Web Image Retrieval Refinement by Visual Contents

Zhiguo Gong, Qian Liu, and Jingbai Zhang

Faculty of Science and Technology
University of Macau
P.O.Box 3001 Macao, PRC
{zggong, ma46620, ma46597}@umac.mo

Abstract. For Web image retrieval, two basic methods can be used for representing and indexing Web images. One is based on the associate text around the Web images; and the other utilizes visual features of images, such as color, texture, shape, as the descriptions of Web images. However, those two methods are often applied independently in practice. In fact, both have their limitations to support Web image retrieval. This paper proposes a novel model called 'multiplied refinement', which is more applicable to combination of those two basic methods. Our experiments compare three integration models, including multiplied refinement model, linear refinement model and expansion model, and show that the proposed model yields very good performance.

1   Introduction

With the explosive increase of the Web, Web images, with huge amount and comprehensiveness in meaning, are becoming one of the most indispensable information representation types on the Web. Comparing with Web pages, it is, however, much difficult to find a model to support efficient and effective Web image retrieval. The main reason is due to the facts that (1) The Web images are used freely in the Web pages, and no standard exists for the relationships between the texts and embedded images in the same Web pages; (2) Web images are quite comprehensive in meaning, and they are created by different persons for different purposes; (3) the qualities of the Web images vary greatly. For those reasons, we can not either use the traditional database models (relational model), or visual-content-based model alone for the Web image retrieval. Those challenges make web image retrieval become an attractive research area. For a web image, two sources can be used as its content descriptions: high level or semantic content, such as what the image means, and visual content, such as color, texture, and shape of the objects. Accordingly, two basic approaches are exploited in the web image retrieval: text-based and visual content-based.

Semantic content-Based or text-based web image retrieval(TBIR)[1, 2] utilizes semantic content and is based on the assumption that the associated text can describe the semantics of Web images. Therefore, the associated text is used to index the images. Such systems, such as google and yahoo, try to correlate

associated terms to the embedded images with respect to their importance and relative positions to the image. However, many words, though close to the image, are irrelevant to its semantics. Further more, many noise images exist, such as logos, banners, buttons, which generate many noises to the result. In contrast to the text-based method, the other method is visual content-based web image retrieval(CBIR) [3, 4, 5, 6, 7, 8]. The method assumes that images of the same kind are related to each other in visual features, such as color and texture, and it uses visual content features to support image retrieval. As a matter of the fact, visual content-based image search systems can only provide satisfied performances if all the images are for the same semantic. However, in the context of Web, large percent of noise images will be in the result. That is, visual feature-based method alone can not work well for Web image retrieval.

To improve the performance, in recent years, the hybrid method[9, 10] is used. The basic idea of the hybrid model is to combine text-based ranking with content-based ranking using linear composition model. However, the linear combining model does not take into account of the co-support of those two rankings and great margin of their performance. In this paper we propose a novel model called 'multiplied refinement' for the integration of these two basic models to overcome the limitations of that model. In our solution, the user can start his image search with words or concepts, then he can refine his search with visual features by sample images by multiplied refinement model. Our experiments reveal that the multiplied refinement model is better than both the linear refinement model and expansion model.

The rest of the report will be organized as follows. Sect. 2 introduces some related work. Sect. 3 presents the architecture of web image retrieval and the technologies of its components: text-based image retrieval and visual content image retrieval. Sect. 4 describes different integration models of visual content-based and text-based. Sect. 5 compares their performances and Sect. 6 gives our conclusion.

2   Related Work

There are two basic approaches to support Web image retrieval, including TBIR and CBIR. TBIR creates text-based image indices using the associate texts of the Web images. And in general, a term's relevance to a Web image is based on its locations in the Web document. One of the early researches of TBIR was reported by Chua et al.[1] and Chang and Lee[11]. Sanderson and Dunlop[12] attempted at modeling the semantic of a Web image as a bag of words extracted from both the owner page of the image and the source pages which link to the owner page. Shen et al.[13] modeled surrounding texts, including image's title, image's caption, page's title and image's alt, into corresponding chainnets with different weights. However, it is hard to support fast access to the Web images.

