A Prototype System using Lexical Chains for Web Images Retrieval Based on Text Description and Visual Features

Dr. K. V. Ramana^{#1}, Chiranjeevi Palaparthi^{#2}, Mahaboob Basha Mula^{#3}, Satya Krishna Pepakayala^{#4} Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Kakinada

Kakinada, Andhra Pradesh, India –533003.

¹vamsivihar@gmail.com ²chiru.palaparthi@gmail.com ³bashajntuk@gmail.com ⁴satyakrishna.cse@gmail.com

Abstract--Content Based Image Retrieval, in the current scenario has not been analyzed adequate in the existing system. Here, we implement a prototype system for web based image retrieval. The system is based on description of images by lexical chains which are extracted from text related images in a web page. In this paper, we provide Relevance Feedback (RF) techniques that aim to the real world user requirements. The relevance feedback techniques, based on image text description are expanded to support image retrieval by combining textual and visual features. All the feedback techniques are implemented and compared with precision and recall criteria. The experimental results prove that retrieval methods that makes use of both text and visual features achieve overall better results than methods based only on image's text description.

Keywords--Relevance Feedback, Lexical Chain, Content Based Image Retrieval.

1. INTRODUCTION

Multimedia contents are growing rapidly and the need for multimedia retrieval is occurring more and more frequently in our day to day life. Earlier approaches to the content-based multimedia retrieval do not adapt the query and retrieval model based on the user's perception of the visual similarity [1, 8]. However, multimedia databases containing a large number of images with high precision is still an arduous manual task. Due to the complexity of multimedia contents, image understanding is difficult but interesting issue in this field. Image retrieval is becoming a province of increasing and essential importance in the present world, as part of Information Retrieval (IR) field. IR field is described as accurate and speedy access to a large amount of information which can be in the form of text documents, image collections, video or other multimedia objects. IR field has been a very productive for many researchers in the past years. As a result, a number of powerful image retrieval techniques have been proposed to deal with such problems. Content Based Image Retrieval (CBIR) is one of the current image retrieval systems. The relevance feedback technique from the IR domains is used in content based image retrieval [3, 9]. The main drawback of CBIR is the inability to present an image conceptually with a set of low level visual features such as color, texture, shape. To solve such problems a prototype system is being proposed in this paper to extract the images using the relevance feedback techniques through the lexical chains.

2. PROPOSED APPROACH

As stated above, the Image Retrieval field is vast but it is explored partially. A fair new idea of lexical chain approach is analyzed in this paper. The other issues that are discussed here include Image collection gathering process, image database scheme, indexing techniques and retrieval issues. The original weighted lexical chain model is proposed, analyzed and compared. Several up-to-date relevance feedback techniques are described in detail and fully implemented and their results are compared [4]. Furthermore, the system is expanded to combine textual with visual features focusing at a particular and rather narrow yet very interesting from the industries point of view, image retrieval problem. The difficulties that may arise in retrieving an image, such as weight assignment or image feature selection are described and analyzed since they may have an impact in the overall performance. Nevertheless, the combination of textual and visual features has been shown to achieve higher precision than text or image features alone.

2.1 Overview of Lexical Chains

A model referred as "lexical chains" is adopted for finding a more organized structure that can represent the image attributes selected and their semantic meaning more adequately. A lexical chain can be defined as a sequence of words semantically related to the image. A lexical chain might contain a few words to some sentences. These lexical chains are categorized as:

A Page Title Lexical Chain (PLC), an Image Title Lexical Chain (ITLC), an Alternate Text Description Lexical Chain (ADLC) and finally a Caption Lexical Chain (CLC) that includes the whole extracted image caption. Normally, only the CLC may exceed the limit of one sentence, since the image caption in some cases may include a few paragraphs. All the other lexical chains are usually confined to a few words or a single sentence; it doesn't match the semantic meaning.

Therefore, two other lexical chains that are constructed from the caption LC, since the caption is the only attribute that may include multiple sentences. The first one, called sentence lexical chain (SLC) represents a single sentence in the image caption. The second one, called reconstructed sentence lexical chain (RSLC) is a new sentence made up of two related sentences. By related sentences, we mean that they share atleast one common word. Once a common word is found, the sentences are split in two, the first half of the first sentence and the second half of the second sentence forming a reconstructed sentence LC. The remaining halves from the second RSLC.



Figure 1: Lexical chain representation of image text description

3. RELATED WORK

The process of obtaining relevant images and multimedia content from the World Wide Web is known as Web based image retrieval. Applications of web based image retrieval are

- Navigation of image collections
- Publishing and advertising
- Medicine and health related application domain
- Architectural and engineering design
- Crime prevention and legal issues

The above applications differ in search strategies and content representations. However they all adopt the following architecture.



