KEYWORD AND IMAGE CONTENT FEATURES FOR IMAGE INDEXING AND RETRIEVAL WITHIN COMPRESSED DOMAIN

Irianto, Y. Suhendro

Department of Informatics, Darmajaya Informatics and Business Institute
Jalan Z.A Pagar Alam 93, Bandar Lampung, Lampung
E-mail: Suhendro@darmajaya.ac.id

ABSTRACT: The central problem of most Content Based Image Retrieval approaches is poor quality in terms of sensitivity (recall) and specificity (precision). To overcome this problem, the semantic gap between high-level concepts and low-level features has been acknowledged. In this paper we introduce an approach to reduce the impact of the semantic gap by integrating high-level (semantic) and low-level features to improve the quality of Image Retrieval queries. Our experiments have been carried out by applying two hierarchical procedures. The first approach is called keyword-content, and the second content-keyword. Our proposed approaches show better results compared to a single method (keyword or content based) in term of recall and precision. The average precision has increased by up to 50%.

Keywords: CBIR, high level, low level features, recall, precision.

INTRODUCTION

Most of Content-based Image Retrieval [1] approaches are aimed to find images that are semantically similar to a given query (often a single example image). In this definition, *semantically similar* is meant in the sense of human visual perception and usually refers to *high-level* features. However, the methods used to satisfy this demand are generally based on numerical feature extraction and metric-based or distance measurement. This approach usually refers to *low-level* features. Now, the problem of most (general-purpose) CBIR approaches is low quality in terms of recall and precision, and as the reason the semantic gap has been proposed [2].

Currently, there are two distinct approaches or methods adopted in image retrieval, namely: content-based and text-based approaches [3]. Content-based image retrieval (CBIR) uses low-level visual features to retrieve images. With this approach, it is unnecessary to annotate images and translate users' queries. However, due to the semantic gap between image visual features and high-level concepts [4], it is still hard to use a CBIR system to retrieve images with correct semantic meanings. Integrating textual information may help a CBIR system to cross the semantic gap and improve retrieval performance.

Recently many approaches have been used to combine text or keyword and content-based techniques for image retrieval. A simple method to conduct with text and content-based retrieval separately and merge the retrieval results has been introduced by the authors [5, 6]. In contrast to the parallel approach, a pipeline approach uses textual or visual information to perform initial retrieval, and then uses the other information to filter out the irrelevant images [7]. In these two approaches, textual and visual queries are formulated by users and do not directly influence each other. Another approach is transformation-based which tries to mine the relations between images and text, and uses the mined relations to transform textual information into visual one, and vice versa [8].

This paper proposes two integrated methods; the first method applies keywords as the first step and applies image content as the second step (simply called *K-C*). The second method applies content as the first step and employs image keyword as second step (simply called *C-K*). The methods are used in such as way that the image keyword and content to be considered as filter to one another to get the best image retrieval performance. The rest of the paper is structured as follows. Section 2 outlines relevant related work, section 3 describes the experimental design, and section 4 presents our experimental results. In section 5 we conclude our experiments, and finally we plan our future work in section 6.

RELATED WORKS

Many existing image retrieval systems are text or keyword based. The solution historically has been to develop text-based ontologies and classification schemes for image description. Text-based indexing has much strength including the ability to represent both general and specific examples of an object at varying levels of complexity. Unfortunately most of the images frequently have little or no accompanying

textual information such as keyword, caption, and/ or description. Some literature reviews pertaining to text-based approaches include [9, 10, 11, 12].

The problems with text-based access to images have promptly increased interest in the research and development of image content based approaches. The approach is referred to as content-based image retrieval (CBIR). Content-based image retrieval relies on the characterization of primitive features such as colour, shape, and texture that can be automatically extracted from images. Commercial CBIR systems that have been used include IBM's Query by Image Content (QBIC) described by [13], Virage's VIR Image Engine [14], and *Excalibur's Image Retrieval (vrw.excalib.com). On the Web, CBIR image retrieval systems include WebSEEK [15], Informedia (*informedia.cs.cmu.edu), *and Photobook (*vismod. www.media.mit.edu).

