International Journal of Fuzzy Logic and Intelligent Systems, vol. 12, no. 4, December 2012, pp. 324-328
http:// dx.doi.org/ 10.5391/IJFIS.2012.12.4.324 PISSN 1598-2645 eISSN 2093-744X

Study of Cross-media Retrieval Technique Based on Ontology

Su Mei Xi[1] • Young Im Cho[2]

[1] College of Information, Shandong Polytechnic University, Jinan, 250353, China
[2]College of Information Technology, the University of Suwon, San 2-2, Bongdam-eup, Hwaseong, 445-743, Korea
E-mail: xsm@suwon.ac.kr, ycho@suwon.ac.kr

Abstract

With the recent advances in information retrieval, cross-media retrieval has been attracting lot of attention, but several issues remain problems such as constructing effective correlations, calculating similarity between different kinds of media objects. To gain better cross-media retrieval performance, it is crucial to mine the semantic correlations among the heterogeneous multimedia data. This paper introduces a new method for cross-media retrieval which uses ontology to organize different media objects. The experiment results show that the proposed method is effective in cross-media retrieval.

Keywords: Cross-media Retrieval, Ontology, Similarity calculation

1. Introduction

With the rapid development of internet and multimedia technology, multimedia information increase quickly which use texts, audios, images, videos, the three dimensional model as the carrier, and big change occurs in information retrieval area. So the importance of multimedia retrieval is becoming more and more obvious.

Currently most practical multimedia retrieval systems adopt text search technology based on keywords, while these keywords can not represent the multimedia information objectively and completely. In order to make up the shortages, scholars began to study the content-based multimedia retrieval, and created more mature theories and some practical prototype systems, such as SpeechBot for audio retrieval, commercial image retrieval system QBIC and video retrieval system VideoQ and so on.

There still exist some drawbacks of content-based multimedia retrieval [1]. For example, when the user using a tiger image as a query his expected results may be some related images, texts, videos and even tiger's voice. That is to say, the returned results are some related multiple media while inputting one kind of medium.

Content-based retrieval only pays attention to the lower features usually (such as color and texture of image, rhythm of audio), so they can not satisfy the mentioned request, and can not correspond to the manner of obtaining information from multichannel of human beings. Researchers put forward to cross-media technique in order to realize the seamless connection among multiple model media.

Because of heterogeneous bottom features, problems such as construction of relationship between the different model media and the similarity calculation have not been solved efficiently. In view of this, ontology is used for organize and relate all kinds of media in order to realize cross-media retrieval.

Ontology-based cross-media retrieval is a new retrieval schema [3], which is based on content-based retrieval and store multimedia semantic information by using ontology. It obtains core set of each concept for each media in ontology by machine learning method after extracting each media object features. Its subordinate concept can be obtained through computing the similarity between the eigenvalue of each retrieval request and each class core. It can use concepts to realize any media type retrieval.

Existing multimedia retrieval systems mostly can only retrieve single mode multimedia database or can operate multi-model media data but not support cross-media retrieval. Figure 1 is the image results about inputting automobile audio by the method of literature [6]. It can be seen that, comparing with the traditional content-based retrieval techniques, because of no restriction of retrieval cases and returning results for cross-media retrieval, its function is more powerful and it is used for wider area. Furthermore, research meaning of cross-media retrieval also shows outstanding performance in the following two aspects.

(1) Human perception system itself can be understood as a cross-media retrieval system. For example, when seeing a tiger image, its bellow will be imagined; when seeing a green apple, its acid taste will be imagined. In this sense, cross-media retrieval system is not only more powerful but also more satisfied to human beings thinking mode. With the development of science technology, in addition to visual, auditory information, input and output of future computer may be expanded to more perception type such as touch and taste, and the meaning of computer cross-media system will be more abundant at that time. Since conforming to the process of human perception and thinking system, as a new platform, the

Manuscript received Dec. 3, 2012; revised Dec. 19, 2012; accepted Dec. 24, 2012.
* Corresponding author: Young Im Cho (ycho@suwon.ac.kr)

cross-media retrieval system not only enriches computer services but also extends computer functions. Hence with the development of computer technology, the research of cross-media retrieval technique will be more important.

