Contemporary Engineering Sciences, Vol. 8, 2015, no. 13, 581 - 587
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ces.2015.5266

Ontology Based Video Retrieval System

J. Jerin Jose

Faculty of Computing
Sathyabama University Chennai-600119, India

Abstract

Multimedia is one of the prior ways of communication channel for mankind. Due to the advancement in technology and enormous growth in mankind civilization there are tremendous amount of multimedia data available. So, obviously we are in need of some technique for retrieving these data. This paper will give an overview of Ontology based Video retrieval system.
Through the ontologies the system will express key entities and relationships describing learning material in a formal machine processable representation. An ontology-based knowledge representation could be used for content analysis and concept recognition, for reasoning processes and for enabling user-friendly and intelligent multimedia content search and retrieval.

Keywords: MPEG7, Video Retrieval, CBIR, Ontology, Computer vision

I. Introduction

Due to today's advancement in technology there are tremendous amount of multimedia data. There are so many concepts and system to handle these issues, but the lack in that entire concept is that a searching system is has to be in such a way that it should think and act like human does. This concept is famously known as computer vision. The requirement needed to make a machine think like human is by providing the clear-cut semantic of all the data's available to the system, so that it knows the knowledge of what really it is searching. To create a searching system as equivalent to human way of knowledge searching system, we need to know how a human organize and search knowledge. In that Ontology [Thomas Hofwebwer-2004] is one of the promising way of balancing human search to machine search. Where, in ontology the concepts, individuals and relation between the concepts of a domain is been organized in a well formed way thus the

complete semantic and syntactic relation between a concept is been provided which ease the way of searching. The tricky part of ontology is lays on how to identify the concepts and its relationship and how to provide the functionality of that concept. In this paper we try to provide an ontological way of binding the semantic gap between high and low level feature which would pave a way for efficient and effective video retrieval system. This paper is organized in such a way that, first we provide an overview of video and video structure, then the general procedure used for video retrieval concepts is been reviewed , related work taken part in video retrieval procedure is analyzed and finally narrated our proposed framework of Ontology based video retrieval system.

II. Background

Digital video signal is actually consists of time varying consecutive still image. Thus the images in digital video is been digitalized both spatially and temporally [Yu-Gang Jiang, Qi Dai, Jun Wang, Chong-Wah Ngo, XiangyangXue and Shih-Fu Chang -2012]. In video the consecutive images is called as frames, thus to have an illusion of motion picture video the number of frame required for one second is normally is about 25 frames but as per NTSC standards we use 30 frames per second. For the pixel arrangement of the frames there are some standards format, CIF (common Intermediate format) uses 352 x 288 and so on. The frame pixel format differs from mobile application, video conferencing, High Definition and Normal broadcasting Television signal. The complete video sequence is sub-divided into Scenes, Shots and Frames. Among these process the shot boundary detection and key frame detection is one of the major concepts to be considered. Content based image and video retrieval concept is one of the very old research concept where yet now active research participation is going on. Some of the notable research work and prototype which is still provide an basic foundation for multimedia retrieval system are discussed here: QBIC [M Flickner, H Sawhney, W Niblack, J Ashley, Q Huang et al-1995] in this prototype system, query would be an image or rough sketch. The feature such as RGB composition of color, coarseness, contrast and directionality feature of Texture and shape feature such as area, circularity, eccentricity, major axis orientation and a set of algebraic moment invariants of shape are used. Visual Seek [John R.Smith, Shih-Fu Chang-1996] in this system the query for the image retrieval system would be the color of the needed image. VIDEOQ [Shih-fu Chang, William Chen, Horace J. Meng, HariSundaram, Di Zhong-1997] is a query by image or sketch based system, in addition of considering color, shape and texture of the image, they mainly concentrate on spatial temporal constrain. MARS [6 Huang, Tom; Mehrotra, Sharad; Ramchandran, Kannan-1996] they try to provide a standard way of multimedia analysis and retrieval system. They too use the basic content of image such as color, texture, shape and layout. VIRE [M. Rautianen et al-2002] this video browsing and Retrieval system (VIRE) uses the image content and textual details of video for indexing. The image content used are color, Motion and audio. IBM’s cue video system [B. Adams et al-2002] This is one of the IBM's

research project for video Retrieval system exhibited for TREC 2002 They uses enormous visual feature such as color histogram, color correlogram, edge orientation histogram, wavelet texture, Tamura texture, co-occurrence texture, motion vector histogram and Mel-frequency coefficients.

III. Proposed Framework

In this session Ontology based video retrieval system was proposed. Here, we are using both image and video as in video retrieval system a key frame of a shot is mainly used, which is again a still image. The Multi-Modal means, here in our proposed system we are using both high and low level feature of an image.
The low level feature of the image is normally the color, shape and texture of the image. In our system after analysis we are using MPEG 7 [Manjunath, BS, Jens-Rainer Ohm, Vinod V Vasudevan & Akio Yamada -2001] and Texton [Song-Chun Zhu,Chen-En Guo, Yizhou Wang & ZijianXu-2005] based features. The High level feature of an image is usually textual wording used around the images. The word "Semantic Gap" is used by the researcher to specify the g0ap between the low level and high level features. In our work we try to reduce those gaps through ontological way of explaining the high and low level features.