Retrieving images based on color similarity is achieved by calculating a color histogram for each image that identifies the proportion of pixels or DCT coefficients within an image holding specific values. Many researches are attempting to segment color proportion by region and by spatial relationships among several color regions [16, 17]. Another content based feature is texture but this is not easy concept to represent. The identification of specific textures in an image is achieved mainly by modeling texture as a two-dimensional gray level variation. The problem here is in identifying patterns of co-pixel variation and associating them with particular classes of textures such as *silky* or *rough*. The author of [18] have extended work in this area through the development of a texture thesaurus that matches texture regions in images to words representing texture attributes.

Queries for shapes are generally achieved by selecting an example image provided by the system or by having the user sketch a shape. The primary mechanisms used for shape retrieval include identification of features such as lines, boundaries, aspect ratio, and circularity, and by identifying areas of change or stability via region growing and edge detection. Although shape, color and texture are undoubtedly important visual features for image representation, there is still little understanding of the best way to implement these attributes for image retrieval. An understanding of what constitutes similarity for image retrieval purposes is also needed. The technology for content-based image retrieval is still in its infancy and opens to research. The focus to date has been primarily on the use of features that can be computationally acquired, but little has been done to identify the visual attributes needed by users for various tasks and collections. Research on shape based image retrieval has been done by many authors [19, 20, 21].

EXPERIMENTAL DESIGN FARMAKOKINETIK

The first stage of content based image retrieval is to extract low-level image features either from entire images or from image regions. We consider *color moments* to represent the color distribution in the HSV colour space. In order to describe the visual content of an image, a visual feature is computed. This feature should be an optimal encoding of key visual features representing generic, low-level information about the image: colours, shapes, textures, and patterns. In this experiment I calculate the distance between image query and image in the database to construct an indexing key by using this equation:

$$d(h_q, h_k) = (\sum_{i=1}^{N} \frac{|hi|}{64}), k \in [6,000]$$

$d(h_q, h_k)$ is the distance between image query (q) and image (k) in the database, it is considered as default method of image indexing. If the image retrieved is exactly the same as the image query, the distance will be zero.

The system has approximately 360 keywords identified for the images which are adopted from WorldNet (wordnet.princeton.edu) for 6,000 images. Each image was visually examined and all its relevant keywords identified. All the embedded keywords in each image describe the image visualization. For example, an image of a bear in the jungle will have keywords such as bear, grass, trees, animal, river, and water attached in it. We create at least 5 keywords in every image by using WordNet.

The major goal of experiments is to evaluate the performance of integrated methods, keyword-content (K-C) and content-keyword (C-K) for image retrieval in the compressed domain. In addition, this experiment is also proposed to investigate and compare the performance of a single method (content or keyword solely) along with the integrated methods. The experiments employ 6,000 images which consist of 10 classes or categories including *bear, motorbike, car, cat, building*, *flower, mountain, model (artist), sky, and texture.*

