**Military Technical College**
**Kobry El-Kobbah,**
**Cairo, Egypt**

**7th International Conference**
**on Electrical Engineering**
**ICEENG 2010**

# Content-based Search for Effective Image Retrieval

*By*

Hsiao T. Pao* Hsin C. Fu**

***Abstract:***

In this paper we present a content-based approach for effectively searching and retrieving images. The proposed content-based image query and retrieval (CBQR) method uses two or more types of query for accessing images --- textual annotation associated with images and visual appearance, such as color, texture and positional features of objects in sample images. A user places a keyword-based query first and then retrieves desired images by visual content-based query. A prototype CBQR system was implemented and is available at http://140.113.216.66/WebImageSearch for online demo, and public evaluation. We also conducted experiments over a categorized Corel image collection and a non-categorized WWW image collection to show the performance of the proposed CBQR method.

***Keywords:***

Image query and retrieval, Text search, Content-based, Content-based query, Model-based search

* Dept. of management science, National Chiao-Tung University, Hsinchu, Taiwan, ROC. *htpao@cc.nctu.edu.tw*

** Dept. of computer science, National Chiao-Tung University, Hsinchu, Taiwan, ROC. *hcfu@cs.nctu.edu.tw*

## *1. Introduction:*

The dramatic growth of digital media at home, in enterprises, and on the web, has over the last decade spawned great interests in developing methods for effective indexing and searching of desired visual contents to open the value of these contents. Conventional text-based search techniques have been widely used in commercial search engines over large content corpora, such as WWW. However, using text-based search techniques on non-textual unstructured content, such as image and video data, is not nearly as mature or effective as on text documents. In fact, this kind of approach works fairly well for retrieving images with text annotation such as named entities, e.g., specific people, objects, or places. However, it does not work well for generic topics related to general settings of objects, as the text annotation rarely describes the background setting or the visual appearance, such as color, texture, shape, and size of the objects. Because of these and other limitations, it is now apparent that conventional text search techniques on their own are not sufficient for effective image and/or video retrieval, and they need to be combined with techniques that consider the visual features of the content. Thus, we need a seamlessly integrated paradigm of text-based and visual content-based query and retrieval in a unified interface.

## *2. Related works:*

The methods for query and retrieve relevant information from multimedia contents (primarily images) can be divided into two main categories:

1) ***Text-Based Methods***: The text surrounding multimedia objects is analyzed and the system extracts those that appear to be relevant. Shen et al. [2] explore the context of web pages as potential annotations for the images in the same pages. Srihari et al. [3] propose extracting named entities from the surrounding text to index images. The major constraint of text-based methods is that it requires the presence of high quality textual information in the surrounding of the multimedia objects. In many situations, this requirement may not be satisfied.
2) ***Image-Based Methods***: More methods focus on extracting semantic information directly from the content of multimedia objects. Wang et al. [4] proposed SIMPLIcity, a system that captures semantics using the robust Integrated Region Matching metric. The semantics are used to classify images into two broad categories which are then used to support semantics-sensitive image retrievals. Recently, Goh et al. [5] propose a confidence-based dynamic ensemble (CDE) to use local and global perceptual features to annotate images. CDE can make dynamic adjustments to accommodate new semantics, to assist the discovery of useful low-level features, and to improve class-prediction accuracy. However, if images have neither dominant regions nor common visual features, these methods

may probably fail to conclude an acceptable result, i.e., a semantic meaning. The ability to select one or a few dominant region(s) with semantic meaning is where our proposed approach differs from these methods.

## ***3. Proposed method and system:***

Our approach to content-based image query and retrieval relies on a seamlessly integrated paradigm of text-based and object-based query and retrieval in a unified interface. An important feature of the proposed method is to allow a user to query image objects without consciously segmenting the objects that should be matched during the image search. Rather, the user can simply choose desired objects with a few clicks of a mouse button. Color and texture characteristics of the nearby area around the mouse cursor are collected as partial features of the query objects. In addition, positional relationships among the selected objects are also included in query features. All the selected query objects and features can be combined with one another using Boolean operations such as union, intersection, and exclusion. In other words, image objects are considered as keywords in document search. Now, how do we extract most or all the candidate objects in an image to form an index of key-objects, which will be used to match with query objects during image search? The proposed query object is mainly a region of close colors and textures; the key-objects of an image can be extracted in the following manner. The proposed key-object extracting method is to partition an image into regions, in which each region contains pixels of close colors and textures. A close color region means the color difference between each pair of pixels in the region is smaller than a pre-determined threshold. The threshold can be a variety of values, such that partitioned color regions can be formed to contain pixels of different colors' closeness, to meet different query objects' needs. The region of close textures can be organized in a similar manner [6].