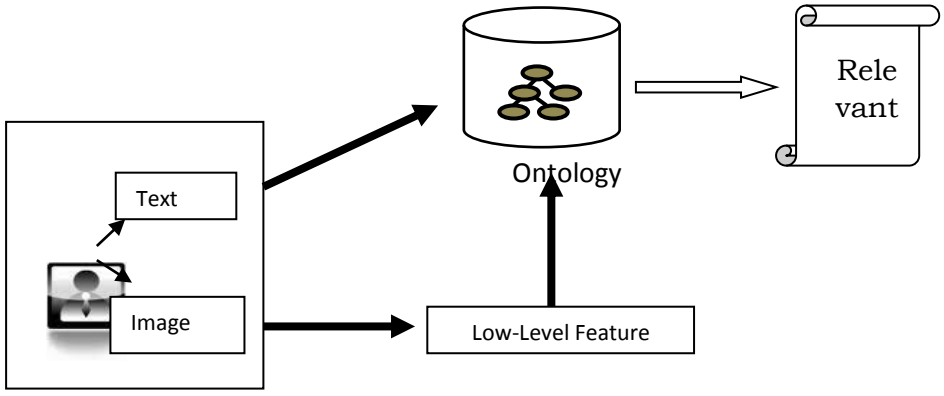

Fig. 1: Proposed System User Interface

Fig.1 shows the general overview of the proposed system's user interface, where a user can give either text or image as a query. If the given query is textual in nature, from the developed domain based multi-modal ontology repository the relevant set of videos and images can be listed. If the given query is in image in nature then MPEG 7 based low level feature is extracted and matched with the multi-modal ontology repository for the relevant set of videos. HereI have created this ontology in such a way that it can be used for both image and video

retrieval. In case of video retrieval, the video frames are analyzed and from the analysis the key-frame of the each shot is determined from the determined key-frames the low level feature is extracted.

There are many basic definitions for ontology [Yildirim, Y., Adnan Yazici, Turgay Yilmaz - 2013], it is one of the organized way of explaining any complex concept. It is one of advance topic of artificial intelligent it is a study where we make the machine way of thinking equivalent to human way of thinking and reasoning. In human memory system hierarchy organization of data is one of the technique which is been equivalent to ontology. Ontology is written in a standard XML format language called OWL, which look just like an XML language. In this our work we need a retrieval system which should be as rational as human thinking, we use the concept of ontology in specifying data. To create ontology we require domain knowledge from where we can create domain ontology. With this domain ontology concept we can reduce the semantic gap between low and high level feature. As the domain we are working on is natural scenery an ontology for those concept is been created as shown in Fig. 2

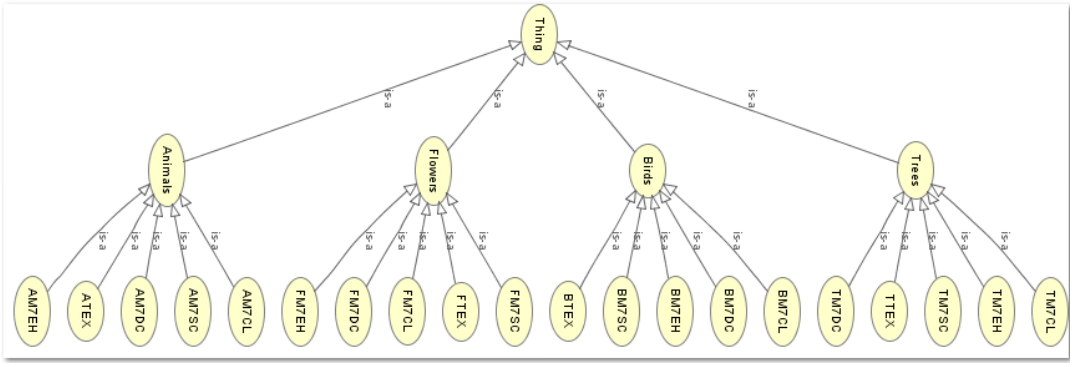

Fig. 2 High level feature Ontology

This ontology view is created by Protégé [Protégé version 4.2.0-2006], where we have created a general ontology of the natural scenery concept with Animal, Flower, Birds and Tree for each image categories its corresponding MPEG 7 based Color layout, Scalable color, Dominant color, Edge histogram and Texton value can also be interpreted. To integrate the high and low level feature the domain ontology of the low level feature has to be integrated to that of above high level feature ontology

In the ontology of natural scenery video and images, below the multi- modal fusion of flower domain class. From Fig.2 the class Flower has five sub-classes as FM7CL, FM7SC, FM7DC, FM7EH and FTEX. That is F for flower, M7 for MPEG 7 and CL – colour layout, SC – scalable color, DC - dominant color and EH – Edge histogram, TEX for Texton.For this ontology a modal set-up is shown

for flower ontology as shown in Table 1. Where the sub-class of flower is merged with corresponding low-level ontology. Thus FM7SC is fused with scalable color of low level ontology, FM7CL is fused with color layout of low level ontology, FM7DC is fused with dominant color of low level ontology and FTEX is fused with Texton of low level ontology.

