

Content-based Search for Image Retrieval

Mohamed M. Fouad
Department of Computer Engineering, Military Technical College
Kobry Elkoppa, Cairo, Egypt
mmafoad@ieee.org

Abstract— In this paper, a content-based image retrieval approach is presented for effective searching. The proposed approach uses two or more types of query for accessing images, textual annotation associated with images and visual appearance, such as colour, texture and positional features of objects in sample images. One can first place a keyword-based query, and then the desired images are retrieved by visual content-based query. The proposed retrieval approach shows clear improvements over competing approaches in terms of retrieval accuracy and visual inspection using Corel gallery and WWW images.

Index Terms— Content-based, image retrieval and query, text search

I. INTRODUCTION

In the last decade, the growth of digital media, at home, in enterprises, and on the web, has spawned great interests in developing methods for effective indexing and searching for desired visual contents. Conventional text-based search approaches have been widely used in commercial search engines over large content corpora, such as WWW. However, using text-based search approaches on non-textual unstructured content, such as image and video data, is not nearly as mature or effective as on text documents. In fact, such approaches work fairly well for retrieving images with text annotation, such as named entities (*e.g.*, specific people, objects, or places) [1]. However, those approaches don't 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 colour, texture, shape, and size of the objects. Because of these and other limitations, it is now apparent that conventional text search approaches on their own are not sufficient for effective image and/or video retrieval. As well, they need to be combined with other approaches that consider the visual features of the content. Thus, a seamlessly integrated technique of text-based and visual content-based query and retrieval is needed in a unified scheme.

The approaches of query and retrieving relevant information from multimedia contents (primarily images) can be divided into two main categories:

1) *Text-based approaches:*

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 limitation of the text-based approaches is its need to the presence of high quality textual information in the surrounding of the multimedia objects.

2) *Object-based approaches:*

These approaches focus on extracting semantic information directly from the content of multimedia objects. Wang et al. [4] propose 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 in order to help extract useful low-level features, and to improve class-prediction accuracy. However, if images have neither dominant regions nor common visual features, such approaches may fail to conclude an acceptable result.

In this paper, we propose an image retrieval approach that combines the text-based concept with the object-based query and retrieval under one umbrella. The proposed approach is able to select one or a few dominant region(s) with semantic meaning compared to the approaches above. The rest of this paper is structured as follows. In Section II, the proposed approach is presented. In Section III, we show the experiments and results. Finally, conclusions are given in Section IV.

II. THE PROPOSED APPROACH

The proposed approach of 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 approach 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. Colour 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 most or all the candidate objects in an image are extracted 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 colours and textures. So, 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 colours and textures. A close colour region means the colour 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 colour regions can be formed to contain pixels of different colours' closeness, to meet different query objects' needs. The region of close textures can be organized in a similar manner [6].

Fig. 1 shows an overall diagram of the proposed content-based image query and retrieval approach; referred to as CIQR that constitutes of three modules: i) image search servers, ii) Meta servers, and iii) CIQR web server. The image search servers periodically search for WWW 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 Meta 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 CIQR 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 allocates a Meta server to service the user for retrieving and passing matched images.

Fig. 2 depicts the flowchart of the proposed CIQR 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. Fig. 3 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 the keyword associated images. As one can notice 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 colour or texture. Then, one may browse the retrieval images to find horses with desired colour and/or texture as visual query objects for secondary search and retrieval.

As shown in Fig. 4, an user, who wants to retrieve images presenting "a white horse to the right of a brown horse grazing in the lush green pastures", may enter search queries by clicking on: i) the white horse on the right, ii) the brown horse on the left, and iii) the green pasture at the bottom (not shown in Fig. 4) of one or several given sample images. Fig. 5-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 colour and texture features.

Furthermore, the object-based query may include objects' spatial relationship as additional visual features. Fig. 5-b illustrates that the retrieval results of Fig. 5-a can then be further refined and improved by integrating the location relationship associated with colour and texture features as query object features.