An overall diagram of the proposed CBQR system is depicted in Figure 1(a). The CBQR system contains three modules: (1) Image search servers, (2) Meta servers, and (3) CBQR web server. The Image search servers periodical search WWW for new images, retrieve and extract visual features of new images, and then index these visual features with the URLs of the new images. Finally, the metadata dispatcher sends the collected information of new images to one of Meta servers. Each Mata server contains a Meta database, which saves received image indexing information, and a Meta agent, which performs maintenance and content searching of the Meta database. The CBQR web server is used to provide a user at client side for query images and to receive visual features from the user. Based on the user's query image and selected regions of interest, the query dispatcher allocate a Meta server to service the user for retrieving and passing matched images.

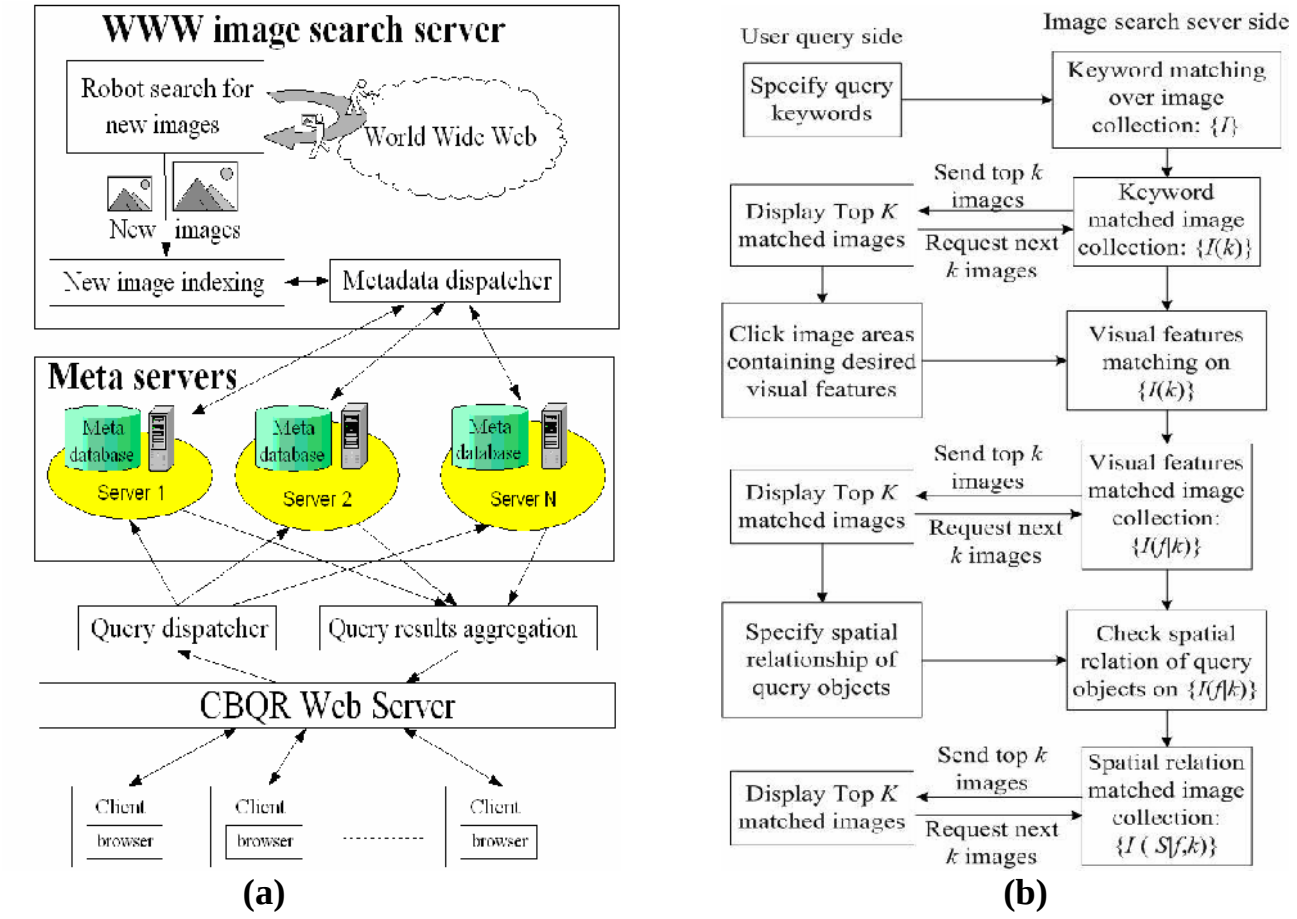



***Figure (1): (a)*** *The overall diagram and* ***(b)*** *the flowchart of the proposed content-based image query and retrieval* ***(CBQR)*** *system.*

Figure 1(b) depicts the flowchart of the proposed CBQR approach, which shows how a user may specify one or several keywords to query and retrieve a set of relevant images from an image collection. Figure 2 illustrates retrieval results of queries on keywords (1) *horse* and (2) *white horse and brown horse* respectively. However, neither of these results is satisfactory, since the keyword based retrieval returns images containing keywords in their surrounding