The experiments can be categorized in two approaches, the first called keyword-content (*K-C*) and the second called content-keyword (*C-K*). The keyword-content approach is carried out when a user composes a query by using keyword as the first step. If the images retrieved are satisfactory (more than 90 % of the images retrieved are appropriate for the query), they are considered the final result, and precision and recall of the first 50 images retrieved are calculated. Otherwise, the images retrieved are stored in a temporary database and queried using content features. An Overview of its experimental design can be shown in Figure 1.
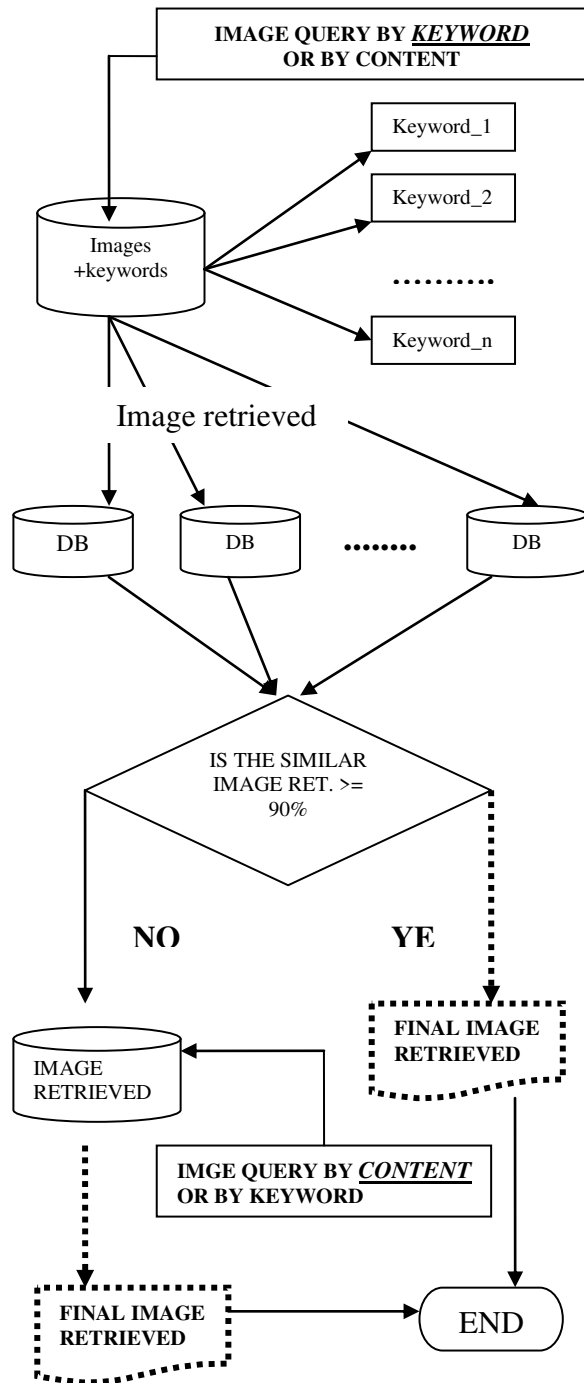
Figure 1. Overview of the system, Query by Keyword-Content (K-C) and Query of Content-Keyword (C-K)

The content-keyword method first makes on image content or visual query into the image database. The images retrieved are calculated their precision and recall based on the first 50 rank images. If the images retrieved are satisfactory, they are considered as the final result, and the precision and recall of the first 50 images retrieved are calculated. Otherwise, all the images retrieved in this are placed in a temporary database which is queried using keyword.

A user composes a keyword query by using words such as bear, river, animal, claws, and carnivore and expects bear images to be retrieved. In each image class, five keywords are used which making a total of 50 queries. We rank only the first 50 images retrieved to calculate the precision and recall, regardless of the number of the images retrieved. Meanwhile in applying content features as a query, we use the query-by-example (QBE) technique with the database. In content based image retrieval we construct 5 images as image queries for each class which compose a total of 50 queries.

An example of image retrieval by keyword using Standard Query Language MySQL can be presented as:

SELECT `image_id` , `image_type`, `contents`,
`image_sizes` , `image_categorys` ,
`image_names` , `keyword`
FROM `IMAGE_DB`
WHERE 1 AND `keyword`
LIKE '%sky%' LIMIT 0, 30

Where *IMAGE_DB* is the image database of 6,000 images and *sky* is the keyword. If the images retrieved are satisfactory, the precision and recall are calculated as the final result of image retrieval. Otherwise, the images retrieved are stored in a temporary database and will be queried using image content.

THE EXPERIMENTAL RESULTS

Like most previous studies on most content based image retrieval, I calculated the effectiveness of the system in term of *precision* and *recall*. Precision or specificity is the ratio of the number of relevant images retrieved to the number of images retrieved, whilst recall or sensitivity is the ratio of the number of relevant images retrieved to the number of relevant images in the database.