Figure 2: Architecture of Web image retrieval system

From the World Wide Web, the web crawler will collect images and their respective HTML documents and pass them to text and image analyzer modules. The extracted data from the above modules undergo the indexing process and are usually stored in the databases. The user submits queries to the database which in turns replies by returning a set of results. The way the documents are presented, the features that are used to calculate the similarity between the user query and the web pages stored in the database and the representation of the results depend entirely on the implemented system. There are three main approaches to WWW image search and retrieval:

- Text based retrieval
- Content Based Image Retrieval (CBIR)
- Annotated image collection

3.1 Text based retrieval

Text based retrieval annotates images with text derived from the HTML documents. This approach is based on the observation that an image in a web page is semantically related to its surrounding text. This can include the image caption, the image file name, and the neighbor text around the image and or several other attributes. The extracted text is then almost always indexed and stored, represented in a specific schema that depends on the implementation of the system. The main idea behind text-based retrieval of images is that words or terms appearing at different locations of an HTML document have different levels of importance to the images. To improve query results, relevance feedback is often offered as the last part of such retrieval systems.

3.2 Content based image retrieval

Image analysis techniques are also to extract a variety of visual features from images. These include histograms, color, texture measurements, image dimensions, shape, orientation, moment invariant features etc. The extracted features are usually indexed and stored in systems database. The user interacts with the CBIR system via visual interface by issuing keyword queries, queries by image examples or queries combining keywords and image examples. Any multimedia object can form a query since the query interface is "query-by-example". Internally, like multimedia objects, a query is also represented as a collection of features. A user may use multiple objects as a query [2]. The objective is to find and retrieve images from the database that satisfy the user's criteria of similarity with the query.

3.3 Annotated image collection

For a diverse range of image consumers, several companies provide specialization in visual content. Image databases can be queried in several ways. Query-by-example is the most widely supported model in research prototypes and commercial products. In this environment, user formulates a query by means of giving an example image selected from a pool of general image categories [6]. The images are indexed and retrieved by user specified keywords and or query by example. These are manually assigned to each image or derived from proprietary techniques and algorithms. The image collections get updated periodically. While end users may use these services, they are especially geared towards companies and professionals who provide high volumes of diverse images.

3.4 Relevance feedback

Relevance feedback is the process of automatically adjusting an existing query using information fed-back by the user about the relevance of previously retrieved documents [3]. To evaluate the effectiveness of the relevance feedback process, it is necessary to compare the performance of the first iteration feedback search with the results of the initial search performed with the original query statements. Normally, recall (R) and precision (P) measures are used to reflect retrieval effectiveness, where recall is redefined as the proportion of relevant items that are retrieved from the collection and precision is the proportion of retrieved items that are relevant [5].

Precision is given by the formulae:

$$Precision (P) = \frac{|Relevant \cap Retrieval|}{|Retrieval|}$$

Recall is given by the formulae:

Recall $(R) =$	Relevant Retrieval
	Relevant

The peculiarity of the image retrieval problem needs other relevance feedback techniques rather than the classic ones. Nevertheless the same problems remain. The biggest one being in the subjectivity of the human perception: the way people perceive and judge is based on subjectivity that varies from person to person. The

other is the computer-centric issue, simply defined as the weight problem and the inability to combine high level concepts with user subjectivity.

Pseudo-relevance feedback (PRF), also known as blind feedback, is a technique commonly used to improve retrieval performance. Its basic idea is to extract expansion terms from the top-ranked documents to formulate a new query for a second round retrieval. Through a query expansion, some relevant documents missed in the initial round can then be retrieved to improve the overall performance [12]. Several pseudo relevance feedback techniques do not require any feedback from the user. But instead, they try to recalculate similarity according to share holding or other criteria. Yet, their performance is expected to be below the user-engaged technique. Nevertheless, most relevance feedback techniques depend on user feedback information, which is based on previously retrieved objects, in order to adjust an existing query to the user's preferences. Several techniques are described in detailed below.

3.5 Semantic accumulation

This method allows the user to pick the most relevant image from the results based on the lexical chains representation. In order to construct a new query, it gathers and accumulates the semantic information of the selected image. In addition to new terms, noise will be added into the query as well. Therefore instead of using the whole image, a single and a most related highest similarity lexical chain is used. The most related lexical chain is calculated from the list's similarity formulae.

$$Similarity = \frac{\begin{array}{c}list1.size\\ \Sigma \end{array} \sum \begin{array}{c}\lambda a_{p} * a_{q}\\ p=0 \end{array} = \frac{p=0}{\sqrt{list1.size}} * MatchScale} \\ \end{array}$$

The procedure for semantic accumulation is described below:

1.	Search using user query
2.	User selects feedback image
3.	Extract most related LC from the selected
	image.
4.	Merge query and lexical chain to obtain new
	query.
5.	Make search with new query (Step 1)

The above steps are represented in the form of a diagram and is shown below



Figure 3: Relevance feedback by semantic accumulation

The semantic accumulation approach has certain drawbacks. Query enrichment process might insert new images to the results. It may also insert many unrelated images in regards to the first query. In practical this method would either narrow the result to a very good set or widen them to a large set of unnecessary relevant images. There is practical no feedback at all since the search would return the same results as before.

