Study on Hybrid Image Search Technology Based on Texts and Contents

H.T. Wang1, F.L. Ma2, C Yan3 and H Pan3
1Network Information Center, Wuhan University of Technology, Wuhan 430070, PR China
2School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, PR China.

Abstract. Image search was studied first here based on texts and contents, respectively. The text-based image feature extraction was put forward by integrating the statistical and topic features in view of the limitation of extraction of keywords only by means of statistical features of words. On the other hand, a search-by-image method was put forward based on multi-feature fusion in view of the imprecision of the content-based image search by means of a single feature. The layered-searching method depended on primarily the text-based image search method and additionally the content-based image search was then put forward in view of differences between the text-based and content-based methods and their difficult direct fusion. The feasibility and effectiveness of the hybrid search algorithm were experimentally verified.

1. Introduction

Text-based image search has been being further studied by many scholars so far. References indicate that many scholars focused on types of image correlation information in the stage for annotation of most of the current text-based image search; on the other hand, less study was carried out to the weight related to the context information[1,2]. The content-based image search existed in some ready-made systems abroad; for example, Web SEEK and Visual SEEK have image search tools with color features (such as texture features based on the wavelet transform and visual features based on the color set) for applications[3,5]. On the other hand, the domestic research on the content-based image search started relatively late, which always focused on some single features not enough to express information of entire images; moreover, those search algorithms based on features of single contents often present unsatisfactory effects. Thus, multi-feature fusion technology of the content-based image search is still to be further studied. The fusion issues focuses on mutual compensation of different factors and integration of their advantages to get a better effect. The traditional fusion methods include the linear fusion, fuzzy analytical hierarchy process[6-9], factor analysis and so on. The fusion issue in the image field is generally to bridge the "gap" between image text and visual information and discuss how to perform their fusion for better description of images[10,11].

2. Text-based image feature extraction

2.1 Processing information related to images. First, the text information in the page including an image is preprocessed. The processing flow diagram is shown in Figure 1.
NLPIR was selected here as the word segmentation tool and Sogou scel was introduced to deal with new words. After completion of word segmentation of a document, stop words may be removed by means of screening of part-of-speech (POS) in accordance with NLPIR Chinese (POS) annotation set.

2.2 Weight-based TF-ID selection of candidate phrases. AHP was selected here as the method for analysis of the weights of information related to images[12].

2.2.1 Establishment of the hierarchy structure model. Text information was selected in 3 positions, namely: webpage title (B1), image title (B2) and image context information (B3). The specific hierarchy structure is schematically shown in Figure 2.

2.2.2 Establishment of the judgment matrix (M). The method for establishment of the judgment matrix is described as follows: its overall objective is analysis of weights of information related images and the scoring mode is adopted for judgment of the relative importance of any two factors among 3 factors.

2.2.3 Calculation of eigenvectors of the judgment matrix by means of Asymptotic Normalization Coefficient.

2.2.4 Consistency checking. During the experiment, the indexes for checking results are generally the C.I. (consistency index) and C.R. (consistency ratio). The specific steps are as follows:
Step 1: calculation of C.I. is performed by the following equation:
\[ C.I. = \frac{\lambda_{\text{max}} - n}{n - 1} \]  
(1)
Where: \( \lambda_{\text{max}} \) represents the maximum eigenwert of the matrix.
Step 2: checking the mean random C.I. in the lookup table.
Step 3: calculation of C.R. is performed by the following equation:
\[ C.R. = \frac{C.I.}{R.I.} \]  
(2)
It is gained from Steps 1 and 3 that: C.I. and C.R. are 0; obviously, both of them are less than 0.1; thus, this experiment passed the consistency checking; and the estimated weights of 3 kinds of text information related to images are presented in Table 1.

| Factor | Type             | English   | Weight |
|--------|------------------|-----------|--------|
| B1     | Webpage title    | Page Title| 0.35   |
| B2     | Image title      | Title     | 0.5    |
| B3     | Webpage text     | Text      | 0.15   |

Weights of context information (Table 1) and the TF- algorithm were here integrated for operation; and the specific equations are as follows:
Weighted term frequency (TF):
\[
(\hat{w}_f) = \sum_{i=1}^{m} \frac{n_i w_i}{n}
\]
Inverse document frequency (IDF):
\[
(\hat{d}_f) = \log \left( \frac{N}{N_i+1} \right)
\]

Where: \(m\) represents the total number of types of context information related images (\(m=3\) here); \(n_i\) represents the occurrence frequency of the key words in the \(i^{th}\) type of context information; \(w_i\) represents the weight of the \(i^{th}\) type of context information; \(n\) represents the total number of candidate phrases in the context information related images; \(N\) represents the size of the corpus; \(N_i\) represents the number of texts including this term; and \(wtf-idf\) represents the weight-based TF-IDF value of keywords.