The average precision of image retrieval by keyword is 65%, furthermore the precision of image retrieval by content as second step of the keyword-content approach show as high as 75% and 24% for the precision and recall, respectively. From the experiments found that *texture* and *model* classes have the highest precision at 98 % and 100%, respectively, whilst the bear images have the lowest precision of 24%.

Experiment showed that keyword based retrieval on texture images gave precision of 100%. In this case, system did not apply content feature as a second step. Another results showed that five classes have improved in precision when the keyword-content method was carried out, specifically images are bear

(24 % to 84%), car (44% to 49%), cat (43% to 68%), flower (60% to 87%), and model or artist (89% to 98%). Further detail can be seen in Figure 4.

The average precision of keyword based retrieval is 65%, whilst keyword-content improves slightly to 75%. This means keyword-content has increased precision by 10%. Bear images demonstrated the highest in increased of precision value which was 60%. Further detail of keyword-content method performance can be seen Figure 2. The C-K method improved precision and recall compared to content feature based retrieval. As shown in Figure 3, the average precision and recall of content–keyword based are 75% and 31%, respectively. Meanwhile, content based retrieval presents 27% and 78% for precision and recall, respectively.

Precision has been increased dramatically by content-keyword method on model images (from 23% to 100%). In general, for all the images the performance of content-keyword method increases up to 50% compared to content based retrieval in term of precision and recall. More detail can be seen in Figure 5. Excellent result has been demonstrated by model and texture images in which their precision increase from 23 % to 100 % and 47% to 100%, respectively. The lowest amount increase in precision is shown by flower and car which are 17% to 40% and 42% to 66%, respectively.
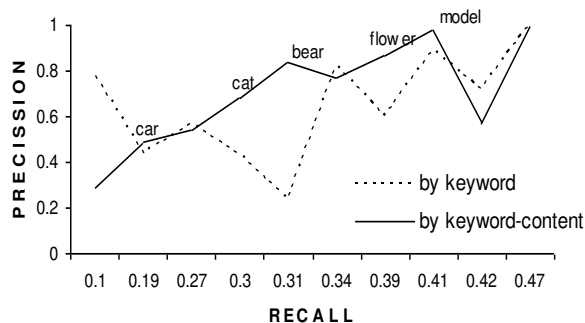

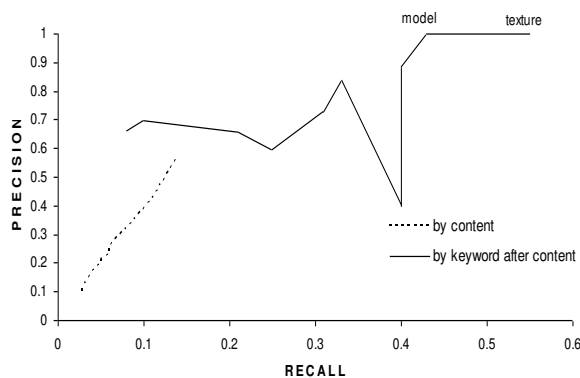**Figure 2. Keyword–Content Based Image Retrieval (applying keyword then content)**

Figure 3. Content-Keyword (C-K) Based Image Retrieval (applying content then keyword)
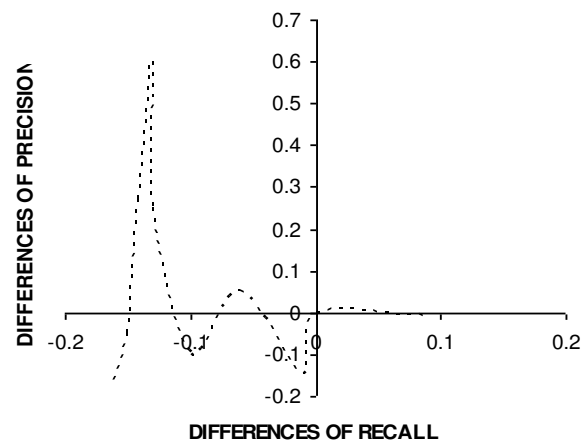