On the other hand, CBIR is also attractive. It utilizes visual contents of images, such as color, texture and shape. Kato[14] was among the first to model visual contents of images. Shortly, John and Chang[5] provided image retrieval

based on color set. More recently, Yanai[15] incorporated visual features such as color and region signatures to model visual contents of images. However, the performance of this method is low and inadequate. To address this problem, Aslandogan and Yu[16] relied on special features to look for human images on the web.

However, Both approaches are pros. and cons.. Therefore, some researches are based on the integration of TBIR and CBIR. Chen et al.[17] and Zhuang et al.[9] utilized expansion model with linear combination of the two approaches. But low precision of CBIR affects the performance.

3   The Technologies of Image Retrieval

Image retrieval includes two basic components, CBIR and TBIR. Its architecture is shown in Fig. 1. There are two basic technologies, including semantic
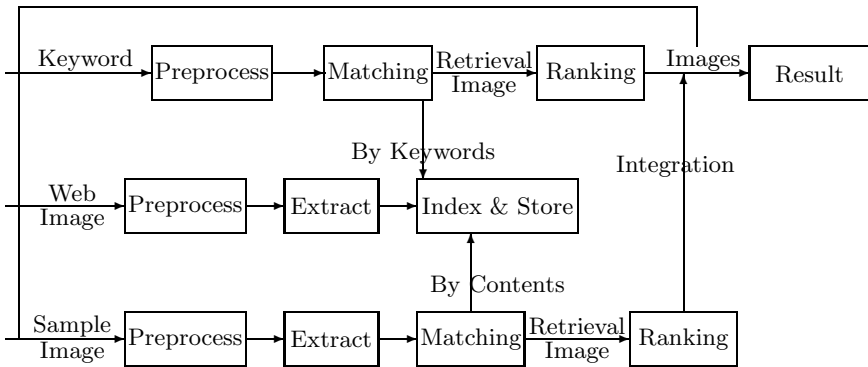

Fig. 1. The Architecture of the System

extractions and visual feature representations of Web images. In this section, we introduce those two technologies in our model.

3.1   Text-Based Image Retrieval

TBIR is to annotate Web images using their associate texts. Then, the techniques for traditional text retrieval can be employed for the Web image retrieval. In order to do so, it is time to determine what parts of the associate texts are used for the extraction and what are the impacts of different parts to the semantics of the images.

Semantic Source. To obtain semantic representation, semantic source (scope of the associated text) must be traced. Based on the relationship between the HTML documents and their containing web images, several parts of the text should be taken into account: image's title, image's alt, image's caption, page's title [2, 13] and other nearest surrounding texts[2]. There are also some other

sources, such as HTML meta data. But these sources may provide false information, which is unrelated to the image and causes some confusion. At length, the five parts, including image's title(STT), image's alt(STA), image's caption(STC), page's title(STP) and nearest surrounding text(STS), fall into better choices.

Semantic Representation. There are several common models to represent the semantics of Web images, including term oriented representation model[2] and ChainNet model[13]. The term oriented model assumes each terms in the associate text are independent with each other and calculates terms' contributions to the images with respect to their locations in the text. That is, different words may have different weights according to the type of its semantic source text. A variable of TFIDF model[2] is used in our system for the calculation of the semantic relevances of terms t_k in any type of text blocks ST_l with respect to image i_j as follows:

$$ntf(t_k)|_{ST_l} = \frac{tf(t_k)|_{ST_l}}{|ST_l|} \qquad (1)$$

In (1), $tf(t_k)|_{ST_l}$ is the frequency of term t in text block ST_l which is any of STT, STA, STC, STP or STS. $|ST_l|$ is the size of block ST_l. Thus, the total semantic relevance of term t_k for image i_j is defined as:

$$ttf(t_k)|_{i_j} = \sum_{k=1}^{L} w_l * ntf(t_k)|_{ST_l} \qquad (2)$$

In (2), L is the total number of the text blocks extracted from the associate text of the image, and w_l is the weight of ST_l, which is defined according to how much that semantic text block contributes to the image i_j in semantics and $\sum_{k=1}^{L} w_l = 1$. The experiment in [2] shows this method provides satisfactory performance.

3.2 Visual Content-Based Image Retrieval

Besides TBIR, visual content-based model is popularly used in traditional image database search. For CBIR, there are several necessary components,including identifying the available visual content features, adopting effective feature representation, automatically extracting the visual features and choosing the discriminating function for the visual features. Now, visual contents include color, texture, objects' shape and spatial frequency, which can be the information of the whole images or the region after partitioning the whole image into several regions. Because segmentation of image is still an open research, the whole image's information is used in this paper.

