CANFIND – A SEMANTIC IMAGE INDEXING AND RETRIEVAL SYSTEM

Chin-Hwa Kuo, Tzu-Chuan Chou, Nai-Lung Tsao, and Yung-Hsiao Lan

Computers and Networking Laboratory (CAN Lab.)
Department of Computer Science and Information Engineering
Tamkang University, Taiwan, ROC
+886-2-2623-8784, chkuo@mail.tku.edu.tw

ABSTRACT

We present CanFind, a semantic image indexing and retrieval system in this paper. To identify the target images of interest in the database in the conceptual level, the presented system makes use of keywords as the input of searching vehicle. The system consists of two subsystems, i.e., *semantic indexing* and *query expansion*. In the semantic indexing, the subsystem includes three main building blocks, namely, keyword extraction, keyword expansion, and keyword weighting. The information of WordNet is used to extend existing keywords associated with images. This design intends to overcome the drawbacks in conventional keyword-based image retrieval system. Next, the resulting word set is filtered by a filter to extract common words from the word set and set up the image indexing for the corresponding image. In the query expansion, corpus is used to help users find relative or precise results in the facing dilemma of too few or too many query results for a given query. The designed semantic image indexing and retrieval system is integrated with IWiLL, a web-based language learning platform to further illustrate the value of the designed system.

Keywords: Keyword expansion, query expansion, semantic image indexing and retrieval, e-Learning

1. INTRODUCTION

Image information retrieval is an active research subject in information technology [1]. Its purpose is to retrieve images or image sequences that are relevant to a query. Such an information system can have many potential applications in many areas, to name just a few, telemedicine, digital libraries, distance learning, tourism, GIS, etc. Research in this area has gone through conventional keyword-based approach and then content-based visual information retrieval (CBR).

Note that the access to visual information is not only performed at a perceptual level, but also a conceptual level. To achieve this objective, one of the research directions in semantic modeling and representation makes use of semantic networks to retrieve target images [2]. Many image database research projects devoted a lot of efforts in this field. MediaNet [3] uses partially annotated collections of multimedia data to enhance the retrieval of multimedia data. The concepts and relationships between the concepts are defined and exemplified by multimedia information such as text, images, video, and audio-visual descriptors. Furthermore, MediaNet models the traditional semantic relationship types such as generalization and aggregation but adds additional functionality by modeling perceptual relationships based on feature descriptor similarity and constraints. In addition, Yang etc. proposed thesaurus-aided approaches to facilitate semantics-based access to images [4]. They constructed the semantic hierarchy, which supports flexible image browsing by semantic subjects. They also formulated a semantic similarity metric to get incorporated with visual similarity to improve the accuracy of image retrieval. In the above two papers, both of them want to build a connection between semantic concept and image low-level features to provide a semantic retrieval function on a CBR image database system.

Our approach here is along the same direction as semantic networks, but we attempt to use semantic concept in the indexing phase instead of using low-level feature. The proposed system consists of two subsystems, i.e., semantic indexing and query expansion. The semantic indexing subsystem includes three main building blocks, namely, keyword extraction, keyword expansion, and keyword weighting. The information of WordNet [6] is used to extend existing keywords associated with images. This design intends to overcome the drawbacks in conventional keyword-based image retrieval system. Next, the resulting word set is filtered by a filter to extract common words from the word set and set up the indexing for the corresponding image.

In the query expansion, BNC (British National Corpus) [7] is used to help users find relative or precise results in the facing dilemma of too few or too many query results for a given query. When too few or too many results occur, the system provides several candidate keywords which are highly related with the input query according to the co-occurrence rates in the BNC. This cue gives users further guideline about their desire. After user selects the desired candidates in this step. The system provides more relative images for the situation of too few results. On the other hand, the system provides the more precise results in the first page for the situation of too many results. Thus, users can allocate the target images in a ranking order. The designed semantic image indexing and retrieval system is integrated with IWiLL, a web-based language learning platform to further illustrate the value of the designed system.

The rest of this paper is organized as follows. In section 2, we describe the semantic indexing subsystem. In section 3, we present the query expansion subsystem. Results and an application are illustrated in section 4. Finally, the conclusions and future works are given in the section 5.