Figure 4. Plot the difference of precision and difference of recall calculated for keyword and keyword-content approach.
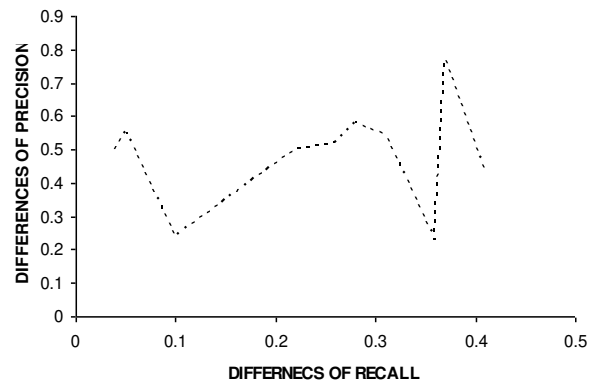

Figure 5. Plot the difference of precision and difference of recall calculated for content and content-keyword approach.
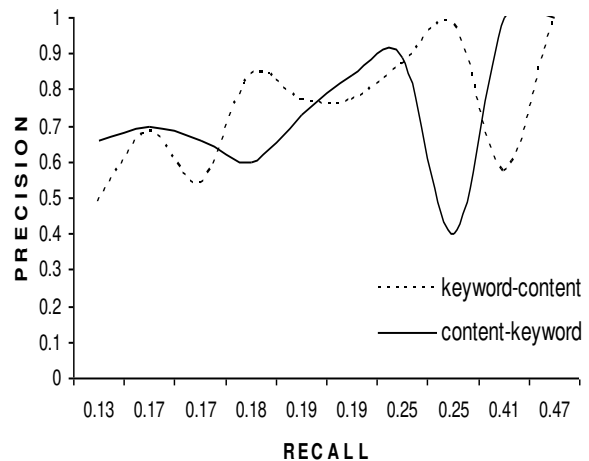

Figure 6. Keyword-Content and Content-Keyword

Interesting result has been found that keyword-content and content-keyword base retrieval present no significant different in average precision which are 74.8% and 75.1%, respectively. Figure 6 provides more detail about our integrated methods. However in

keyword-content based retrieval, the keyword based retrieval as the first step demonstrates 65%. While in content-keyword based retrieval, during the content based retrieval as the first step shows that precision at 26%.

CONCLUDING AND REMARKS

This paper shows the usefulness of proposed methods of integrated keyword and content for image retrieval. By using the first five keywords in the images, the content-keyword (C-K) system has increased the average precision from 65% to 75%. Whilst, keyword-content (K-C) shows no significantly difference in average precision which is 74.8% and 75.1%, respectively. The content-keyword method also demonstrates an excellent precision of 100 % for *model (artist)* and *texture* images.

FUTURE WORK

My Future works in conducting with integrated methods of image retrieval in compress domain will be how to apply keyword and content features in such manner so that user could query by using a keyword and an image example at the same time, other word implementing keyword and content features method in serial. However the future system will automate the decision step.

In the near future ontology will be used to determine specific relationships among features in images that can be used effectively in classifying those images. This ontology provides the pattern description for image annotation construction. It also provides the vocabulary and background knowledge describing features of the image objects.

