Combination of Keyword and Visual Features for Web Image Retrieval System

Nyein Myint Myint Aung

***Abstract*—This paper presents the implementation of web image retrieval system using keyword-based search and visual image features. In order to correctly correlate terms to a web image, the associated text of the web image is partitioned into text blocks according to the structure of the text with respect to the web images. Then, keywords are extracted and stored in image indexing database which will later be used in keyword based retrieval. When user enters keyword, result images are generated by image indexing and searching algorithms. Most of the web image search systems are based on only keyword based searches. But most of the result images in the keyword search are not relevant to the query. To reduce the influence of those irrelevant images, visual image features are used. Firstly, image features are extracted and then they are stored in the image indexing database. And then these features are used to cluster images for relevant and non-relevant. Combination of keyword search and visual feature extraction will result in producing more relevant images by removing non-relevant images. Extracting visual features can improve the system performance.**

***Index Terms*—Image retrieval system, keyword-based search, visual image features.**

I. Introduction

Web images have been becoming one of the most important information types on the Web. Thus, how to effectively gather, manage and reuse this valuable resource are among the most attractive research topics in the area of Web information retrievals. Image search is a specialized data search used to find images. To search for images, a user may provide query terms such as keyword, image file/link, or click on some image, and the system will return images "similar" to the query. The similarity used for search criteria could be meta tags, color distribution in images, region/shape attributes, etc. In Image Meta Search system , the search of images are based on associated metadata such as keywords, text, etc. Content-based image retrieval (CBIR) aims at avoiding the use of textual descriptions and instead retrieves images based on similarities in their contents (textures, colors, shapes etc.) to a user-supplied query image or user-specified image features [1].

Most of the web image search systems are based on only keyword based searches. Those systems use surrounding blocks of text to index the corresponding images. However, no standard work exists as how to correlate text blocks to web images.

Most works agree upon that the image name, ALT tag of the image, title of the Web page, and the close text are important to index the corresponding Web image. But the scopes of the close texts of the web images are different from work to work. On the other hand, content-based image retrieval (CBIR) has been introduced and developed to support image search based on visual features. Although these features could be extracted from images automatically, they are not accurate enough to represent the semantics of images. So, this paper presents the combination of visual feature-based model with keyword-based model.

The rest of the paper is organized as follows. Section II reviews related work. Section III discusses the framework of the proposed web image retrieval system. Section IV presents our evaluations and Section V concludes the paper.

II. Related Work

Most of the web image search systems are based on only keyword based searches. Those systems use surrounding blocks of text to index the corresponding images. Users can search web images just in the same ways as for general web search engines. Web images, even with close visual features, may have great differences in their semantics.

In [2] describe content-based web image retrieval system. In [3] presented web image search system based on web links. In this system, two web images are supposed to be similar if they are co-cited by many web pages. It is actually extended from some well known link-based web page schemes to the context of web images retrieval.

In “Image Retrieval by Hypertext links” by [4], links are used to track back source pages of the image container page. Then, container page and its source pages are used together to index the corresponding image. In [5] proposed the development of a world wide web image search engine that crawls the web collecting information about the images it finds, computes the appropriate image decompositions and indices, and stores this extracted information for searches based on image content. This approach avoids the search time problem of labeling. Results of above systems are a bunch of images which may include non-relevant images.

A relevance feedback scheme for both keyword and visual feature-based image retrieval is proposed by [6]. For each keyword, a statistical model is trained offline based on visual features of a small set of manually labeled images and used to propagate the keyword to other unlabeled ones. Besides the offline model, another model is constructed online using the user provided positive and negative images as training set. Support vector machines (SVMs) in the binary setting are adopted as both offline and online models. For this approach, labeling of image is necessary and it may effect the relevance of the image to query.

Manuscript received September 10, 2013; revised December 15, 2013.

Nyein Myint Myint Aung is with University of Technology, Yatanarpon Cyber City, Myanmar (e-mail: thae.thae.star@gmail.com).

DOI: 10.7763/IJIET.2014.V4.456

III. Framework of the Proposed Wed Image Retrieval

The framework of the proposed web image retrieval is shown in Fig. 1. As can be seen from Fig. 1, the system consists of 3 components: the pre-processing component, image feature extraction component and indexing component, and the image retrieval component.

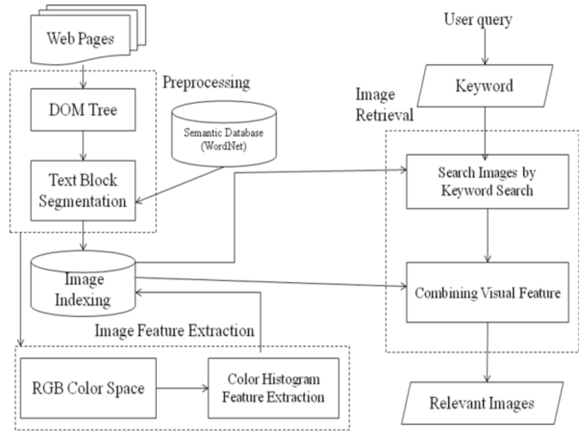

Fig. 1. Framework of the proposed web image retrieval.