(2) Multimedia objects are essentially complementary, so cross-media retrieval system meets the requirements of complementary characteristics of multimedia data. Firstly, multimedia objects are complementary in information, for example, a bird's yell and a bird's image reflecting the bird's attribute from different aspects, although the same semanteme they belong to but including different information. These two kinds of information are complementary. Image information can not substitute of voice information and vice versa. For this point, cross-media retrieval mechanism is necessary to realize the retrieval of complementary information.

Furthermore, multimedia objects are complementary in semanteme. When they are ambiguous that analyzed separately, which appear simultaneously, the semanteme will be more explicit when they are regarded as whole. For this point, the multimedia semantic complementarity needs to be considered adequately when mining the multimedia semanteme, that is to say, when mining the multimedia objects semanteme, different kinds of multimedia objects contents need to be considered comprehensively for spanning the "semantic gap" preferably.

2. Related Works

Cross-media retrieval is a relatively new research direction. Because of the heterogeneous of extracted lower features of various media data, complexity of cross-media retrieval is higher than traditional multimedia retrieval. For example, for content-based multimedia retrieval, we only need to compare the similarity among single mode features, but for cross-media retrieval, we need to calculate the similarity among multi mode features, such as m dimension feature of image and n dimension feature of audio. At present the main two methods about cross-media retrieval complexity processing are as follows:

(1) First methods, they analyze the feature correlation between different media types, and construct cross-media association table to maintain correlation. Among them, the typical method is to map the eigenvector of different type into a single space by using CCA (canonical correlation analysis) method, and computing the similarity among them in the new space [4]. Implied external relationship may play a supplementary role, for example, finding the correlation among different media by the hyperlinks of web pages. Such approach is better to promote the development of cross-media retrieval, but it is not very precise for describing the connection degree among media.

(2) Second methods, they construct knowledge base for multimedia data. Literature [5] analyzed the multimedia data of digital library, indexed different type media separately, and to realize cross-media retrieval. But this method is mainly based on keywords, it has big workload and can not describe the information that multimedia include properly. Therefore, this paper uses the knowledge base manner of literature [5], based on the content-based multimedia retrieval, organizing multimedia information by using ontology, analyzing the content and semanteme of multimedia, and achieving media span at semantic level.

3. Ontology-Based Cross-media Retrieval

3.1. System Framework of Cross-media Retrieval

Cross-media retrieval system framework is designed around ontology in this paper, as shown in figure 2. We preprocess (like smooth, splitting, etc.) the user's retrieval query, and then generate the specific dimension eigenvalue by calling feature extraction module. Then we traverse all concept nodes of ontology; compare the similarity between this eigenvalue and all other concepts to determine the semantic information about the media represented. According to the user's demand, we return any kinds of media file as the result of this node. So, we can realize cross-media retrieval.

The storage process is similar to the above process. After preprocessing and feature extraction, the new media are compared with the rules of ontology in order to determine the concept which the media represent, laying it under the proper node and adjust the inner structure of this node.

3.2. Architecture of Ontology

The above analysis shows that ontology plays an important role in various links of the whole strategy. Therefore the architecture of ontology should be valued.

In order to show the dependency between the concepts, here our ontology structure is a kind of layered structure, in which each node represents an abstract concept and the range of the concept of lower node representing is belong to the range of the higher. The total ontology consists of all ontology nodes and their connections. Each node its structure is shown as figure 3.

Here each concept is divided into several parallel parts according to the difference of media type, and each different media type can represent one same concept. Then there are several core sets under each media type, which represents a side of this concept in this kind of media. We can describe one concept from different sides. So, we need several different core sets.

The media that satisfy the above rules also attach on the core set, and are sorted according to the similarity degree with the current concept. Under the core set there are cores, which describe all conditions under some side, and the core is represented by eigenvalue of each kind of feature type.

In general, the higher condition matching degree is the closer similarity of the media and this concept is. For example, panda concept can be described by several media. Among them, in image area, panda can be described from multi sides, such as the appearance is black white and hairy body. Maybe there are some other descriptions, and each description is a core set, while the histogram corresponding to black white and the texture feature corresponding to hairy both are a class core under this core set separately.

For good scalability of the above ontology structure it can be used in large cross-media retrieval system. Because of realizing cross-media in concept level in this paper, which corresponding to the rules that people knowing the world, it can make the correlation more precise among various media type.

3.3. Retrieval Process and Expansion

For a new image R of user submitting, similar images and other type media can be returned.

