As shown in Figure 1, the video retrieval ontology system proposed in this paper is largely composed of four steps; the semantic scene creation, the video analysis, the video ontology creation, and the semantic scene search.

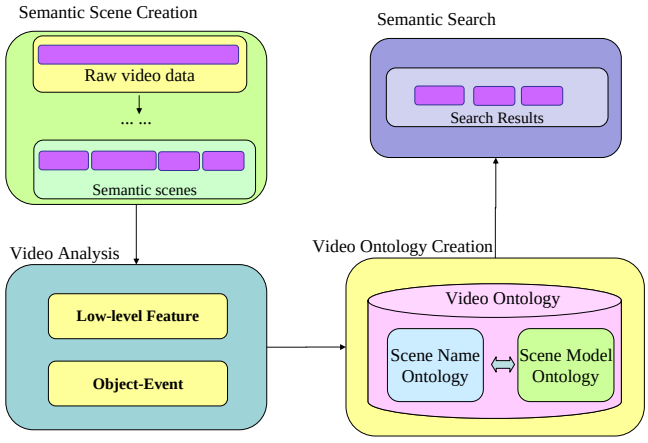

Figure 1. Ontologies system for semantic video retrieval.

This work was supported by grant No.(R01-2006-000-10876-0) form the Basic Research Program of the Korea Science & Engineering Foundation

First, the semantic scene creation step is a preprocessing for modeling and searching video in the scene unit. In this paper, the basic unit is defined as the semantic scene. The basic unit for the ontology implementation and retrieval is used. The video analysis step analyzes video data using low-level feature and high-level feature provided in MPEG-7 Standard. The video ontology is constructed based low-level feature and object-event information obtained in the video analysis. Finally, the video retrieval shows retrieval result of the semantic scene using the implemented video ontology.

2. 1 Semantic scene creation

Video data primarily has composed of three parts; frame, shot and scene. A frame is one of the many single photographic images in a motion picture. A shot is a continuous strip of motion picture that runs for an uninterrupted period of time. Shots are generally filmed with a single camera and can be of any duration. Frames, shots, scenes, and sequences form a hierarchy of units fundamental to many tasks in the creation of moving-image works.

In order to model and retrieve the video in the scene unit, it is necessary to segments into the semantic scene unit. In this paper, the basic unit is defined as the semantic scene. Genre of overall video data before semantic scene creation is defined. In general, one genre is decided for the whole video. It must be used as the information for determining the domain when the scene name ontology is implemneted. After decision of video genre the semantic scene creation such as Figure 2 is passed.

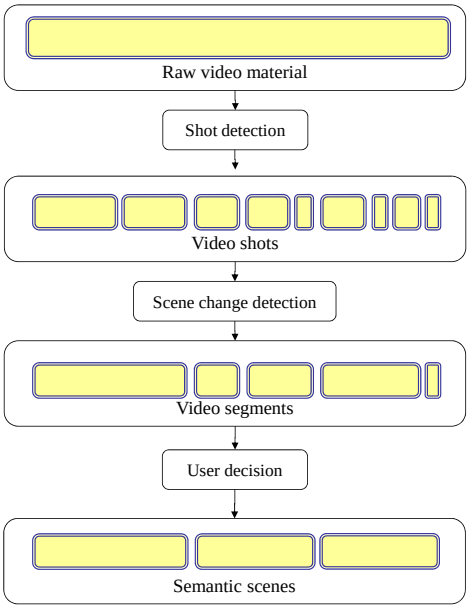

Figure 2. Semantic scene creation procedures.

First of all, the video is automatically segmented in the shot unit. Segmented shots are automatically grouped using the scene change algorithm to create video segments. However, because these video segments are semantically insufficient yet, they are grouping in the semantic scene unit through the user decision.

2. 2 Video analysis

Video analysis acquires low level feature and high level feautre from semantic scene. The basic unit of all semantic scene is shot. A single frame from the shot is chosen. Seleted frame is the keyfram. The keyframe includes low level features for scene model ontology such as color, shape and texture. In addition, The high level feature such as the object and event is extracted from the key frame. The object is described with a kind and the event is defined by using a direction and velocity.

2. 3 Video ontology creation

In order to obtain the desired result of retreival result by using the video ontology, the video ontology data must be implemented in an early stage. The background knowledge implementation is largely composed of the scene name ontology and the scene model ontology as shown in Figure 3.