textual information but not showing keyword associated images. As we can see that the keyword *horse* is too broad for the given query topic, while the keywords *white horse and brown horse* return a few images containing white and brown horses but most of them not showing horses as per user's desired color or texture. Then, the user may browse the retrieval images to find horses with desired color and/or texture as visual query objects for secondary search and retrieval.

As shown in Figure 3, a user who wants to retrieve pictures presenting "a white horse to the right of a brown horse grazing in the lush green pastures" may enter search queries by clicking on: (1) the white horse on the right, (2) the brown horse on the left, and (3)

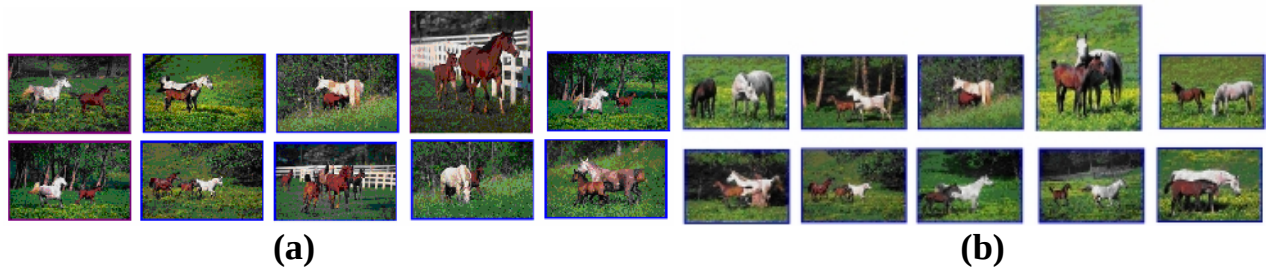
the green pasture at the bottom (not shown in Figure 3) of one or several given sample images. Figure 4(a) then shows that the combination of the previous two methods returns a more relevant set of results by prioritizing keywords retrieval matches that are also visually consistent with color and texture features.

**(a)** **(b)**

***Figure (2):*** **(a)** *Results of text retrieval using horse as keyword.* ***(b)*** *Retrieval results of multiple keywords white horse and brown horse.*

Further more, the object-based query may include objects' spatial relationship as additional visual features. Figure 4(b) illustrates that the retrieval results of Figure 4(a) can then be further refined and improved by integrating the location relationship associated with color and texture features as query object features.