A. Preprocessing Step

In the first step, web page is represented as a DOM tree, with nodes as HTML tags. The HTML DOM views an HTML document as a tree-structure. The tree structure is called a node-tree. All nodes can be accessed through the tree. Their contents can be modified or deleted, and new elements can be created [7].

The node tree below shows the set of nodes, and the connections between them. The tree starts at the root node and branches out to the text nodes at the lowest level of the tree. The nodes in the node tree have a hierarchical relationship to each other. The terms parent, child, and sibling are used to describe the relationships. Parent nodes have children. Children on the same level are called siblings (brothers or sisters).

- In a node tree, the top node is called the root
- Every node has exactly one parent node, except the root which has no parent
- A node can have any number of children
- A leaf is a node with no children
- Siblings are nodes with the same parent

Merge all the child nodes under a parent node bottom up recursively until there is no words similarity between those child nodes or another web image is found under this node. Then, web page will be partitioned into blocks and the image is indexed using the terms of text block which contains that image. In this paper, semantic similarity between terms is computed in order to get overall semantic relevance between text blocks. Vectors are prepared for texts within blocks as in vector space model. Semantic similarity between terms is computed as follows:

This formula is proposed by Wu & Palmer, the measure takes into account both path length and depth of the least common sub-summer [8]:

$$sim(s,t) = \frac{2 \times depth(LCS)}{depth(s) + depth(t)} \qquad (1)$$

where,

s and *t*: denote the source and target words being compared.

Depth(*s*): is the shortest distance from root node to a node S on the taxonomy where the synset of S lies .

LCS: denotes the least common sub-summer of *s* and *t*.

Vector Space Model

Vector space model or term vector model is an algebraic model for representing two texts as vectors. The distance between the two vectors is an indication of the similarity of the two texts. The cosine of the angle between the two vectors is the most common distance measure. Normally, TF-IDF values are used as weights in vector space model. In this system, weights are filled with the multiplication of TF-IDF and semantic similarity values computed in above Equation (1). Cosine similarity algorithm is computed as in Equation (2), where weights filled from vector space model are applied into the Equation (2).

$$sim(X1, X2) = \frac{\sum\left(tx1_j, tx2_j\right)}{\sqrt{\sum t^2 x1_j}, \sqrt{\sum t^2 x2_j}} \qquad (2)$$

where,

$X1$ = segment 1

$X2$ = segment 2

$tx1j$ = weight of term j in segment 1

$tx2j$ = weight of term j in segment 2

j = 1 number of distinct terms in segment 1 and segment 2

In the normal cosine similarity algorithm, weights are computed with tf-idf (TF x IDF) of each term. In this system weights are filled with (TF x IDF x SR), where SR is computed using Equation (1).

B. Image Feature Extraction Step

In the phase of Visual Feature Extraction, color images of RGB color space and visual features are extracted through color histogram.

Color Space

A color model is an abstract mathematical model describing the way colors can be represented as tuples of numbers, typically as three or four values or color components. A more common approach to comparing the color content of an image to that of another images is that of comparing color histograms.

RGB Color Space

An RGB color space is any additive color space based on the RGB color model. All possible colors can be made from three colorants for red, green and blue. RGB is a convenient color model for computer graphics because the human visual system works in a way that is similar to an RGB color space. In this system, histogram values are computed for each pixel in the image file. Histogram is a set of values (double array). When the histogram was made, it was need to choose a bin size. How large (or small) the bin size should be?

- If we choose too small bin size, a bar height at each bin suffers significant statistical fluctuation due to paucity of samples in each bin.
- If we choose too large bin size, a histogram cannot represent shape of the underlying distribution because the resolution isn't good enough.

Red, Green and Blue values of each pixels are grouped into 4 regions (0-63, 64-127, 128-191, 192-255), and computed according to following equations.

$$binCount = number\ of\ regions \tag{3}$$

Index for each color value is computed as follows:

$$idx = \frac{colorValue \times binCount}{255} \tag{4}$$

After computing the idx for each color value (R, G, B), then histogram indexes are computed as follows:

$$idx = i1 + i2 \times b1 + i3 \times b2 \times b3 \tag{5}$$

where,
$i1$ = red index,
$i2$ = green index,
$i3$ = blue index
$b1$ = bincount for red,
$b2$ = bincount for green,
$b3$ = bincount for blue

Histogram values are increased by 1 according to index value. Then all histogram values are normalized to get value between 0 and 1.

C. Image Retrieval Step

When user enters keyword, images are indexed and searched by using vector space model. The results of keyword search may generate irrelevant images. The irrelevant images are filtered out based on visual feature distribution using *k*-Means algorithm.