REFERENCES

1. Del Bimbo, A., 1999, *Visual Information Retrieval*, Morgan Kaufmann Publisher, San Francisco, CA.
2. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., and Jain, R., 2001, Content-Based Image Retrieval at the End of the Early Years, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22nd, pp. 1349-1380.
3. Goodrum, A.A., 2000, Image Information Retrieval: An Overview of Current Research, *Information Science*, 3(2). pp. 63-66.
4. Eidenberger, H. and Breiteneder, C., 2006. Semantic Feature Layers in Content-based Image Retrieval: Implementation of Human World Features, *Proceedings of International Conference on Control, Automation, Robotic and Vision.*
5. Besançon, R., Hède, P., Moellic, P.A., and Fluhr, C., 2005. Cross-Media Feedback Strategies: Merging Text and Image Information to Improve Image Retrieval, 5th *Workshop of the Cross-Language Evaluation Forum, CLEF 2004*, LNCS 3491, pp.709-717.
6. Jones, G.J.F., Groves, D., Khasin, A., Lam-Adesina, A., Mellebeek, B., and Way, 2005, Multilingual Information Access for Text, Speech and Images, *5th Workshop of the Cross-Language Evaluation Forum*, CLEF 2004, LNCS 3491. Springer-Verlag GmbH, pp. 653-663.
7. Baan, J., van Ballegooij, A., Geusenbroek, J.M., den Hartog, J., Hiemstra, D., List, J., Patras, I., Raaijmakers, S., Snoek, C., Todoran, L., Vendrig, J., de Vries, A., Westerveld, T., and Worring, M., 2001, Lazy Users and Automatic Video Retrieval Tools in (the) Lowlands, *Proceedings of the Tenth Text Retrieval Conference (TREC 2001),* National Institute of Standards and Technology, pp. 159-168.
8. Lin, W.C., Chang, Y.C. and Chen, H.H., 2005, Integrating Textual and Visual Information for Cross-Language Image Retrieval, *Proceedings of the Second Asia Information Retrieval Symposium.*
9. Rasmussen, E., 1997, Indexing images, *Annual Review of Information Science and Technology*, 32, 169-196.
10. Lancaster, F., 1998, *Indexing and abstracting in theory and practice,* 2nd edition, Library Association, London.
11. Ruthven, I., 2008, Interactive Information Retrieval, *Annual Review of Information Science & Technology*, Vol. 12, Issue 1, pp. 43-91.
12. Cawkell, A., 1993, Indexing collections of electronic images: A review, *British Library Research Review,* p. 15.
13. Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Qian Huang, Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P., 1995, Query by image and video content: the QBIC system, *IEEE Computer,* 28(9), pp. 23-32.
14. Bach, J.R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Jam, R., and Shu, C.F., 1996, The Virage image search engine: an open framework for image management, *Storage and Retrieval for Image and Video Databases IV*, Proc SPIE 2670, pp. 76-87.
15. Kherfi, M.L., Ziou, D., and Benardi, A., 2004, Image Retrieval from the World Wide Web: Issues, Techniques, and Systems, *ACM Computing Surveys*, Vol. 36, Issue 1, pp. 23-35.
16. Shi, Y., Ding, Y., Zhang, R., and Li, J., 2009, Structure and Hue Similarity for Color Image Quality Assessment, Electronic Computer Tech-

nology, 2009 International Conference on, Macau, pp. 329-333.

17. Carson, C., Belongie, S., Greenspan, H., and Malik, J., 1997, Region-based image querying, *Proceedings of IEEE Workshop on Content-Based Access to Image and Video Libraries*, San Juan, Puerto Rico, pp. 42-49.

18. Ma, W. and Manjanath, B., 1998, Netra: A toolbox for navigating large image databases, *Proceedings of IEEE International Conference on Image Processing* (ICIP97), 1, pp. 568-571.

19. Ma, L., Zhou, Q., Chelberg, D., and Celenk, M., 2004, Shape-based Image Retrieval with Rele-vance Feedback, *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 779 -782.

20. Lin, H.J., Kao, Y.T., Yen, S.H., and Wang, C.J., 2008, A Study of shape-Based Image Retrieval, *Proceedings of the 24th International Conference on Distributed Computing Systems Workshops (ICDCSW'04).* 0-7695-2087-1/04.

21. Xu, X., Lee, D.J., Antani, S., and Long, L.R., 2006, Pre-Indexing for Fast Partial Shape Matching of Vertebrae Images, *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems* (CBMS'06) 0-7695-2517-1/06.