***Figure (3):*** *Query by object-based model: for a brown horse on the left, and for a white horse on the right.*



**(a)** **(b)**

***Figure (4):*** ***(a)*** *Results of visual based retrieval using white and brown colors and textures as query objects;* ***(b)*** *Adding spatial relationship to color and texture features for white and brown horses query.*

## ***4. Experimental results:***

In order to adding more surrounding textual information of retrieved images, a background process is always performed at the end of each image query and retrieval. For the retrieved images with very high confidence, the user's initial query words will be assigned to these images as their associated keywords. In addition, for the retrieved images with medium high confidence, a user assisted keyword assignment interface is used to further increase the surrounding keywords of these images.
The CBQR prototype system is available at *http://140.113.216.66/WebImageSearch* for online demo and public evaluation. The CBQR prototype system contains two image collections: (1) a set of 5000 images collected from 10 categories of Corel photo gallery, and (2) a set of 20,000 images randomly collected from WWW. Thereafter, these two image collections are called Corel5k and WWW20k respectively. The 10 categories in Corel5k are *butterfly*, *bus*, *elephant*, *flower*, *building*, *dinosaur*, *mountain*, *Africa*, *beach*, and *food*. Text title and brief description may or may not be available for the images inWWW20k. Two types of experiments were conducted to evaluate the performance of CBQR approach. The first type of experiment intends to show the retrieval performance of CBQR approach on different categories of images. The performance was evaluated according to the averaged image retrieval accuracy versus a sequence of queries, which are applied in the following order: keyword, color, texture, and spatial relationship of ROI. The accuracy [7] is defined as the ratio of relevant images in the top *T* retrieved images. The averaged accuracy is simply the average of the accuracies measured for the 1600 randomly selected test queries.

***Experiment I-1*** — queries on a particular object or category: Query and retrieval experiments were conducted on *butterfly*, *bus*, *elephant*, *flower*, *building*, and *dinosaur* images from Corel15k and WWW20k image collections. The retrieving performance is shown in Table 1.
***Experiment I-2*** — query categories without a dominant object or common visual features: This type of query and retrieval experiments were conducted on categories, like *mountain*, *Africa*, *beach*, and *food* from Corel15k and WWW20k image collections. The retrieving performance is shown in Table 2. The second type of experiments intends to show the performance comparison with respect to the results in [1]. To evaluate the performance of CBQR approach, we use the averaged accuracy versus query iterations, (Iter. #). The iterations refer to the query sequences as in Experiment I for CBQR approach, and refer to the iterations of relevance feedback as in [1].

***Experiment II-1*** — query a particular object or category: We conducted this type of experiments by querying and retrieving six categories of images, *butterfly*, *bus*,

*elephant*, *flower*, *building*, and *dinosaur* from Corel15k and WWW20k image collections. The image retrieval experimental results of query on butterfly are shown in Table 3. For the query on butterfly, the CBQR achieves 85% and 77% of accuracies over the Corel15k and WWW20k image collection respectively. By repeating four cycles of user relevant feedback, Hsu [1] achieves an 80% of accuracy on a similar type of query on the *butterfly* category, with over 9400 images selected from a Corel photo gallery. A more interesting comparison would be made if the CBQR retrieval with comparable image collection were available.

***Experiment II-2*** — query on background scene: It is very important for an image query system to have capability to retrieve relevant images according to the query on background scenes. Thus, we conducted experiments to retrieve four different types of images, such as *mountain*, *Africa*, *beach*, and *food* from both Corel15k and WWW20k image collections. The experimental results of query on mountain are shown in Table 3. Similar to Experiment II-1, CBQR approach is slightly superior to the GBR-P-S approach on the Corel image collection, and is inferior to the GBR-P-S approach on the WWW image collection. Notice that the retrieval accuracy of these experiments do not imply that the performance of CBQR approach is superior or inferior to the GBR-P-S approach [1], since the MIRQ approach uses regions or object-based visual features directly from users' query targets, which usually reflect user's desires.

***Experiment II-3*** — query on combined objects or regions: This experiment shows query on a certain category, which has neither dominant regions or objects, nor common visual features. For instance, *kitchen* can be a representative example in this category. We conducted query and retrieval for *kitchen* relevant images over both Corel15k and WWW20k image collections. Although a kitchen image has neither dominant regions nor common visual features, a kitchen in the real world usually contains several different kitchen utensils and furniture, such as a microwave oven, a refrigerator, a dishwasher, and kitchen chairs etc. Thus, the CBQR approach allows a user to select a few kitchen objects or regions as user's queries for retrieving images associated with kitchen category. The GBR-P-S approach achieves an 18% of accuracy on the kitchen category retrieval over their Corel image collection [1]. As shown in Table 3, CBQR approach achieves 75% and 70% of accuracies from querying and retrieving the Corel15k and WWW20k image collections respectively.