Color Extraction. Color is the basic and most straight-forward characteristic of the image and most extensively used in CBIR. There are several important issues for color extraction, including appropriate color space and effective color representation. Generally, there are many different color spaces, such as RGB,

CMY, HSL. Among those spaces, HSL represents color by three variables: hue, lightness and saturation, and is more similar to human vision system principles. More importantly, HSL is its tractability, perceptually uniform, and possible and easy transformation from popular RGB space to HSL space.

There are several choices for color representation: color histogram, color coherence vector[22], color correlogram, color moments and color set[8]. Global color histogram is effective and easy to compute and robust to translation, rotation and scale. In global color histogram, each bin represents the number of pixels which has the same color. For an image, there may be plenty of bins which makes for a tremendous increase in the cost of computing the similarity of two images and also leads to inefficient index. Thus, it is necessary to preprocess images and quantization is an effective way.

Texture Extraction. Texture is an innate property of all surfaces and refers to visual patterns of homogeneity. It is discriminable and important structure of the image. There are three basic approaches to extract texture: spectral approaches, structural approaches and statistical approaches. In recent years, wavelet, as one of structural approaches, is popular used in image processing. Based on wavelet transform, the useful information is the statistics of coefficients in each frequency in wavelet transform processing, and mean and variance of the energy distribution of the coefficients for each frequency at each decomposition are used to construct the vector. This representation of texture vector is:

$$\overrightarrow{fvt} = \{\frac{\mu_{11}}{\delta_{\mu_{11}}}, \frac{\sigma_{11}}{\delta_{\sigma_{11}}}, \frac{\mu_{12}}{\delta_{\mu_{12}}}, \frac{\sigma_{12}}{\delta_{\sigma_{12}}}, \cdots, \frac{\mu_{ij}}{\delta_{\mu_{ij}}}, \frac{\sigma_{ij}}{\delta_{\sigma_{ij}}}, \cdots, \frac{\mu_{NM}}{\delta_{\sigma_{NM}}}, \frac{\sigma_{NM}}{\delta_{\sigma_{NM}}}\} \qquad (3)$$

In (3), N is the level of the transform and M is the number of frequencies of each level, and that number is four denoting one approximation frequency and three detail frequencies. μ_{ij} and σ_{ij} is respectively the mean and variance of the frequency j in level i. $\delta_{\sigma_{ij}}$ and $\delta_{\mu_{ij}}$ are standard deviations of σ_{ij} and μ_{ij} respectively in the entire database.

Dissimilarity Functions. Jan Puzicha et al.[18] summarize dissimilarity functions and propose four categories, including heuristic histogram distances, non-parametric test distances, information-theoretic distances and ground distances. Among the functions of those kinds, Euclidean distance is the effective and common dissimilarity function. In fact, Euclidean distance is effective and easy to calculate. The calculation of Euclidean distance, for vector v_1 and v_2, whose form is $\{b_1, b_2 \ldots, b_L\}$, is:

$$d = \sum_{l=1}^{L} (v_1.b_l - v_2.b_l)^2 \qquad (4)$$

In (4), L is the dimensions of vector. In our methods, those two vectors can be color histogram vector and texture vector.

4   Integration Models of Visual Content-Based Image Retrieval and Text-Based Image Retrieval

In our prototype system, an image query can be $q=(q_t, q_i)$, where q_t is the query description for TBIR and q_i is the sample images for CBIR. In the case of q_i=NULL(users do not provide), the query is TBIR. if q_t is NULL, it is CBIR. If users provide both q_t and q_i at the same time, that is the combined image retrieval, which has a common situation where after the retrieval with $q=(q_t,$ NULL$)$, q_i is selected from the result for CBIR. There, the structure of retrieval results for TBIR and CBIR are defined as follows:

$$RS_{\text{TBIR}} = \{(i_1, R_1), (i_2, R_2) \ldots (i_j, R_j) \ldots (i_N, R_N)\}$$
$$RS_{\text{CBIR}} = \{(i_1, S_1), (i_2, S_2) \ldots (i_j, S_j) \ldots (i_N, S_N)\}$$