Algorithm 1. ***Cross-media Retrieval Algorithm***

(1) Extracting characteristics of R by calling feature extraction algorithm;

(2) Traversing the entire Ontology from top to down, repeating step (3) for all nodes, computing the similarities between R and each node, the node of which has the biggest similarity of R is regarded as the subordinate concept of the R, and we note this ontology node as N, go to step (4);

(3) Computing the similarity of R and some ontology node, traversing all core sets of this node, computing the similarity of the extraction features according to algorithm 2 and each core set, the biggest value among these similarities can be regarded as the similarity of R and this concept;

(4) Returning the nearest k_1 results by using content-based image retrieval method among all images under node N, the result set is recorded as C_1;

(5) For other type media under node N, returning the front k_2 results of the queue, the result set is recorded as C_2;

(6) $C_1 \cup C_2$ are results of returning to the user and feedback can be executed according to user's selection.

In addition to retrieve other media that similar to the target medium directly, query expansion can be done according to the user's demand, or we can use the following method to process the retrieval demand.

(1) If the user expected media consist of several concepts simultaneously, the weights of every concept can be assigned by the user. All the media under these concepts will be listed according to algorithm 1 when retrieving, and then the results can be obtained by synthesizing every weight of concept.

(2) If the user expected media consist of some concepts except other concepts, we can let the user assign the weight of each concept, and process them according to method (1), negative value for the weight of exclusive concepts.

3.4. Similarity Computation

Fundamentally, a core set consists of n class cores, and the rule that represents is an n-dimensional vector set. Each vector is the average of all media eigenvalues under this core set, i.e. class core. We design the following algorithm for calculating the similarity between some medium and the core set.

Algorithm 2. ***Similarity calculation of medium and core set algorithm***

(1) According to the selected m features, calling the corresponding feature extraction procedure, obtaining the eigenvalue vector set v_1 that consists of m vectors.

(2) Contrasting to the feature type of core set, deleting the corresponding features' vector value of the exclusive core set, obtaining the eigenvalue vector set v_2 that consists of m' vectors.

(3) Calculating the distance (Euclidian distance) D_i ($0< i \le m'$) of m' vectors and the core set separately, and defining $d = \frac{1}{m'}\sum_{i=1}^{m'} D_i$ as the similarity of this medium and the core set.

3.5. User Feedback

When returning the final results to the user we expect the user's feedback in order to record satisfactory results, adjust the inner structure of ontology node and make next retrieval result more accurate. For example, the user retrieves an image R, who expects the satisfactory images and audio. We can return the retrieved images and audio media to the user and add R to the ontology by using algorithm 1 to compute the distance between the core set and the concept node. For the returned audio files, the correct results can be marked by the user and recorded in R. When the image R is retrieved in next time, the above results can be returned together.

4. Experiment Results and Analysis

This paper implements a cross-media retrieval system by VC++ 6.0, which supports cross-media retrieval between image and audio. Minitype ontology is constructed previously, which includes 8 concepts, as shown in figure 4. Among these 8 concepts, there are 100 images and 50 audio fragments (all files come from the Internet), and they are divided into two groups randomly. The first group includes 60 images and 30 audio fragments, which are used to create ontology core set. The other 40 images and 20 audio fragments are regarded as second group, mixed together as test collection.

Firstly we arrange the audios and images of first group to the corresponding ontology nodes manually, then extract the lower eigenvalues, cluster these eigenvalues according to the k-means algorithm, and form several cores to store. Next, we delete the multimedia objects of first group in database, so the used data in the system are only the learned rules through them. And then the second group audios and images are added into the system according to the text method. Finally we submit an image or audio file arbitrarily, and retrieve the relevant images and audios separately.

In this experiment, a lion image is inputted to the system firstly, retrieving images and audios of lion, as shown in figure 5. And then an aircraft start audio fragment is inputted, retrieving audios and images of aircraft, as shown in figure 6. It can be seen that, the ontology method that this paper proposed can realize cross-media retrieval efficiently.