0-7803-7761-3/03/$17.00 ©2003 IEEE

2. SEMANTIC INDEXING SUBSYSTEM

As shown in Fig.1, the fundamental building blocks of designed semantic indexing subsystem include keyword extraction, keyword expansion and keyword weighting mechanisms. In the keyword extraction phase, the stop words and punctuations are eliminated inside the annotated text. We extract the rest words, use lemmatization for each word and count their occurrence frequencies within the annotation. And then use WordNet to expand the keywords. WordNet is an English lexical database. It includes many kinds of semantic relationship between word senses, for example, synonym, hypernym, hyponym, antonym, and so on. Because different people describe the same thing sometimes using different words, it is reasonable to expand the keyword set of images in our system. Note that we only use synonym and hypernym to expand to keywords because using these two relationships among word senses would not expand some keywords that are not appeared in the corresponding images. But if we use hyponym, antonym, irrelative keywords may be added.

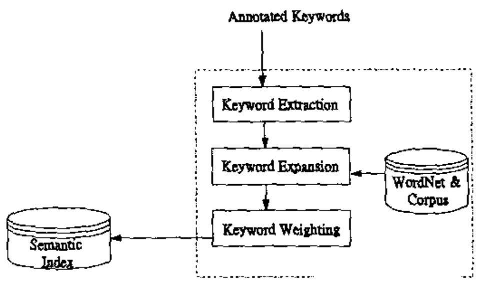

Figure 1: Semantic indexing subsystem

Next, the occurrence frequency of each word in BNC is used. In order to eliminate the unsuitable expansion in the keyword expansion phase, low-frequency words in BNC are filtered. Table 1 presents an example of keyword expansion, and filtering process. The filtering process removes the italic and bold words in Table 1. In this example, if an image is annotated with keyword, child or children, users can search it by lots of keywords, e.g. child, children, kid, youngster, minor, person, individual, someone and so on.

Table 1. Keyword expansion and filtering process

child	After keyword expansion process	
	synonym	kid, youngster, minor, ***shaver, nipper, tiddler, tike, tyke, fry, nestling***
	hypernym	person, individual, someone, somebody, ***mortal***, human, soul
	After the filtering process	
	synonym	kid, youngster, minor
	hypernym	person, individual, someone, somebody, human, soul

In order to rank the output, the original annotated keywords and expanded keywords are weighted by different way. On one hand, the weights of the original annotated keywords are set by the frequencies in original annotation for each image, and on the other hand, the expanded keywords are weighted by following equations.

$$w(t') = \begin{cases} 0.8 \times w(t) & \text{, if } t' \text{ is synonym of } t \\ 0.5 \times w(t) & \text{, if } t' \text{ is hypernym of } t \end{cases}$$

After the three steps, i.e. keyword extraction, expansion, and weighting, we build a semantic index for each object in our image database.

3. QUERY EXPANSION SUBSYSTEM

The query keyword used by users is the most significant but not always sufficient in the query phase. For example, if a user uses "sea" to be the query input, he only can get the object indexed by "sea" in conventional keyword-based database system. If there is a picture that comprises a small "island" in the "sea", and the annotator only uses "island" to describe the picture, we cannot obtain this picture in the above scenario. So we use a query expansion process to help users expand the query keyword set. In query expansion process, we refer to a thesaurus including the relevant degree between words. The thesaurus is built automatically by the modified method of the traditional methods for automatic thesaurus construction [8] with computing co-occurrence rates. We use parts of BNC, i.e., a three million words corpus, and the collocation computation technique to construct the thesaurus. Our method is as follows.

For each two different words w_i, w_j in corpus

$$\text{MI-score}(w_i, w_j) = \log \frac{f_{i,j} \times c}{f_i \times f_j}$$

f_i: is the frequency of w_i in corpus.

f_j: is the frequency of w_j in corpus.

$f_{i,j}$: is the frequency of the co-occurrence of w_i, w_j in corpus. Here the co-occurrence means that the distance of two words is less than N words. (In the designed system, N is set for 20.)

c : The total word number of corpus.

MI-score [9] is chosen as the relevant degree between words in the designed system. High MI-score means the co-occurrence rate of the two keywords is high. After experimental investigation, we choose the MI-score 10 to be the threshold for selecting the candidate keywords for query expansion. For examples, see Table 2, the high relevant degree word sets of "sea" and "bomb", i.e. MI-score is above 10, are listed in the Table 2. Therefore, if the user input the keyword "sea" and select the function of query expansion, the system will response a list with high relevant degree of the words with "sea", e.g. ocean, sand, harbor, island, boat, shore. After user select some of the candidates, the system will search with more related relevant images and more precision images in the first page.