(5)

In 5, supposed the image collection $\{i_1, i_2 \ldots i_N\}$ has totally N images. RS_{TBIR} is the resultant set of TBIR, RS_{CBIR} is the resultant set of CBIR. Item (i_j, R_j) means the relevance between image i_j and some keyword is R_j, and item (i_j, S_j) means the similarity between image i_j and some sample image is S_j. Before integration, RS_{TBIR} can prune the images with R_j less than some threshold, generally, that value is 0 and we call the rest prune-RS_{TBIR}, and RS_{CBIR} also prune the image with S_j less than some threshold for the low performance of CBIR and we call the rest prune-RS_{CBIR}. Therefore, the integrated set may be composed of three parts, including common items which is in both prune-RS_{TBIR} and prune-RS_{CBIR}, TBIR-only items which is in prune-RS_{TBIR} but not in prune-RS_{CBIR}, and CBIR-only items which is in prune-RS_{CBIR} but not in prune-RS_{TBIR}. In our prototype system, the retrieval result set is $\{i_1, i_2 \ldots i_M\}$ with M images. If that set includes those three parts, that is, the integrated set is $(RS_{\text{TBIR}}, RS_{\text{CBIR}})$, and then ranks items again based on integrated relevance, that model is *expansion model*. If only common items and TBIR-only items is in that set, and prunes CBIR-only items, that is, the integrated set is still (RS_{TBIR}) but the sequence of their items may be changed based on integrated relevance, that model is *refinement model*. Each model can use different integrated method to obtain the integrated relevance, such as *linear method and multiplied method*. It is linear combined method, if the integrated relevance is produced by the formula $RS_{\text{new}} = \alpha * R_j + \beta * S_j$, where RS_{new} is new integrated relevance, and α and β are coefficients and $\alpha+\beta=1$. Without more words, in that linear formula, S_j of TBIR-only items and R_j of CBIR-only items are 0. And if the formula, $RS_{\text{new}} = R_j * (1 + S_j)^\gamma$, where RS_{new} and γ is coefficient, is used to compute the integrated relevance, it is multiplied combined method. In that multiplied formula, the similarity S_j is added by 1 to avoid the relevance based on TBIR multiplied by zero for TBIR-only items. Therefore, there are three available models: linear expansion model, linear refinement model and multiplied refinement model.

Linear expansion model is reported by Chen et al.[17] and Zhuang et al.[9]. However, expansion model makes the same disposal for common items and others. As we know, the precision of those two parts varies greatly and more greatly

in CBIR. However, that model treats TBIR and CBIR equally, that is, items with same value of their contribution in TBIR and CBIR is handled with same importance. In fact, CBIR often produces many irrelevant results because of the comprehensive semantics of Web images. In other words, even though an image is much similar to the sample image in visual features, it may have a far distance in semantics. To overcome the limitation, Guojun Lu and Ben Williams[10] provided the linear refinement model to integrate CBIR into TBIR. And different from the expansion model, CBIR-only items are pruned. Refinement model makes different disposal of common items and CBIR-only items. However, linear expansion model and linear refinement model make integration based on linear method, which is not sensitive to the co-support of retrieval sets of CBIR and TBIR. More importantly, linear method disregards the speciality for two image retrieval sets: the correlation between the terms and images which is more apparent in refinement.

In detail, in the refinement, image retrieval starts with the query $q=(q_t, \text{NULL})$, then users can refine their retrieval with visual features by sample Web images with the query $q=(\text{NULL}, q_i)$. From this process, we know that the relevance of the keywords to some resultant image obtained by keywords is original, and more importantly, the refinement by CBIR means that under the situation some sample image is completely semantic to the keyword and other images possess different similarity between them and the sample image, what is the new relevance of the keywords to some resultant image. Without question, the correlation of keywords and images is implied and new relevance can be obtained by multiplying original relevance by the similarity of the sample image and other resultant images. That is the novel multiplied refinement model and that model overcomes the limitation of expansion model and linear method. To evaluate the performance of different models, the prototype system is implemented, which is similar to Fig. 1.