III. EXPERIMENTS AND RESULTS

In order to add 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 confidence, the keyword assignment interface, with user help, is used to further increase the surrounding keywords of these images. The proposed CIQR approach contains two image collections: i) a set of 5000 images collected from 10 categories of Corel photo gallery, and ii) 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 in WWW20k. Two types of experiments were conducted to evaluate the performance of CIQR approach. The first type of experiment intends to show the retrieval performance of CIQR 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 region of interest (ROI).

The retrieving performance of the proposed approach is evaluated by determining the ratio of relevant images in the top k retrieved images [7]. We'll refer to this ratio as the retrieving accuracy (RA). Hence, the average retrieving accuracy (ARA) can be simply defined as the average of the accuracies measured for the 1600 randomly selected test queries.

Experiment I-1: queries on a particular object or category:

This sort of experiments uses *butterfly*, *bus*, *elephant*, *flower*, *building*, and *dinosaur* images from Corel15k and WWW20k image collections. The corresponding RA is shown in Table I.

Experiment I-2: query categories without a dominant object or common visual features:

This sort of query and retrieval experiments uses categories, such as *mountain*, *Africa*, *beach*, and *food* from Corel15k and WWW20k image collections. The corresponding RA is shown in Table II. These experiments are to show the effectiveness of the proposed approach compared to the results in [1]. To evaluate the performance of the CIQR approach, the ARA is then used versus query iterations, (*i.e.*, Iter. #). The iterations refer to the query sequences as in Experiment I for the CIQR approach, and refer to the iterations of relevance feedback as in [1].

Experiment II-1: query on a particular object or category:

In this experiment, only six categories of images are queried and retrieved, such as *butterfly*, *bus*, *elephant*, *flower*, *building*, and *dinosaur* from Corel15k and WWW20k image collections. The corresponding RA of query on *butterfly* is reported in Table 3. Shown is that the CIQR achieves 85% and 77% of accuracies over the Corel15k and WWW20k image collections, respectively. By repeating four iterations 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 CIQR 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, CIQR 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 does not imply that the performance of CIQR approach is superior or inferior to the GBR-P-S approach [1], since the CIQR 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 CIQR 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 III, CIQR approach achieves 75% and 70% of accuracies from querying and retrieving the Corel15k and WWW20k image collections respectively. From three different experiments, the proposed CIQR approach achieves comparable accuracy performance as the current leading approaches [1]. However, the experimental results tell us several things, i) the CIQR approach has the capability of searching and retrieval relevant images from a large collection, such as the WWW20k in less than one second, ii) the CIQR approach can retrieve relevant images closely associated with users' query regions or objects, and iii) the CIQR 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.

For visual inspection, retrieval results of two query examples are illustrated in Fig. 6 and Fig. 7 for a zebra and a tiger as query objects, respectively. 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 both cases.

IV. CONCLUSIONS

In this paper, a hybrid framework of a combination queries for an image retrieval system is introduced to achieve a satisfactory level of performance. The proposed approach combines multiple modalities and retrieval methods. Among the images collected from

WWW, one can see a quite few images that are illustrated for the representation or summary of their associated videos. Therefore, a user may select search queries to retrieve desired images and associated videos from WWW. The results show that the proposed approach achieves clear improvements in terms of retrieving compared to competing approach.

TABLE I: RETRIEVING ACCURACY (%) OF EXPERIMENT I-1 FOR SEARCHING ON I) COREL5K AND II) WWW20K.

Image	Keywords		Colour		Texture		Spatial Relation	
	I	II	I	II	I	II	I	II
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 II: RETRIEVING ACCURACY (%) OF EXPERIMENT I-2 FOR SEARCHING ON I) COREL5K AND II) WWW20K.

Image	Keywords		Colour		Texture		Spatial Relation	
	I	II	I	II	I	II	I	II
Mountain	34	23	44	28	54	34	59	40
Africa	39	20	45	29	55	42	62	53
Beach	21	19	25	21	30	26	33	28
Food	26	13	35	19	47	27	58	32

ACKNOWLEDGMENT

Many thanks for anonymous referees for their constructive comments.