<table>
<tr><td rowspan="5">FLOWERS</td><td>FM7DC</td><td>DOMINANT COLOR</td><td rowspan="3">COLOR</td><td rowspan="5">RECOURSES</td></tr>
<tr><td>FM7SC</td><td>SCALABLE COLOR</td></tr>
<tr><td>FM7CL</td><td>COLOR LAYOUT</td></tr>
<tr><td>FM7EH</td><td>EDGE HISTOGRAM</td><td rowspan="2">TEXTURE</td></tr>
<tr><td>FTEX</td><td>TEXTON</td></tr>
<tr><td colspan="2">HIGH-LEVEL ONTOLOGY</td><td colspan="3">LOW-LEVEL ONTOLOGY</td></tr>
</table>

Table 1 Multimodal ontology Fusion

IV. Conclusion

In this paper, the concepts of video retrieval system were analyzed. To improve the strategy behind content based video retrieval system, in this paper a proposal with respect to ontology were enlightened. The MPEG 7 based low level feature content of the retrieved key-frame are extracted. The extracted features are included in the ontology. The precision of this proposed framework is about 78%. The precision rate can be improved by enhancing the ontology with domain knowledge.

References

[1] B. Adams et al., “IBM Research TREC-2002 Video Retrieval System”, Proc. Text Retrieval Conference TREC 2002 Video Track, Gaithersburg, MD, 2002.

[2] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang et al "Query by Image and Video Content: The QBIC System"Computer (1995) Volume: 28, Issue: 9, Publisher: AAAI Press, Pages: 23-32.
http://dx.doi.org/10.1109/2.410146

[3] Huang, Tom; Mehrotra, Sharad; Ramchandran, Kannan"Multimedia Analysis and Retrieval System (MARS) Project "Papers presented at the 1996 Clinic on Library Applications of Data Processing, March 24-26, 1996 Urbana-Champaign.

[4] John R.Smith, Shih-Fu Chang. "VisualSEEk: a Fully Automated Content-Based Image Query System". In ACM Multimedia, Boston, MA, November 1996", Pages: 87-98. http://dx.doi.org/10.1145/244130.244151

[5] Manjunath, B. S., Jens-Rainer Ohm, Vinod V. Vasudevan & Akio Yamada 2001, 'Color and Texture Descriptor', IEEE Transaction on Circuits and Systems for video Technology, vol.11, no.6, pp. 703-715.
http://dx.doi.org/10.1109/76.927424

[6] M. Rautianen et al., "TREC 2002 Video Track Experiments at MediaTeam Oulu and VTT", Proc. Text Retrieval Conference TREC 2002 Video Track, Gaithersburg, MD, 2002.

[7] Shih-fu Chang, William Chen, Horace J. Meng, HariSundaram, Di Zhong," VideoQ: An Automated Content Based Video Search System Using Visual Cues (1997)" In Proceedings of ACM Multimedia, Pages: 313-324.
http://dx.doi.org/10.1145/266180.266382

[8] Song-Chun Zhu, Chen-En Guo, Yizhou Wang & ZijianXu 'What are Textons?" International journal of Computer vision (2005), vol.62, no.1-2, pp.121-143. http://dx.doi.org/10.1023/b:visi.0000046592.70770.61

[9] Thomas Hofwebwer,"Logic and Ontology entry", Stanford Encyclopedia of Philosophy, First published Mon Oct 4, 2004; substantive revision Tue Aug 30, 2011.

[10] Yildirim, Y, Adnan Yazici & TurgayYilmaz 2013, 'Automatic semantic content extraction in videos using a fuzzy ontology and rule based model', IEEE Transactions on Knowledge and Data Engineering, vol.25, no.1, pp. 47-61.
http://dx.doi.org/10.1109/tkde.2011.189

[11] Yu-Gang Jiang, Qi Dai, Jun Wang, Chong-Wah Ngo, Xiangyang Xue and Shih-Fu Chang. Fast Semantic Diffusion for Large-Scale Context-Based Image and Video Annotation. IEEE Transactions on Image Processing, Vol. 21, No. 6, pp. 3080 – 3091, 2012. http://dx.doi.org/10.1109/tip.2012.2188038

[12] Protege version 4.2.0 developed by Stanford University and University of Manchester 2006.

Received: March 9, 2015; Published: June 18, 2015