In the prototype system, TBIR extracts the associated text in STT, STA, STC, STP and STS, and calculates the relevance based on the variation of TFIDF model. In CBIR, color and texture are utilized. In detail, color histogram and statistic of the wavelet transform are used as the feature vectors. For color histogram, HSL color space is better choice and quantization method is used to get color histogram, that is, hue is divided into eighteen levels, saturation and lightness are divided into three levels respectively, and grey color is scaled into four levels. Therefore, the color vector is 166-dimension(18*3*3+4). For texture, Daubechies wavelet transform is used. As we know, four frequencies are obtained after one time wavelet transform: one is composed of approximation coefficients(LL) and three are composed of detail coefficients, including horizontal coefficients(LH), vertical coefficients(HL) and diagonal coefficients(HH). And wavelet transform can be continued further based on the data of those four frequencies. There are two popular methods for that continuing wavelet transform: pyramid-structured wavelet transform(PWT), which only decomposes LL, and tree-structured wavelet transform(TWT), in which all frequencies will be decomposed. For PWT, some information is lost and for TWT, the decomposition of

HH is unstable. Therefore, the composite method is to decompose the frequencies except HH in each level of the transform again. The mean and variance of each frequencies in each level are utilized as components of the feature vector. The prototype system makes use of wavelet transform four times and produces 320-dimension(2*4*(1+3+9+27)) texture vector.

5   Performance Evaluation

In our prototype system, more than 12000 web images from 50000 web pages are gathered after noise images, such as icons, banners, logos and any image with size less than 5k, removed. In the experiments, 20 terms with their 60 relevant images are used to obtain the optimal parameters and 10 terms with their 20 relevant images are used for testing. In the prototype system, the performance is evaluated by average precision (AP) objective in (6),

$$AP_{jk} = \frac{1}{R_{jk}} \sum_{k=1}^{R_{jk}} \frac{k}{N_k} \tag{6}$$

where R_{jk} is the number of relevant images in the result and N_k is the number of the results when there are up to k revelant results.

The first step is to determine the parameters for each integration. There are five parameters, two for linear expansion model, two for linear refinement model and one for our novel model. Based on the prototype system, Tables 1 can be obtained. Tables 1 are for linear expansion model, linear refinement model and multiplied refinement model. For the left table, the smaller the parameter, β, for the similarity of visual contents is, the better the integration's precision is, which is due to its low precision and single-independence of linear method, that

Table 1. Coefficients of Different Models

Linear Expansion Model

β	α	AP
0	1	0.428571
	α:0.98	0.429175
β:0-0.1	α:0.96	0.429375
α:0.9-1	*α:0.94*	*0.430828*
	α:0.92	0.429134
0.1	0.9	0.428252
0.2	0.8	0.4003912
0.3	0.7	0.391616
0.4	0.6	0.380326
0.5	0.5	0.374529
0.6	0.4	0.323
0.7	0.3	0.290917
0.8	0.2	0.263213
0.9	0.1	0.197241

Linear Refinement Model

β	α	AP
0	1	0.428571
0.1	0.9	0.429186
0.2	0.8	0.429308
0.3	0.7	0.430311
0.4	0.6	0.432207
0.5	0.5	0.43651
β:0.5-0.6	α:0.44	0.437063
α:0.4-0.5	α:0.42	0.438132
0.6	*0.4*	*0.438*
β:0.6-0.7	α:0.38	0.438625
α:0.3-0.4	α:0.36	0.438209
0.7	0.3	0.434484
0.8	0.2	0.430433
0.9	0.1	0.435272

Multiplied Refinement Model

γ		AP
0		0.428571
1		0.436518
2		0.442337
3		0.443291
3	3.4	0.44434
to	3.6	0.444258
4	3.8	0.445431
4		0.447749
4	4.2	*0.448586*
to	4.4	0.447794
5	4.6	0.447237
5		0.446812
6		0.446704
7		0.445797

is, the images with larger value in each collection may be contribute more largely to resultant ranking than common items with smaller value in both collections. In our prototype system, better β is 0.06 for its maximum AP 0.430820. For the middle table, β for the similarity of visual contents is actually for common items for refinement model prunes CBIR-only items before. Therefore, better value of β is larger than that of linear expansion model. In our prototype system, better β is 0.6 for its maximum AP 0.438. From the comparison of linear expansion model and linear refinement model, the precision of common items and of CBIR-only items in CBIR vary largely through better value of β. For the right table, 4.2 for γ is optimal with maximum AP 0.448586.