Table 2. High relevant degree words of "sea" and "bomb"

sea		bomb	
Keyword	MI-score	Keyword	MI-score
ocean	10.959067	mortar	12.135934
sand	10.698609	explosion	11.016353
harbor	10.530977	terrorist	10.772177
island	10.530973	commando	10.529053
boat	10.518034		
shore	10.096508		

In the case that given a query results in many image outputs, the images that the user really wants may be scattered in all pages. This is a common situation of the present keyword-based image retrieval system. Therefore, we provide another set of keywords which are highly related with the first input to guide a user to construct their desires further. With the second input keyword, the system is able to allocate the high ranked images in a place that facilitates user browsing. Our experimental indicates that the above is very convenient tool to help users to find the requested images in this case:

On the other hand, given a query input results in a few image outputs, the user may want more candidate images to select. By the same query expansion process as mentioned before, the system provides more image outputs that match the second keyword expansion. Therefore, it produces more image outputs to user.

In the keyword expansion phase, we do not use this thesaurus because words relevant to the original keyword in this thesaurus may not represent the corresponding images. But in the query expansion phase, users will be helped to retrieval more relative precision images by the above process. However, by using the keyword expansion and query expansion, the proposed system will provide twofold semantic image retrieval functions.

4. RESULTS AND APPLICATIONS

To illustrate the designed scheme, we implemented the CanFind, a web-based image retrieval system. It includes an image upload user interface, when a new image is uploaded into the system; the annotation about the image is requested. Users are requested to describe the uploading images with metadata and main data. The metadata includes category, type (still or moving), color (black and white or color), and source. The main data consists of subject (what), object name (who), event (how), place (where), time (when) and description [5]. These fields guide the users to provide more enriched descriptions. The purpose of these textual fields is to establish a high-level conceptual image retrieval mechanism. Once we gather relative keywords of each image, users can use keywords to find desired images.

Fig. 2 and Fig. 3 show examples without/with keyword expansion. For instance, when we input keyword "house" to search images, the system retrievals 29 images without keyword expansion and 65 images with keyword expansion, respectively. On the other hand, when user inputs keyword "girl" to the system. And system responses very limited results. A semantic search that makes use of query expansion will generate related keyword set. In this case, it includes the candidates as follows: boy, daughter, nice, rake, wonderful, consent, guitar, pretty, sad and folk, see Fig. 4. These candidates appear frequently with keyword "girl" in the BNC. When user selects some of these candidates, the system retrieval more relative and precise images closed to users' request. The system displays these images in a ranking order.

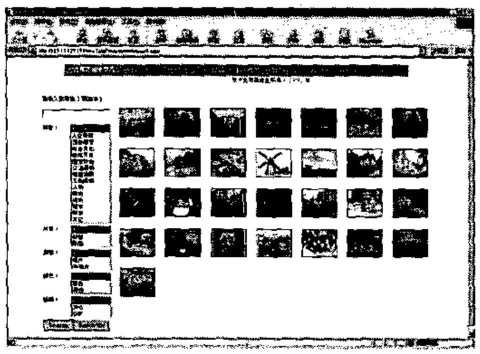

Figure 2. Query with "house" without keyword expansion

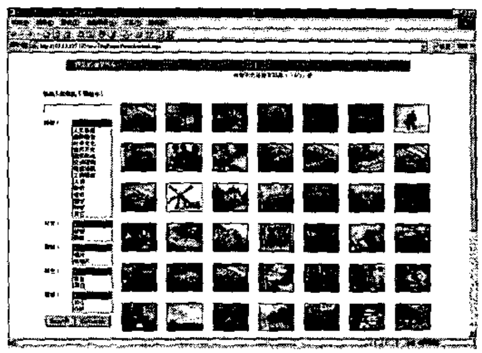

Figure 3. Query with "house" with keyword expansion

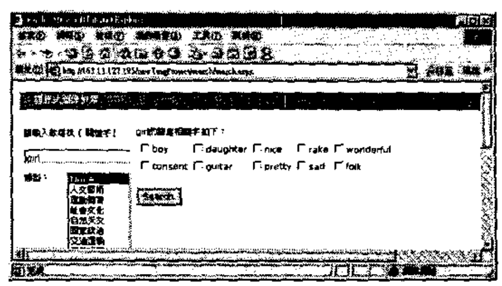

Figure 4. Query expansion with keyword "girl"

The proposed semantic multimedia database system is applied in the Intelligent Web-based Interactive Language Learning system (IWiLL) [10][11] and conventional classroom learning. The IWiLL system comprises management subsystem and authoring tool subsystem. In the management subsystem, teachers can browse, share and assign lectures to classes, and annotate and correct student assignments. In the authoring subsystem, teachers can create, edit, modify and remove lectures on the web. Teachers use the authoring tool just like editing in a common word processor. Teachers extract and choose suitable multimedia materials for teaching vocabulary, scenario study, and listening improvement. They can freely insert and delete words, images, and movie clips, and other interactive components, and they can change their properties by clicking the button on the authoring toolbars showed in Fig. 5.