REFERENCES

- [1] C.T. Hsu, C.Y. Li. Relevance feedback using generalized Bayesian framework with region-based optimization learning. *IEEE Trans. on Image Processing*, 2005, 14(10):1617–1631.
- [2] Heng Tao Shen, Beng Chin Ooi, Kian-Lee Tan. Giving meanings to www images. *The 8th Intern. Conf. on Multimedia*, 2000, 39–47.
- [3] Rohini K. Srihari, Zhongfei Zhang, Aibing Rao. Intelligent indexing and semantic retrieval of content-based documents. *Information Retrieval*, 2000, 2(3):245–275.

- [4] James Ze Wang, Jia Li, Gio Wiederhold. SIMPLiCity: Semantics-sensitive integrated matching for picture Libraries. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001, 23(9):947–963.
- [5] Kingshy Goh, Beita Li, Edward Y. Chang. Semantics and feature discovery via confidence-based ensemble. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2005, 1(2):168–189.
- [6] H. Tamura, S. Mori, T. Yamawaki. Texture features corresponding to visual perception. *IEEE Trans. on Systems, Man, and Cybernetics*, 1978, SMC-8(6):460–473.
- [7] Z. Su, H. Zhang, S. Li, S. Ma. Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning. *IEEE Trans. on Image Processing*, 2003, 12(8):924–937.

TABLE III: PERFORMANCE COMPARISON OF IMAGE QUERY AND RETRIEVAL OF BUTTERFLY, MOUNTAIN AND KITCHEN RELEVANT IMAGES AMONG CIQR COREL5K, CIQR WWW20K AND GBR-PS [1].

	Iter. #1	Iter. #2	Iter. #3	Iter. #4
butterfly images				
CIQR Corel5k	62%	71%	74%	83%
CIQR WWW20k	46%	55%	63%	66%
GBR Corel5k [1]	55%	65%	70%	80%
mountain images				
CIQR Corel5k	34%	44%	51%	57%
CIQR WWW20k	23%	28%	32%	36%
GBR Corel5k [1]	40%	50%	55%	55%
kitchen images				
CIQR Corel5k	19%	50%	75%	82%
CIQR WWW20k	16%	47%	71%	75%
GBR Corel5k [1]	18%	17%	16%	20%



Mohamed M. Fouad, received the Bachelor engineering (honors, with great distinction) and Masters' engineering degrees from the Military Technical College (MTC), Cairo, Egypt, in 1996 and 2001, respectively. As well, he received the Ph.D. degree in Electrical and Computer engineering from Carleton University, Ottawa, Ontario, Canada, in 2010. He is currently a faculty member with the Department of Computer Engineering, MTC. His research interests are in online handwritten recognition, image registration, image reconstruction, image retrieval, super-resolution, video compression, and multiview video coding.