The next experiment is to compare the performance of different models and the result in Fig. 2 can be obtained. Figure 2 shows that all models improve
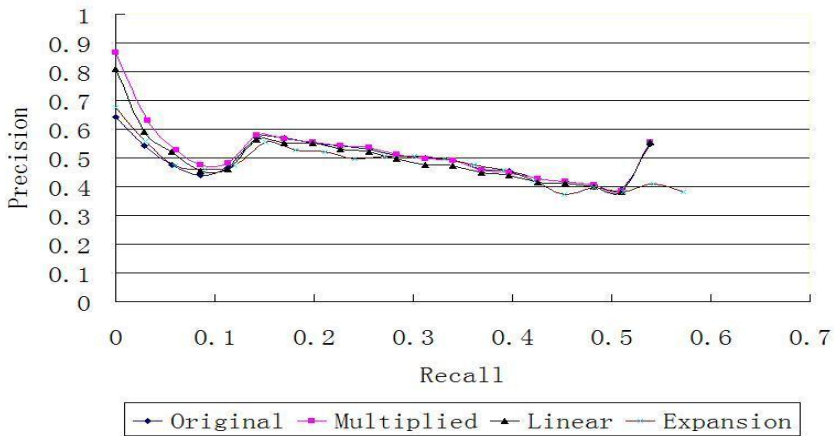

Fig. 2. Comparison between Different Models

the performance. In Fig. 2, original retrieval's AP is 0.428571, its maximum recall is 0,53856 and the AP before the recall with 0.113381 is 0.444375. Linear expansion model, at the beginning of the recall, gives a little improvement. Out of question, this model can obtain higher recall, 0.572158. In our prototype system, its AP is 0.430828, its maximum recall is 0.572158 and the AP before the recall with 0.113381 is 0.457954. The expansion model is more effective when TBIR has less recall. For each refinement, the maximum has no change. Therefore, it is more effective when TBIR has higher recall but less AP. In our prototype system, linear refinement model provides better improvement. Its AP is 0.439073 and the AP before the recall with 0.113381 is 0.525908. More importantly, multiplied refinement model provides best improvement. Its AP is 0.450086 and the AP before the recall with 0.113381 is 0.551451. From Fig. 2, the notable improvement cannot be shown. But the remarkable improvement is the beginning of the retrieval. As we know, users most focus on the first K items of each retrieval. Therefore, those items are most important and the

improvement of those items is more interesting and useful. From our experiment, the APs of original retrieval, Linear expansion model, linear refinement model and multiplied refinement model before the recall with 0.113381 are 0.444375, 0.457954, 0.525908 and 0.551451, respectively. From those values, we know all model make satisfied improvement and the improvement of multiplied refinement model is the best.

To make clear comparison of different integration model , More statistics is provided in Table 2, where 'O', 'M', 'L' and 'E' represent original retrieval, the retrieval of multiplied refinement model, the retrieval of linear refinement model and the retrieval of linear expansion model, respectively. Table 2 shows

Table 2. Examples for Comparison between Different Models

Keyword	Top 16 Images				Top 48 Images				Top 96 Images			
	O	M	L	E	O	M	L	E	O	M	L	E
dv	8	12	9	9	37	36	36	36	46	49	48	51
notebook	8	11	10	10	19	25	24	19	26	34	21	24
game	5	9	7	7	30	40	37	36	56	55	53	41

that each integration model can produce better performance before the top 48 images. At the 100 images, the variation is small, even if some original result may be better because of the noise of the integration. Among those integration, our novel multiplied refinement model show best performance because it overcomes the limitation of expansion model and linear method.

Fig. 3. Original Result of "Notebook"

Fig. 4. Original Result of "DV"

Fig. 5. Original Result of "Game"

Fig. 6. Refined Result of "Notebook"

Fig. 7. Refined Result of "DV"

Fig. 8. Refined Result of "Game"

For the detail of the performance of multiplied refinement model, we give some examples where there are top 16 images of search results to show. The result of text-based retrieval of *notebook, dv and game* is respectively in Fig. 3, Fig. 4 and Fig. 5. In our novel refinement model, Fig. 6, Fig. 7 and Fig. 8 are obtained after refinement. Without more words, the refinement produces better performance. Take the images of "DV", there are 12 related images compared to 8 images before refinement.