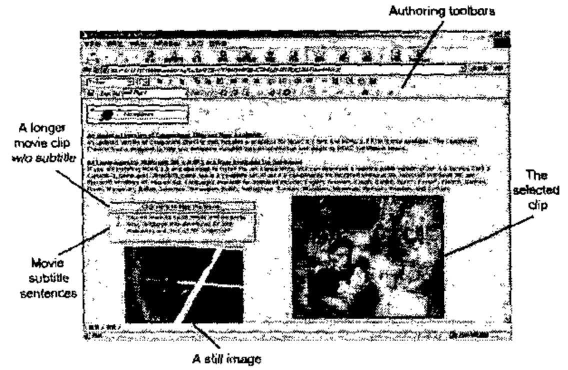

Figure 5. IWiLL web-based authoring tool

When teachers want to insert an image into the lecture, the methods described before are used. After the teachers compose all the materials including texts, images, movie clips and other learning components e.g. essay writing and discussion board [10][11], an online rich lecture is produced. Once the teachers assign the authored lectures to their classes, the students can study the multimedia lectures on the web and interact with teachers and their classmates in the same environment.

5. CONCLUSIONS AND FUTURE WORK

We have proposed a semantic multimedia database system in this paper. The designed system provides not only a new trend for indexing of multimedia database system, but also help end-users expand their query. In order to provide semantic multimedia retrieval features, we integrate keyword extraction, expansion and weighting schemes to construct extended index for the corresponding images. Moreover, on the purpose of helping users with convenient and efficient query, we also supply the users with query expansion. The proposed system has been applied in the IWiLL system, which is used by teachers and students over thirty schools and universities in Taiwan.

We will pursue our research further in the methodology of keyword expansion and query expansion and the evaluation of effectiveness of such an approach. These are the fundamental building blocks of the proposed system. Meanwhile, annotation contents also play a crucial role in our approach. A systematic and comprehensive way that guides an annotator to describe suitable keywords of the corresponding images is of benefit. Now, we are working on this issue.

ACKNOWLEDGEMENTS

The present work is partly supported by National Science Council Taiwan ROC under the contract No. NSC90-2213-E-032-013.

6. REFERENCES

[1] Bimbo, Alberto, "*Visual Information Retrieval,*" Morgan Kaufmann Publishers, 1999.

[2] Wasfi Al-Khatib, Y. Francis Day, Arif Ghafoor, and P. Bruce Berra, "Semantic Modeling and Knowledge Representation in Multimedia Databases," *IEEE Transactions on Knowledge and Data Engineering,* 1999, pp. 64-79.

[3] A. B. Benitez, J. R. Smith, and S. -F. Chang, "MediaNet: A Multimedia Information Network for Knowledge Representation," *Proceedings of IS&T/SPIE 2000 Conference on Internet Multimedia Management Systems*, Vol. 4210, Boston, MA, Nov. 6-8, 2000.

[4] Jun Yang, Liu Wenyin, HJ Zhang, YT Zhuang, "Thesaurus-Aided Image Browsing and Retrieval." *Proceedings of IEEE International Conference on Multimedia & Expo (ICME) 2001*, pp. 313-316.

[5] Shatford, Sara. "Analyzing the Subject of a Picture: A Theoretical Approach." *Cataloging & Classification Quarterly 6* (Spring 1986), pp. 39-62.

[6] WordNet, http://www.cogsci.princeton.edu/~wn/, 2002

[7] British National Corpus, http://www.hcu.ox.ac.uk/BNC/, 2002

[8] Lancaster, F. Wilfrid and Amy J. Warner. "*Information Retrieval Today,*" John Wiley & Sons, Arlington, Virginia, 1993, pp. 259-263.

[9] Church, Kenneth, William Gale, Patrick Hanks, and Donald Hindle. "*Using statistics in lexical analysis, in Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon,*" Uri Zernik, ed., Hillsdale, NJ, 1991, pp. 115-164.

[10] Kuo, Chin-Hwa, David Wible, Meng-Chang Chen, Li-Chun Sung, Nai-Lung Tsao, and Chia-Lin Chio. "Design and Implementation of an Intelligent Web-based Interactive Language Learning System," *Journal of Educational Computing Research*, Vol. 26, 2002.

[11] Kuo, Chin-Hwa, David Wible, and Nai-Lung Tsao. "On Designing a Web-based English Writing Environment and Learner Corpus," *IEICE Transactions on Information and Systems*, Vol. E84-D, No. 8, August 2001.