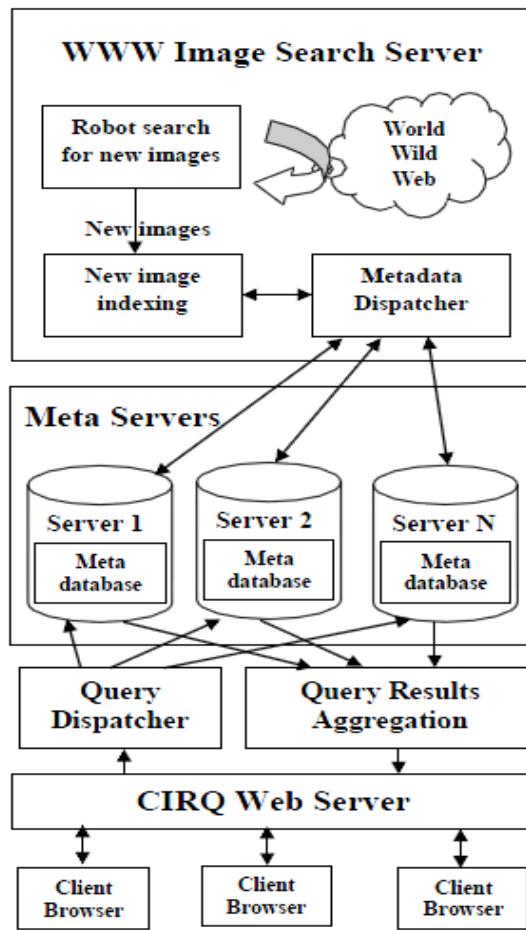


Figure 1. Framework of the proposed content-based image query and retrieval (CIQR) approach.

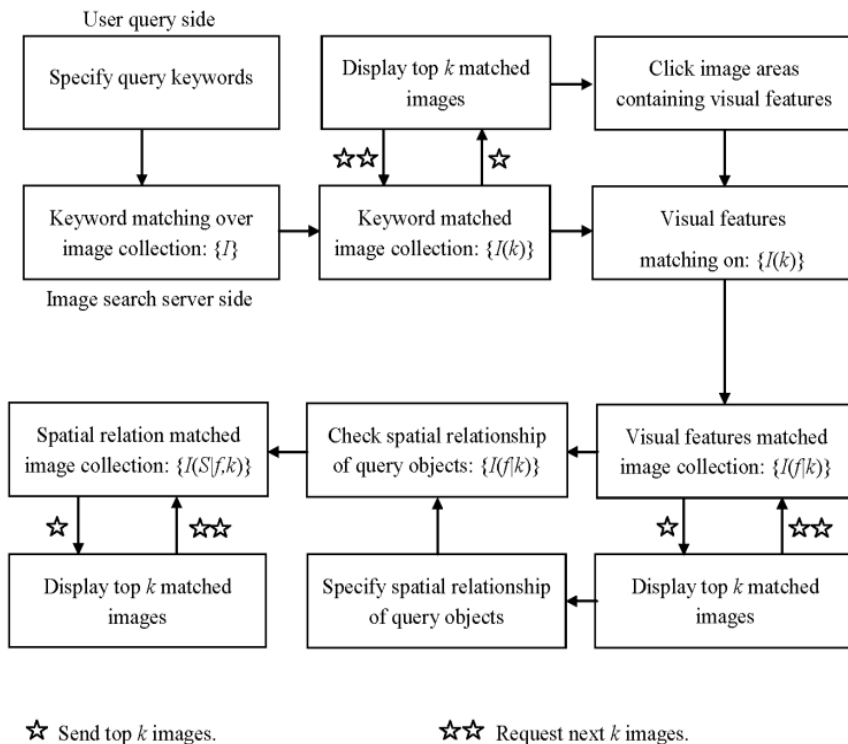


Figure 2. Flowchart of the proposed content-based image query and retrieval (CIQR) approach.



Figure 3. Results of text retrieval using (a) one keyword: horse, (b) multiple keywords: white horse and brown horse.



Figure 4. Query by object-based model: for a brown horse on the left, and for a white horse on the right.



Figure 5. a) Results of visual based retrieval using white and brown colours and textures as query objects, b) Adding spatial relationship to colour and texture features for white and brown horses query.