### 2.3 Extraction of text keywords by integrating topical features

#### 2.3.1 Selection of subject elements.
**POS:** Nouns and verbs shall be selected to lower the calculation complexity.

**TF:** those terms closely related to the theme generally occur more frequently in a paper. A TF calculation method was utilized here for normalization of TF. The specific equation is as follows:

\[
df(w_i) = f_i(w_i) / \sum_{j=1}^{n} f_j(w_j)
\]

Where: \(tf(w_i)\) represents the relative TF of Word \(w_i\) in Document \(d_j\); \(f_i(w_i)\) represents the occurrence frequency of Word \(w_i\) in Document \(d_j\); and \(n(d_j)\) represents the number of notional words in the document.

**Document frequency (DF):** DF refers to the number of documents including a certain word; and DF is calculated by:

\[
df(w_i) = \frac{n(w_i)}{N} \times \log \left( \frac{n(w_i)}{N} \right) - (1 - \frac{n(w_i)}{N}) \log \frac{n(w_i)}{N}
\]

Where: \(df(w_i)\) represents the DF theme expression of Word \(w_i\); \(n(w_i)\) represents the number of documents including Word \(w_i\); and \(N\) represents the total number of documents.

**Position of a word:** the weight of the position of a word is calculated by:

- As for Word \(w_i\) in a title: \(f_p(w_i) = 1 + 4 \times \frac{f(t_i)}{s(t)}\)
- As for Word \(w_i\) in texts: \(f_p(w_i) = 1\)

Where: \(f_p(w_i)\) represents the weight of the position of Word \(w_i\); \(f(t_i)\) represents the occurrence frequency of Word \(w_i\) in a title; and \(s(t)\) represents the number of notional words in the title of Document \(d_t\).

#### 2.3.2 Co-occurrence analysis of phrases.
The co-occurrence phenomenon of phrases refers to the statistical relationship between phrases or words. The following keywords were extracted here. The specific weight is calculated by:

\[
g(w_i) = n(w_i) / (n(w_i) + 1)
\]

Where: \(n(w_i)\) represents the number of fixed collocations of Word \(w_i\) in the document space; and \(g(w_i)\) represents the theme expression of Word \(w_i\).

#### 2.3.3 Calculation of theme semantic features.
After word segmentation and removal of stop words, some significant nominal words remains; moreover, evaluation of expression weights of words primarily depends on their TFs and positions; and co-occurrence information centralized in texts may have a reference value. In general, the theme expression of information (such as TF) is dominant to reflect the correlation between words or phrases and the theme to a great extent. The co-occurrence
The phenomenon of phrases reflects whether those words centralized in texts are keywords. The scoring equation is as follows:

\[ W(w_i) = tf(w_i) \times f_j(w_i) \times df(w_i) \times (1 + g(w_i)) \]  

(11)

Where: \( tf(w_i) \) represents the relative TF of Word \( w_i \) in Document \( d_j \); \( df(w_i) \) represents the DF weight of Word \( w_i \); \( f_j(w_i) \) represents the weight of Word \( w_i \); and \( g(w_i) \) represents the theme expression of Word \( w_i \).

The total of the weight of a candidate centralized word with statistical features and its theme weights is regarded as the final scores of the candidate word:

\[ p(d, w_j) = wtf-idf + W(w_j) \]  

(12)

Where: \( wtf-idf \) represents the weight-based TF-IDF value of the keyword; and \( w(w_j) \) represents the theme semantic feature value of the word.

3. Content-based Image Search

3.1 Content-based image feature selection. The following factors were taken into account here for selection of the image feature set:

- How to select the feature from various features for expression of the image contents as can as possible;
- How to efficiently extract the image features.

As for experimentally extraction of image features, an image set including 110 images (resolution: 1024*740; and bit depth: 24 bits) acted as the test objects and each image feature was extracted for 15 times; and then the measured data were averaged. Figure 3 show comparison of extraction rates for various features, where the vertical axis represents various features and the horizontal axis represents time whose unit in second for extraction of a feature.

![Figure 3. Comparison of extraction rates for image features](image)

Figure 3 indicates that extraction of features of the Color Correlogram takes a long time and extraction of features of other items spend equivalent time. Considering the generation efficiency of the eigenvalues of images, fusion of features may occur for those items except the Color Correlogram.

3.2 Content-based multi-feature fusion scheme. Our fusion scheme is calculated by:

\[ w = \sum k_i \times s_i \]  

(13)

Where: \( w \) represents the weight of images; \( k_i \) represents the weight of the \( i^{th} \) image feature solved by means of the genetic algorithm; \( s_i \) represents the degree of similarity based on the \( i^{th} \) image feature.