From three different experiments, the proposed CBQR approach achieves comparable accuracy performance as the current leading approaches [1]. However, the experimental results tell us several things. They are: (1) the CBQR approach has the capability of searching and retrieval relevant images from a large collection, such as the WWW20k in less than one second, (2) the CBQR approach can retrieve relevant images closely associated with users' query regions or objects, (3) the MIRQ approach allows users to combine several queries by using Boolean operations to organize a category, which has neither dominant regions or objects, nor common visual features, such as kitchen,

laboratory, etc.

***Table (1):** Average retrieval rates of Experiment I-1 for searching on Corel5k and WWW20k (values for WWW20k is printed in bold face).*

| Image Cat. | Keywords | Color | Texture | Spatial relation |
|---|---|---|---|---|
| butterfly | 63/**49**% | 70/**56**% | 80/**65**% | 88/**72**% |
| bus | 35/**23**% | 39/**26**% | 44/**31**% | 49/**35**% |
| elephant | 37/**21**% | 38/**25**% | 41/**30**% | 42/**32**% |
| flower | 39/**28**% | 45/**34**% | 55/**43**% | 63/**50**% |
| building | 28/**21**% | 35/**27**% | 46/**33**% | 55/**40**% |
| dinosaur | 86/**37**% | 90/**42**% | 95/**47**% | 99/**52**% |

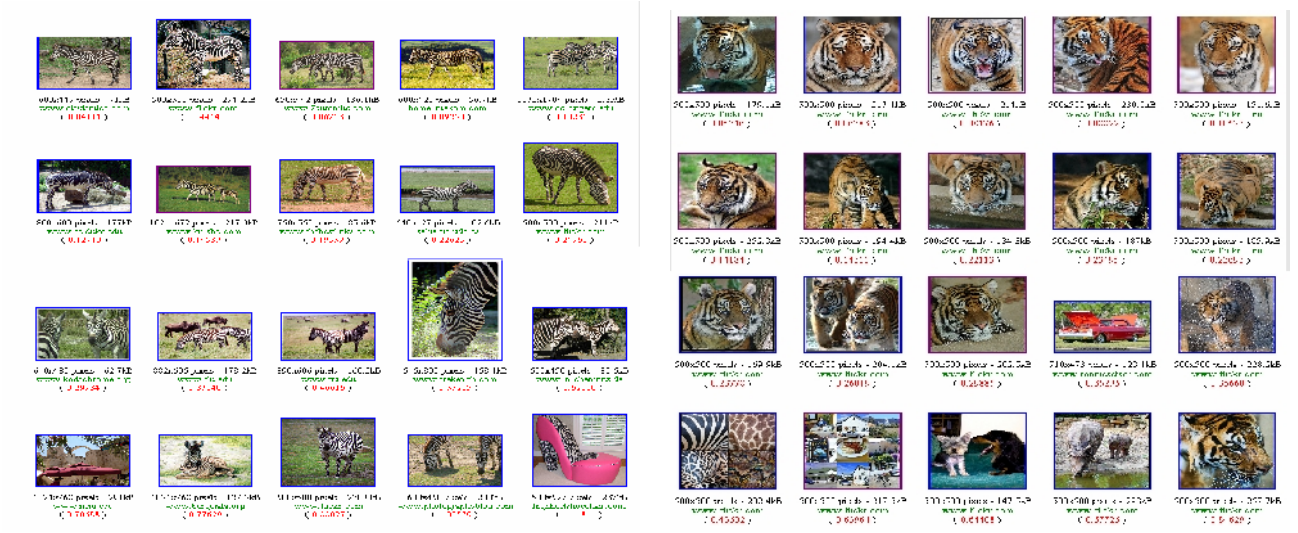
***Table (2):** Average retrieval rates of Experiment I-2 for searching on Corel5k and WWW20k (values for WWW20k is printed in bold face).*

| Image Cat. | Keywords | Color | Texture | Spatial relation |
|---|---|---|---|---|
| mountains | 34/**23%** | 44/**28%** | 54/**34%** | 59/**40%** |
| Africa | 39/**20%** | 45/**29%** | 54/**42%** | 62/**55%** |
| beach | 21/**19%** | 25/**22%** | 30/**26%** | 33/**28%** |
| food | 26/**13%** | 35/**19%** | 47/**27%** | 58/**32%** |