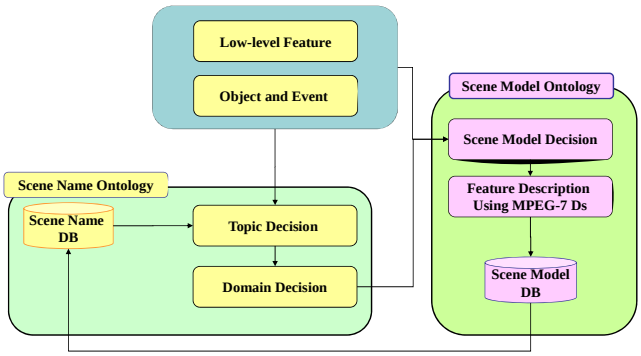

Figure 3. Video ontology construction processing.

Scene name ontology structurizes and stores the terms of the object and event that one subject can contain according to the genre in advance for terms for determining the domain. Scene model ontology is implemented by describing the feature of semantic scene in the metaldata.

2. 3. 1 Scene name ontology

Scene name ontology is the dictionary of terms to be used for contents of the object and event. Through the hierarchical structure of terms, the high-level concept and low-level concept is defined. Then, the new concept can be deduced by defining the hierarchical relations of these terms.

The video indexing process using the scene name ontology indexes the subject of semantic scene by using previousely-implemneted scene name ontology. For this reason, the scene name ontology must be implemented before the video indexing.

In terms of the internal structure of scene name ontology, the terms list of subjects to be used at the scemantic scene indexing is stored. These terms are

configured with the objects and event list tree as shown in Figure 4. Because the subject which defines the scene in the domain is formed with one terms, an inclusion relation of terms is defined. For determining the subject of semantic scene, the genre information and the object & event information are used. Indexing and retreival can be done within the range of genre domain in which the subject is contained through utilizing the indexed subject information.

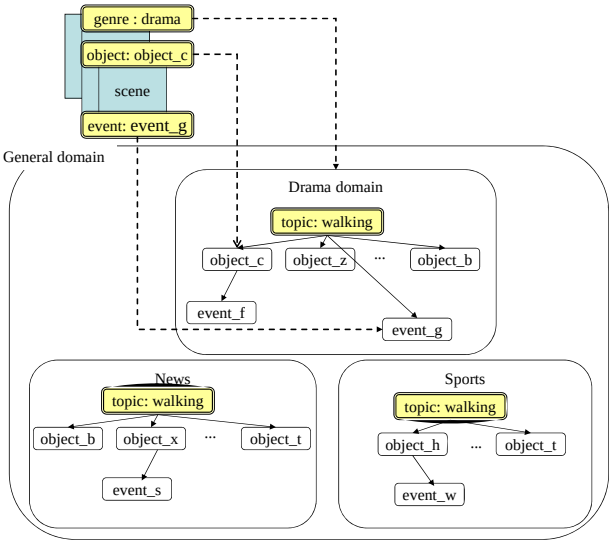

Figure 4. Graph of the word properties in scene name ontology.

Figure 4 shows the example of that the subject of one semantic scene is determined in the scene name ontology. By using ‘drama’, the genre information obtained from the semantic scene segmentation process, ‘Drama’ domain is determined. Then, it is indexed with the subject of ‘walking’ which is contains ‘object_c’ and ‘event_g’, the object and event information in the tree.

2. 3. 2 Scene model ontology

Scene model ontology is intended to defining and storing the relationships between the low level information that these subjects can have and the high level information such as the object and event. That is, the scene model is semantically defined in the relation of the event and object in the scene.

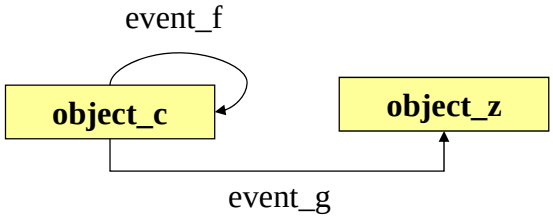

Figure 5. Relation of object and event.

In particular, the event is not simply defined with one key frame, but is defined with the relationship of objects which exist in each key frame image. That is, the direction of the event is defined as the movement of the same object between shots, and the velocity is defined based on the time interval between shots. Figure 5 shows the diagram of the relationship between the object and the event. 'object_c' has 'event_f', and 'object_c' & ‘object_z' has 'event_g'.

The language to be used for storing in the ontology is stored by using the RDFS(Resource Description Framework Schema), the metadata structure which can define this relationship. RDF provides the standardization for the semantic, Structure and Syntax for connecting various metadata. RDF expresses three features, and there are the Object, Attribute and Value. Also, RDF Schema describes the features of resources, and has the function of describing and defining the relations between resources and the restrictions. In Figure 6, the metadata storage structure is simplified. The storage structure in Figure 6 becomes the basic metadata structure that the semantic scne is stored in the scene model ontology.

```
<genre category=id>
  <scene model = topic>
   <relation name=“rel_id“>
    <object type = id>
     <color>  value </color>
     <texture>  value </texture>
     <shape>  value </shape>
    </object>
    <event type= id>
     <object type=“direction” >  value </object >
     <object type=“velocity”>  value </object >
    </event>
   </relation>
  </scene>
</genre>
```

Figure 6. Scene model framework using XML.