K-Means Algorithm

The *K*-Means algorithm is an algorithm to cluster n objects based on attributes into *k* partitions, *k* < *n*. It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data. It assumes that the object attributes form a vector space. The objective it tries to achieve is to minimize total intra-cluster variance, or, the squared error function:

$$V = \sum_{i=1}^{k} \sum_{x_j \in s_j} \left(x_j - \mu_i \right)^2 \tag{6}$$

where there are k clusters Si, i = 1, 2, ..., k, and μ_i is the centroid or mean point of all the points xj in Si. The K-Means clustering was invented in 1956. The most common form of the algorithm uses an iterative refinement heuristic known as Lloyd's algorithm. Lloyd's algorithm starts by partitioning the input points into *k* initial sets, either at random, or using some heuristic data. It then calculates the mean point, or centroid, of each set. It constructs a new partition by associating each point with the closest centroid. Then, the centroids are recalculated for the new clusters, and algorithm repeated by alternate application of these two steps until convergence, which is obtained when the points no longer switch clusters (or alternatively, the centroids are no longer changed) [9].

IV. Experiments

We use the Google image search function to gather the images. Altogether, 3114 images with associated web pages are downloaded and used in our evaluations. In this section, we compare keyword-based search approach with combined search approach. Two standards recall and precision, classically used in information retrieval, are employed to evaluate. Precision is the ratio of the retrieved relevant images and the number of all retrieved images, i.e.,

$$p = \frac{Total\ Number\ of\ Retrieved\ Relevant\ Images}{Total\ Number\ of\ Retrieved\ Images} \tag{7}$$

Recall is the ratio of the retrieved relevant images and total number of relevant images, i.e.,

$$R = \frac{Total\ Number\ of\ Retrieved\ Relevant\ Images}{Total\ Number\ of\ Relevant\ Images} \tag{8}$$

In the following Table I to Table II, we shows the experimental results for some example keywords using keyword-based search approach and combined search approach.

TABLE I: Precision and Recall Result for "Desktop Computer" Keyword

Search Type	No. of Relevant Return	Total No. of Relevant Count	Total Return	Precision (%)	Recall (%)
Keyword Search	60	61	126	47.62	98.36
Combined Search	60	61	106	56.60	98.36

TABLE II: Precision and Recall Result for "Monitor" Keyword

Search Type	No. of Relevant Return	Total No. of Relevant Count	Total Return	Precision (%)	Recall (%)
Keyword Search	201	206	229	87.77	97.57
Combined Search	201	206	226	88.94	97.57

Table III shows the overall results applying the evaluation methodology from above two equations. In combined search, we used k-means algorithm to cluster images based on their color features. In order to determine the appropriate value of *k*, the clustering technique is executed iteratively with *k* ranging from two to a reasonable maximum number. Then, we set *k*=20.

TABLE III: Overall Experimental Results

Search Type	Precision(%)	Recall(%)
Keyword Search	57.257	89.707
Combined Search	79.845	84.046

V. Conclusions

This paper presents image retrieval system using combination of keyword based search and visual features based search. It first partitions web pages into several text blocks based on their semantic cohesions, blocks which contain web images as associated texts for the corresponding web images. Finding semantic relation in text block segmentation improves relevance level of keyword search since texts around image plays an important factor in image search system. Result images of keyword search system may contain irrelevant images and applying visual feature distribution algorithm to cluster images can filter out irrelevant images and thus returning only relevant images. Therefore, precision of this system can be relatively higher than keyword-based approaches.

REFERENCES

[1] Wikipedia, *Image Retrieval*.

[2] T. Gevers and A. W. M. Smeuldres, “The PicToSeek WWW image search system,” in *Proc. IEEE int’l.Conf. Multimedia Computing and Systems,* Florence, Italy, 1999

[3] R. Lempel and A. Soffer, “PicSHOW: Pictorial authority search by hyperlinks on the Web,” *ACM Transactions on Information Systems*, 2002.

[4] V. Harmandas, M. Sanderson, and M. D. Dunlop, “Image retrieval by hypertext links,” in *Proc. SIGIR-97,20th ACM Int’l. Conf. Research and Development in Information Retrieval*, Philadelphia PA, USA, 1997.

[5] S. Sclaroff, L. Taycher, and M. L. Cascia, “Image Rover: A content-based image browser for the World Wide Web,” in *Proc: IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1997.

[6] F. Jing, M. J. Li, H. J. Zhang, and B. Zhang, “Relevance Feedback for Keyword and Visual Feature- Based Image Retrieval,” in *Proc. Third International Conference, CIVR 2004*, vol. 3115, Dublin, Ireland, July 21-23, 2004, pp. 438-447.

[7] Wikipedia, *HTML Dom Node Tree*.

[8] N. D. Thanh and S. Troy, *Measuring Similarity between sentences*.

[9] Saharkiz, “K-Means Clustering Used in Intention Based Scoring Projects,” *Code Project*, 3 Jan., 2009.

Nyein Myint Myint Aung received the bachelor of computer science from University of Computer Studies, Yangon in 2006 and Master from the University of Computer Studies, Kyaing Tong in 2009. She is working as tutor in Computer University, Pyay since 2010. Her current research interest includes information retrieval and image processing.