***Table (3):** Performance comparison of image query and retrieval of butter y, mountain and kitchen relevant images among CBQR Corel5k, CBQR WWW20k and GBR-P-S9400. The accuracy is measured as the ratio of relevant images in top 20 retrieved images.*

| Accuracy of retrieved *butterfly* images | | | | |
|---|---|---|---|---|
| | Iter. #1 | Iter. #2 | Iter. #3 | Iter. #4 |
| CBQR Corel5k | 62% | 71% | 74% | 83% |
| CBQR WWW20k | 46% | 55% | 63% | 66% |
| GBR-P-S9400 | 55% | 65% | 70% | 80% |
| Accuracy of retrieved *mountain* images | | | | |
| CBQR Corel5k | Iter. #1 | Iter. #2 | Iter. #3 | Iter. #4 |
| CBQR WWW20k | 34% | 44% | 51% | 57% |
| CBQR WWW20k | 23% | 28% | 32% | 36% |
| GBR-P-S9400 | 40% | 50% | 55% | 55% |
| Accuracy of retrieved *kitchen* images | | | | |
| CBQR Corel5k | Iter. #1 | Iter. #2 | Iter. #3 | Iter. #4 |
| CBQR WWW20k | 19% | 50% | 75% | NA |
| CBQR WWW20k | 16% | 47% | 71% | NA |
| GBR-P-S9400 | 18% | 17% | 16% | 15% |

For visual inspection, retrieval results of two query examples are illustrated in Figure 6. The query image of each example is at the upper-left corner of each image set, and the rest images are the top 20 query results. As we can see, most of query results match with the query images in the cases of zebra and tiger.



**(a)** **(b)**

***Figure (5):*** The retrieval results of the proposed **CBQR** system by using **(a)** a zebra and **(b)** a tiger as query objects.

## ***5. Conclusions and future works:***

Figures 3-5 and Tables 1-3 validate the combination queries that for an image retrieval system to achieve a satisfactory level of performance, it needs to effectively combine multiple modalities and retrieval methods. Among the images collected from WWW, we see quiet a few images are illustrated for the representation or summary of their associated videos. Therefore, a user may carefully select search queries to retrieve desired images and associated videos from WWW. One of the possible future works of the proposed CBQR approach is to collect more images from the WWW, such that users may be able to query and retrieval more desired images and/or videos.

## ***References:***

[1] C.T. Hsu and C.Y. Li, “Relevance feedback using generalized bayesian framework with region-based optimization learning,” IEEE Trans. on Image Process., vol. 14, no. 10, pp. 1617–1631, Oct. 2005.

[2] Heng Tao Shen, Beng Chin Ooi, and Kian-Lee Tan, "Giving meanings to www images," in MULTIMEDIA '00: Proceedings of the eighth ACM international conference on Multimedia, New York, NY, USA, 2000, pp. 39–47, ACM.

[3] Rohini K. Srihari, Zhongfei Zhang, and Aibing Rao, "Intelligent indexing and semantic retrieval of content-based documents," Inf. Retr., vol. 2, no. 2-3, pp. 245– 275, 2000.

[4] James Ze Wang, Jia Li, and Gio Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture LIbraries," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 9, pp. 947–963, 2001.

[5] Kingshy Goh, Beitao Li, and Edward Y. Chang, "Semantics and feature discovery via confidence-based ensemble," ACM Trans. Multimedia Comput. Commun. Appl., vol. 1, no. 2, pp. 168–189, 2005.

[6] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," IEEE Trans. on Systems, Man, and Cybernetics, vol. SMC-8, no. 6, pp. 460–473, 1978.

[7] Z. Su, H. Zhang, S. Li, and S. Ma, "Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning," IEEE Trans. on Image Process., vol. 12, no. 8, pp. 924–937, Aug. 2003.

***Acknowledgement***

This research was supported in part by the National Science Council under grant NSC95-2221-E009-218 and in part by the Microsoft under grant FY08-RES-THEME-092.