As it is derived as the basic structure, one semantic scene and the whole of video can be defined.

2. 4 Semantic scene search

The semantic scene retrieval to search the tree construt in scene name ontology is presented. Semantic scenes of matching result include not only concept of the topic keyword but also sub-concept of the topic keyword. After the query is inputed, the topic decides using object and event in the scene name ontology. The result of searching a semantic scene shows the scene that has a keyword such as the topic decision. Also, the scene name ontology is composed of domains according to the genre of video. The reason why the ontology is configured according to the genre, is because the used terms can be perfectly different when the genre is different inspite of the same text. Accordingly, by

suggesitng the range of retrieval by genre, the disadvantage that the range of retrieval is extended by the retrieval result covering up to the similar semantic terms can be solved.

3. Simulation

The search capabilities of the semantics based search system proposed in this paper and the contents-based search system is evaluated. Simulation data used in our simulation is five dramas of 60 minutes with 306 shots and 100 semantic scenes. The metadata of implemented ontology are parsed to be easily retrieved and then stored the database with structured table format.

A performance evaluation of the video retrieval system is experimented in terms of two query type; keyword query and the semantic query. Keyword query has simply a concrete meaning and semantic query has an abstract meaning.

Precision is the rate of the related scenes to the retrieved scenes. The precision may be a good measure of search performance. Our simulation results show that the conventional method shows better performance in simply text matching by keyword query. However the proposed method shows 48% better performance in semantic query type than the previous method since the relation concepts are employed through the ontology.

The result of recall shows the actually-retrieved rate among all related experiment data of the video retrieval. Therefore, among all experiment data, there are only parts which are related to the keyword question in the existing papers. Because the retrieval is based on only the text match, if the meaning is same but the text is different, the terms are not retrieved.

On the other hand, the proposed method retrieves most of related scenes because the conceptual contents are implemented through the video ontology. By analyzing the meaningn of the terms which define the scene not the simple text match like the keyword retrieval, semantically-similar scenes with the different form of terms are all shown as the retrieval result. It shows the satisfactory retrieval to users.

4. Conclusion

If the low level feature is well extracted, video ontology system shows the ontology unit as the semantic retrieval results as well as the scene unit. The video ontology supplements the structural disadvantage of metadata that the existing MPEG-7 Standard has. The existing MPEG-7 structure is the storage structure in which the semantic information and low level information are independent.

The precision of retrieval was improved through the domain restriction by the scene name ontology. And the video can be semantically recognized based on the scene unit through the scene unit ontology. If the large amount of video ontology is implemented, the video will seem to be automatically analyzed.

References

[1] M. Flickner et al. "Query by image and video content: The QBIC system," *IEEE Computer*, vo1 28, no 9, pp.23-32, September 1995.

[2] J.M. Martinez. Overview of the mpeg-7 standard. Technical Report 5.0, ISO/IEC, Singapore, 2001.

[3] C. Fellbaum, WordNet: *An Electronic Lexical Database* ,Bradford Book, 1998.

[4] D. Lenat and R. V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project.* Addison-Wesley. 1990.

[5] L. Hollink, A.Th. Schreiber, J. Wielemaker, and B. Wielinga, "Semantic annotation of image collections," *In Proc. of the Semantic Workshop*, USA, 2003.

[6] A. Hoogs, J. Rittscher, G. Stein, and J. Schmiederer., "Video Content Annotation Using Visual Analysis and a Large Semantic Knowledgebase," *In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol 2, pp. 327-334, 2003.

[7] G.C. Stein, J. Rittscher, and A. Hoogs. "Video annotation using a semantic database extended with visual knowledge." *In Proc. of ICME*, vol 1, pp.161-164, 2003.

[8] L. Hollink, M. Worring, A.Th.Schreiber, "Building a Visual Ontology for Video Retrieval", *In Proc. of the ACM Multimedia*, pp.479 - 482, November 2005.

[9] S. Bloehdorn, K. Petridis, C. Saathoff, N. Simou, V. Tzouvaras, Y. Avrithis, S. Handschuh, I. Kompatsiaris, S. Staab, and M. G. Strintzis, "Semantic Annotation of Images and Videos for Multimedia Analysis," *Lecture Notes in Computer Science*, vol. 3532, pp. 592-607 , May 2005.