3.6 Semantic integration differentiation

This approach is an improved version of the semantic accumulation technique. Selecting one image at a time is rather tedious and time consuming. User can select both relevant and irrelevant images. This technique integrates the relevant feedback images to construct a new query. Based on the feedback irrelevant sections, the

system differentiates the irrelevant images from the return results. In order to make new enriched query, the system extracts most semantically related lexical chains from each relevant image from the previously submitted query. In order to form a negative query, the system extracts the least semantically relevant lexical chain for the images marked as least relevant. Results returned from the new query are matched against the negative one and dropped if more similar to the negative lexical chains.

The procedure for semantic integration differentiation is described below:

- 1) User selects a number of relevant and irrelevant images
- 2) Extract the most related lexical chain from each relevant images
- 3) Combine query and extracted lexical chains to obtain a new query
- 4) Search using new query
- 5) Extract the most unrelated lexical chains from the irrelevant images and combine in a query-form lexical chain
- 6) For each image remove it from results if more similar to the lexical chain obtained from the irrelevant images.
- 7) Go to step 1

The above steps are represented in the form of a diagram and is shown below



Figure 4: Semantic integration differentiation

The main drawback in this technique is the fact that the query enrichment might have a negative effect, inserting too many new terms that might be irrelevant with the initial query. In the semantic accumulation feedback, there is absolutely no weight rearrangement at all. Instead of weight rearrangement, relevance feedback is achieved by expanding the results each time with related queries.

3.7 Rui-Huang relevance feedback technique

According to Rui, all images are treated as a multimedia objects with certain features while there may be more than one representation for one specific feature. For example if all relevant images have similar values for the color attribute then this attribute is a good indicator of the user's needs. If its values are very different then probably it is not a good indicator. Therefore its weight must be properly rearranged [11]. The inner weights are calculated as 1/deviation. The outer weights are calculated directly from the user judgment as the normalized sum of the user scores for the best NRT results. The steps described below, are initiated as

- 1) User marks some images
- 2) Obtain best results
- 3) Calculate inner, outer weights
- 4) Recalculate similarity score for all images
- 5) Go to step 2 or Exit

Outer weight assignment is done as in the algorithm below:

 $RT = [RT_1 \dots RT_i \dots RT_{NRT}]$ set threshold results.

 $RT^{j} = [RT_{1} \dots RT_{i} \dots RT_{NRT}]$ set threshold for j attribute

Sort results according to score for each attribute and select the NRT first. Therefore we have as many RTs as presentation (e.g. text or visual) that are considered in the total score calculation.

For all attributes

 $W_i^{\ j} = W_i^{\ j} + SCOREi$ If marked from user $W_i^{\ j} = W_i^{\ j} + 0$ If not marked from the user

Set $W_i^{\ j} = 0$ if $W_i^{\ j} < 0$ $W^T - Sum of all new weights$

 $W^{T} = \sum W_{i}^{j}$

Normalize all new weights $W_i^{\ j} = W_i^{\ j} / W^T$

After the weight adjustment process image similarity is calculated by the formula

$$\begin{split} & \textbf{Score}_{i} = \textbf{W}_{text}^{outer} * \textbf{Score}_{text} + \textbf{W}_{visual}^{outer} * \textbf{Score}_{visual} \\ & \textbf{Score}_{i} = \sum_{0}^{nr.attrib} (\textbf{W}_{j}^{inner} * \textbf{Score}_{j}) \end{split}$$

Where i: text or visual

j: number of attributes for visual text

Rui method is expected to have the best result since user preferences are quite detailed yet it depends heavily on the information provided i.e better results will appear if the user judges more and more images.

3.8 FALCON-Related Relevance Feedback

While the above feedback tries to calculate the ideal weights and try to find the ideal query, FALCON on a dissimilarity function that recalculates the score for all images in the initial set, based on user judgment. A new, dissimilarity measure [7] is calculated: the sum of the power of the distance between an image and the user provided images divided with the number of these good images. Therefore, similarities to the query are those images that have a small distance from the set of good points. Distance of a candidate image is considered as the difference between the image's score and the user selected good image.