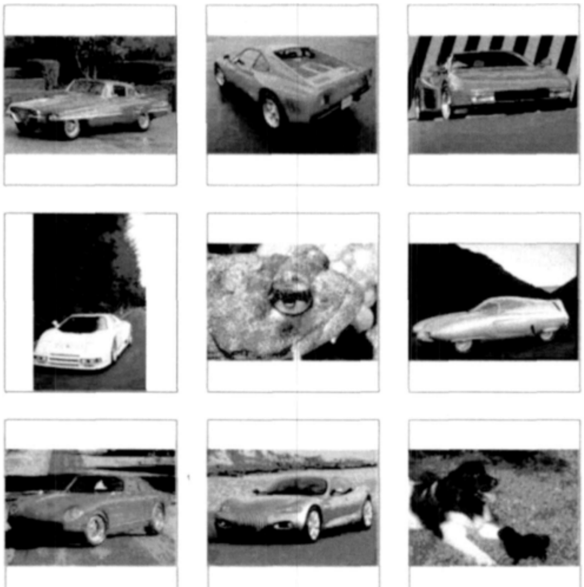


Figure 1. Front 9 returning results of submitting automobile voice

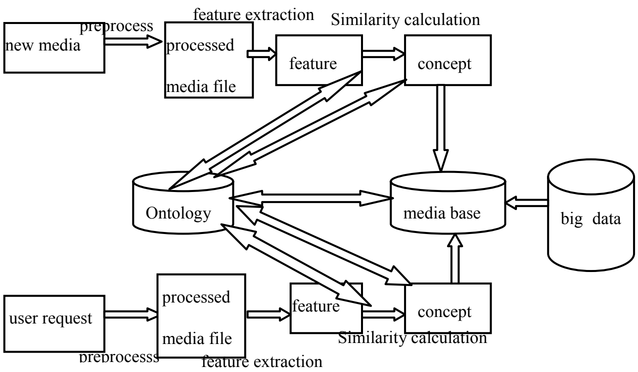


Figure 2. Cross-media retrieval system framework

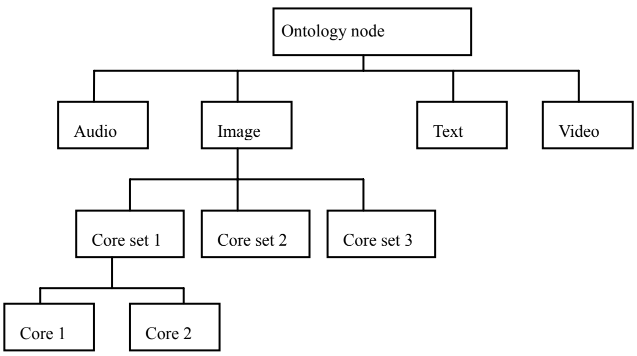


Figure 3. Node structure of ontology

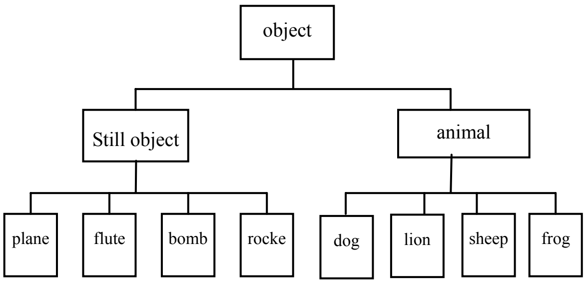


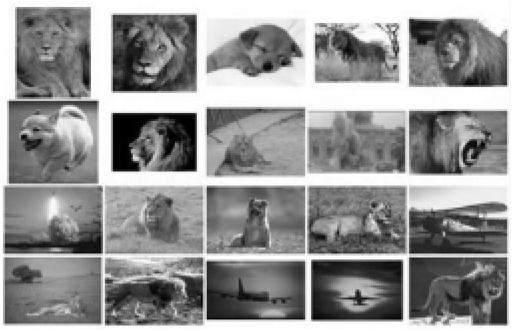
Figure 4. Architecture of ontology

5. Conclusions and Future Work

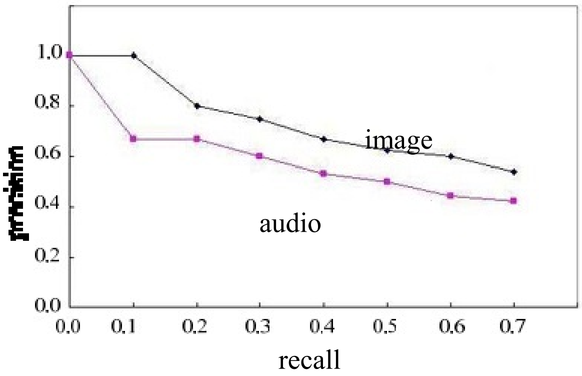
Core to the ontology philosophy is the ability to extend a generic framework in such a way to accommodate requirements for media markup. This paper uses ontology to manage different media and realize the relationship semantically among several media. Specific work is as follows: we design a feasible cross-media retrieval framework and an ontology structure used for cross-media. Similarity calculation and retrieval can be realized on this basis. However, this method also need be improved. For example, we need better theoretical support for how constructing a domain ontology. The current concepts are extracted by hand, so the application of ontology can not be expanded in a large scale. Further-more, the constructs provided for motivation and trait description may be employed both for annotation and analysis.

With the emergence of the semantic web, it is increasingly important for annotation of data into machine-readable formats. The ontology provides a shared conceptualization of this domain which will act as an enabling technology for this evolution of the web.

As a new research direction, related research about cross-media has just started, and cross-media retrieval is only one issue of cross-media research area. Some other works need us to consider: 1) constructing more reasonable system framework of cross-media knowledge representation; 2) more efficient cross-media inference method; 3) how to store and manage the large scale cross-media data efficiently; 4) how to realize efficient and accurate cross-media retrieval and mine semantic relationship efficiently based on large scale cross-media data.



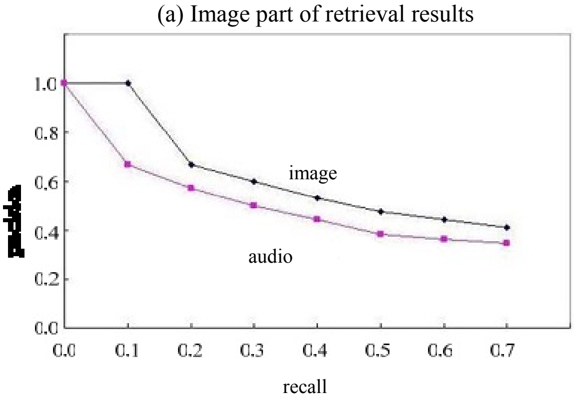