6   Conclusion

Much attention [9, 10] has been devoted to Web image retrieval. And two basic approaches for image retrieval are TBIR and CBIR, which utilize the associated text and visual features, respectively. Each can be utilized independently but has its limitations. Therefore, different integration models are tried. Expansion model and refinement model are two common models, and multiplied method and linear method are two common methods. Therefore, the available models include linear expansion model, linear refinement model and multiplied refinement model. However, expansion model doesn't consider low precision of CBIR and linear method disregards the correlation of TBIR and CBIR. Therefore, this report has proposed a novel model—multiplied refinement model to integrate CBIR into TBIR to overcome those limitations and the performance is better than others.

References

1. Chua T. S et al: A Concept-based Image Retrieval System. Proceedings of 27th Annual Hawaii International Conference on System Science, Maui, Hawaii, January 4-7 1994. (1994) 590-598
2. Zhiguo Gong, Leong Hou U, Chan Wa Cheang: An Implementation of Web Image Search Engines. ICADL. (2004) 355-367
3. Jonathan Ashley et al: The Query By Image Content (QBIC) System. SIGMOD Conference. (1995) 475
4. Ediz Saykol, Ugur Güdükbay,Özgür Ulusoy: Integrated Querying of Images by Color, Shape, and Texture Content of Salient Objects. ADVIS. (2004) 363-371
5. John R. Smith, Shih-Fu Chang: Single Color Extraction and Image Query. ICIP-95. (1995)
6. John R. Smith, Shih-Fu Chang: Automated Image Retrieval Using Color and Texture. Pattern Analysis and Machine Intelligence (PAMI). (1996)
7. John R. Smith, Shih-Fu Chang: Tools and Techniques for Color Image Retrieval. Storage and Retrieval for Image and Video Databases (SPIE). (1996) 426-437
8. John R. Smith, Shih-Fu Chang: TVisualSEEk: A Fully Automated Content-Based Image Query System. ACM Multimedia. (1996) 87-98
9. Yueting Zhuang, Qing Li,Rynson W. H. Lau: Web-Based Image Retrieval: A Hybrid Approach. Computer Graphics International. (2001) 62-72
10. Guojun Lu, Ben Williams: An Integrated WWW Image Retrieval System. http://ausweb.scu.edu.au/aw99/papers/lu/paper.html. (1999)

11. C. C. Chang, S. Y. Lee: Retrieval of similar pictures on pictorial databases. Pattern Recogn. **24** (1991) 675–681
12. V. Harmandas, Mark Sanderson, Mark D. Dunlop: Image Retrieval by Hypertext Links. SIGIR. (1997) 296-303
13. Heng Tao Shen, Beng Chin Ooi, Kian-Lee Tan: Giving meanings to WWW images. MULTIMEDIA '00: Proceedings of the eighth ACM international conference on Multimedia. (2000) 39–47
14. Kato T: Database Architecture for Content-Based Image Retrieval. Proceedings of Society of the Photo-Optical Instrumentation Engineers: Image Storage and Retrieval, 1662. 1992. San Jose, California, USA, SPIE. (1992)
15. Keiji Yanai: Generic image classification using visual knowledge on the web. ACM Multimedia. (2003) 167-176
16. Y. Alp Aslandogan, Clement T. Yu: Multiple evidence combination in image retrieval: diogenes searches for people on the Web. SIGIR. (2000) 88-95
17. Zheng Chen et al: Web mining for Web image retrieval. JASIST. **52** (2001) 831-839
18. Jan Puzicha et al: Empirical Evaluation of Dissimilarity Measures for Color and Texture. ICCV. (1999) 1165-1172
19. Petra Nass: The Wavelet Transform. http://www.eso.org/projects/esomidas/doc/user/98NOV/volb/node308.html. (1999)
20. M. K. Mandal, T. Aboulnasr: Fast wavelet histogram techniques for image indexing. Comput. Vis. Image Underst. **75** (1999) 1077-3142
21. Wikipedia: HSL color space. http://en.wikipedia.org/wiki/HLS_color_space.
22. Greg Pass, Ramin Zabih, Justin Miller: Comparing Images Using Color. ACM Multimedia. (1996) 65-73