A distance is calculated for all the images marked from the- user. The dissimilarity functions is given by

$$D_G = \frac{1}{K} * \sum_{i}^{k} d(x_j, g_i)$$

Where $x_j = Candidate$ image score

 $g_i =$ Score of user marked relevant image

d (x_i , g_i) dissimilarity function implemented as $x_i - g_i$

k - Number of user provided relevant images

The steps of the algorithm are described below:

1.	User chooses set of relevant images.
2.	Construct or expand the set of good points
3.	For all images calculate distance from good points
4.	Calculate dissimilarity for all images based on function
5.	Display
6.	User adds other images to the good set or Exit
7.	Go to step-2

The sum of ath power of this distance divided by the number of user selected images provides the dissimilarity score of candidate image. The set of good point is expected to change since user might add other relevant images. Therefore the dissimilarity score for each image changes as well. Here we must note that the changes of the value of a parameter are expected to change results as well.

3.9 Pseudo Relevance Feedback

In all the above methods, the information provided from the user is of crucial importance for the whole feedback process. Nevertheless, the normal user sometimes may find it rather disturbing to mark and evaluate even some of the results. Therefore some pseudo relevance feedback techniques are called to fill the gap. Pseudo meaning

that if no information at all is available from the user then the system itself is called to rearrange the results by some techniques: shareholding, clustering, automatic weight recalculation etc.

3.10 K-Means Clustering

The same principles of the well known clustering algorithm can be applied as a pseudo relevance feedback technique considering the fact that we expect the upper part of the results to be quiet relevant and the lower part rather irrelevant. We can apply the K-means algorithm [10] in order to try and cluster the results into good, don't care and irrelevant to the query. The number of centers may vary, but in our implementation we adopt a 3center approach. We can define some initial center for good, don't care and irrelevant and run the algorithm several times. In order to improve result we have certain options like displaying only the good results, or the good and don't care etc.

The algorithm is given below:

1.	Find center through thresholding or use
	given centers
2.	For each result include in the category

- whose center is closest and adjust the center properly.
- 3. Go through clustering again with derived center.
- 4. Display one or more categories.

4. CONCLUSION

In this paper, we implemented a fairly new idea of image retrieval based on text description and visual features. At last as mentioned, it is shown that the incorporation of text, visual features and user preferences is surely to provide very high precision and recall criteria. Rui's algorithm, as discussed provides a very efficient and user friendly way to improve the results. In the future there are some remaining issues to investigate, the size of or database and the plethora of images in it guarantee that the results are very good and the retrieval speed of this method is rather low making it quite interesting for further research and implementations.

5. REFERENCES

- [1] D.H. Kim and C.W. Chung, "QCluster: Relevance Feedback Using Adaptive Clustering for Content Based Image Retrieval," Proc. ACM SIGMOD, pp. 599-610, 2003.
- K. Porkaew, K. Chakrabarti, and S. Mehrotra, "Query Refinement for Multimedia Similarity Retrieval in MARS," Proc. ACM [2]
- Int'l Multimedia Conf. (ACMMM), pp. 235-238, 1999.
 Y. Rui, T. Huang, and S. Mehrotra, "Content Based Image Retrieval with Relevance Feedback in MARS," Proc. IEEE Int'l Conf. Image Processing, pp. 815-818, Oct.1997. [3]
- Y. Rui, T. Huang, M. Ortega, and S. Mehrotra, "Relevance Feedback: A Powerful Tool for Interactive Content Based Image [4] Retrieval," IEEE Trans. Circuits and Systems for Video Technology, Vol. 8, no. 5, pp. 644-655, Sept.1998.
- G. Salton and C. Buckley, "Improving Retrieval Performance by Relevance Feedback," J. Am. Soc. Information Science, Vol. [5] 41, no. 4, pp. 288-297, 1990.
- K. Vu, K.A. Hua, and N. Jiang, "Improving Image Retrieval Effectiveness in Query-by-Example Environment," Proc. 2003 [6] ACM SYMP. Applied Computing, pp.774-781, 2003.
- [7] L. Wu, C. Faloutsos, K. Sycara, and T. R. Payne, "FALCON: Feedback Adaptive Loop for Content-Based Retrieval," Proc. 26th Int'l Conf. Very Large Database (VLDB), pp. 297-306, 2000.
- R. Burnelli and O. Mich, Image retrieval by examples, IEEE Transactions on Multimedia, 2(3): 164-171, September 2000 [8]
- G. Salton and M. J. McGill, introduction to Modern Information Retrieval. McGraw-Hill Company, 1983. [9]
- Wilkin G.A, Xiuzhen Huang, "K-Means Clustering Algotithms: Implementation and Comparison," Vol., no., pp 133-136, Aug [10] 2007
- [11] C. Buckley and G. Salton, "Optimization of relevance feedback weights," in Proc. Of SIGIR'95.
- Xiangji Huang, Yan Rui Huang, Miao Wen, Aijun An, Yang Liu, Poon J, "Applying Data Mining to Pseudo-Relevance [12] Feedback for High Performance Text Retrieval," Data Mining, Vol., no., pp 295-306, Dec. 2006.