(a) Image part of retrieval results



(b) Retrieval results

Figure 5. Retrieval results of inputting an image

(a) Image part of retrieval results



(b) retrieval results

Figure 6. Retrieval results of inputting an audio file

Acknowledgment

This work is supported by two projects of Shandong Province Higher Educational Science and Technology Program (J12LN09), China, and by natural science foundation Project of Shandong Province, China (No. zr2011fm028).

References

[1] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based Multimedia Information Retrieval: State of the Art and Challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications*, Feb. 2006.

[2] G. Zhou, K. M. Ting, F. T. Liu, and Y. Yin, "Relevance feature mapping for content-based multimedia information retrieval," *Pattern Recognition*, vol.45, pp.1707-1720, 2012.

[3] Y. Zhou and J. Hu, "Cross-Media Topic Analysis and Information Retrieval," *Proc. of 9th Int. Conf. on Fuzzy Systems and Knowledge Discovery*, pp.1211-1215, 2012.

[4] A. Smeulders, M. Worring, S. Santini S, et al., "Content-based Image Retrieval at the End of the Early Years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, no.12, pp.1349-1380, 2000.

[5] J. Yang, Y. Zhuang, and Q. Li, "Search for Multi-modality Data in Digital Libraries," *Proc. of the 2nd IEEE Pacific Rim Conf. on Multimedia: Advances in Multimedia Information Processing*, Berlin, Germany: Springer, pp. 482-489, 2001.

[6] Y.T. Zhuang, Y. Yang, and F. Wu, "Mining Semantic Correlation of Heterogeneous Multimedia Data for Cross-media Retrieval," *IEEE Transactions on Multimedia*, vol.10, no.2, pp. 221-229, 2008

Su Mei Xi
2001 B.A. Shandong University of Science and Technology, China
2009 Master Shandong University.
2012 current Ph.D Course student.
Interest areas: neural network, fuzzy logic, artificial intelligence, information retrieval, ubiquitous computing, etc.

Young Im Cho
1994 Ph.D Korea University, Korea
1996 Senior Researcher of Samsung Electronics
2012 current Professor of the University of Suwon.
Interest areas: Intelligent system, ubiquitous system, information retrieval, neural network, fuzzy system, etc.