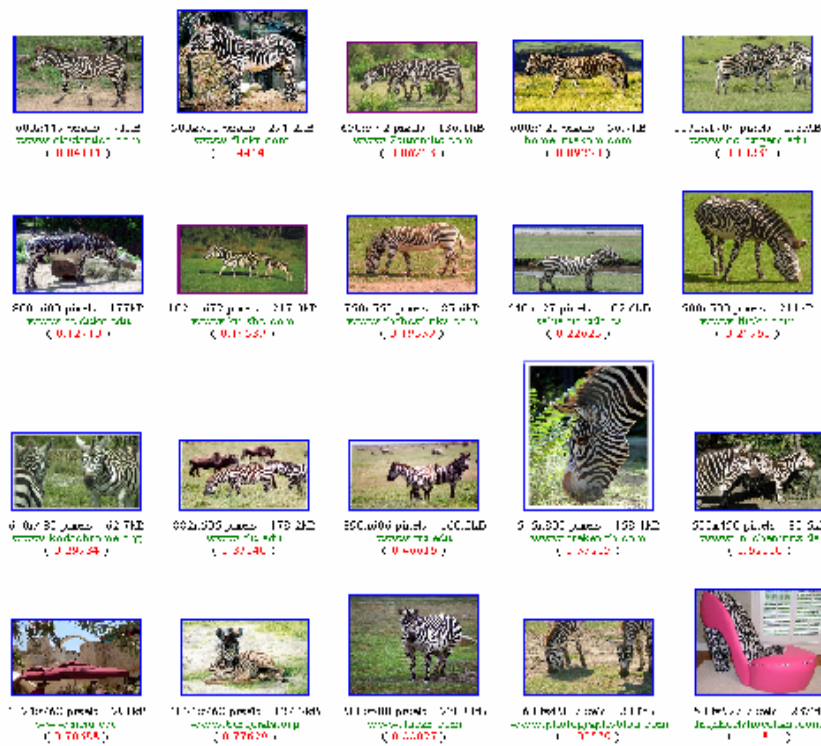


Figure 6. The retrieval results of the proposed CIQR system using a zebra as query object.

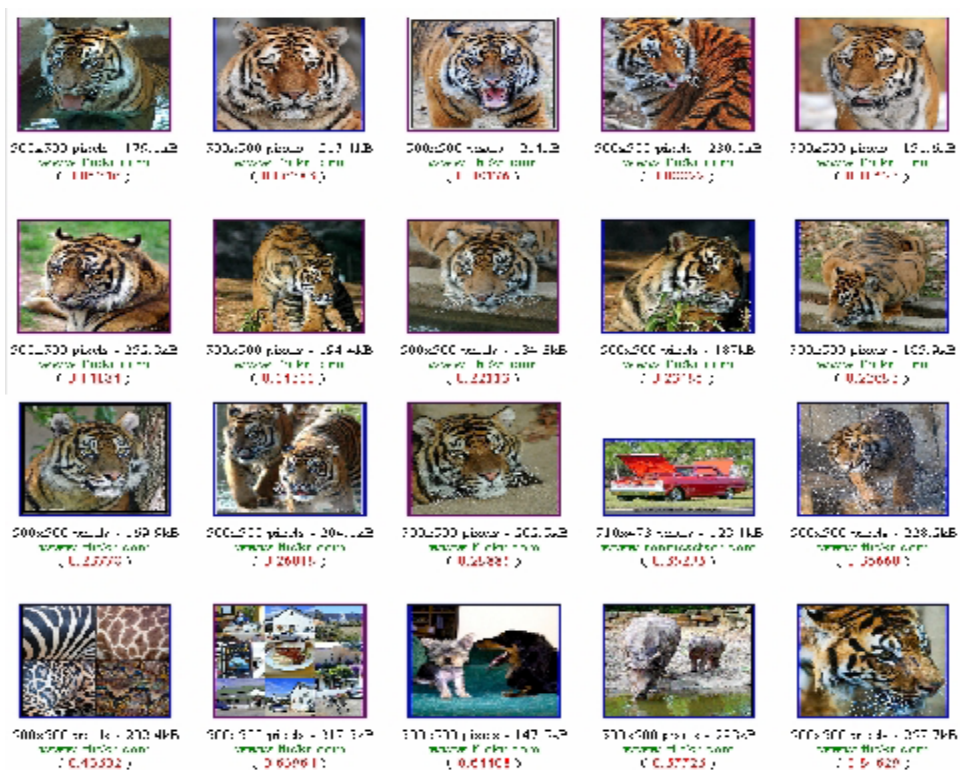


Figure 7. The retrieval results of the proposed CIQR system using a tiger as query object.