To determine the individuals and genes: the weights vector of our genes (image features) is \( (k0, k1, k2, k3, k4, k5, k6, k7, k8) \); and individuals include genes and their presenting fitness.

To determine the criterion for judgment of the fitness of individuals: the fitness is calculated by:
\[ \text{fitness} = \sum \left( \frac{1}{2} \right)^{\text{index}} + 1 \]  

(14)

Where: index is the subscript of a similar image in the search results (the subscript of the similar image in the first place is defined as 0).

To determine the elimination criteria: while our applying the genetic algorithm, only some individuals with the maximum fitness remain to improve the calculation efficiency and appropriately lower the computational complexity.

To determine the gene recombination rules: the specific calculation equation is as follows:

\[ k_{\text{child}} = \left( k_{\text{father}} + k_{\text{mother}} \right) / 2 \]  

(15)

To determine the gene mutation criteria: the weights after gene recombination are as follows:

\[ k_1 = k_1 - \text{random} \]  

(16)

\[ k_2 = k_2 + \text{random} \]  

(17)

In summary, the method for solving the weights of image features based on the genetic algorithm is schematically in Figure 4.

![Flow diagram of the scheme for solving weights of image features](image)

Figure 4. Flow diagram of the scheme for solving weights of image features

The actual implementation is a continuous cyclic process. While the iterative process has been completed, the individual whose fitness peaks is selected from those survival individuals, whose genes are selected as the weights vector of image features; and such vector is integrated into the equation for solving the fusion similarity.

4. Hybrid image search scheme based on texts and contents and application

Considering the algorithms for processing texts and images and their time and space complexities, the text-based search is dominant and the content-based image search is subsidiary for sufficient fusion of their advantages and further improvement of the search efficiency; moreover, duplicate and precision checking rates of the image search engine shall be guaranteed.

4.1 Experiments for extraction of text features of images and results analysis
4.1.1 Experimental setup. As for our experiments for extraction of text features of images, the experimental data are from the Palace Museum Network (http://www.dpm.org.cn/). The specific experimental steps are as follows:

Step 1: Crawling and analysis of 1,000 ceramics images and related text information: they are divided into 5 categories (200 images per category), which are “Bottles”, “Bowls”, “Pots”, “Flagons” and “Dishes”; and the related information is preprocessed.

Step 2: the weight of each keyword is calculated by means of the weight-based TF-IDF algorithm, ten of which are extracted as the text feature candidate set of images.

Step 3: Calculation of topical eigenvalues of words.

Step 4: To determine the regulatory factor (η) for mixing the statistical and topical features. η is among 0-1 for our experiment and its step size is 0.1. 100 documents are selected at random. While 5 keywords are extracted per document in accordance with the algorithm, the average annotation precision (avg_P), recall (avg_R) and comprehensive evaluation index (F) changes along with the growth of η (Figure 5: x axis – η and y axis - avg_P, avg_R and F). Figure 5 indicates that fusion has the best effects while η=0.6.

Figure 5. avg_P, avg_R and F vs. η

Step 5: calculation of overall scores of words based statistical and topical features in accordance with Eq. (16); and these scores are sorted from small to big. The top 5 scores are extracted as text features of images.

4.1.2 Experimental evaluation indexes. avg_P, avg_R and F are utilized for measurement of the precision of extraction of text features of images, which are calculated by:

\[ \text{avg}_P = \frac{1}{N} \sum_{i=1}^{N} P(I_i) \]  \hspace{1cm} (18)

\[ \text{avg}_R = \frac{1}{N} \sum_{i=1}^{N} R(I_i) \]  \hspace{1cm} (19)

\[ F = \frac{2 \times \text{avg}_P \times \text{avg}_R}{\text{avg}_P + \text{avg}_R} \]  \hspace{1cm} (20)

Where: N represents the size of the measurement image set; and \( P(I_i) \) and \( R(I_i) \) represents the precision and recall of annotation of the Measurement Image \( I_i \), respectively, whose calculation equations are as follows:

\[ P(I_i) = \frac{|A_i \cap G_i|}{|A_i|} \]  \hspace{1cm} (21)

\[ R(I_i) = \frac{|A_i \cap G_i|}{|G_i|} \]  \hspace{1cm} (22)
Where: $A_i$ represents the annotation gained based on the above algorithm; $G_i$ represents the standard keyword of the image; $|A_i \cap G_i|$ represents the number of correct annotations; $P(I_i)$ represents the annotation precision of a certain image; and $R(I_i)$ represents the integrity for annotation of a certain image.

4.1.3 Check experiment settings and summary. For verification of feasibility and effectiveness of the algorithm for capturing text features of images by integrating topical and statistical features of words, the following check experiments are set currently; and the corresponding comparison results are as follows.

| Algorithm                      | avg_P  | avg_R  | F     |
|-------------------------------|--------|--------|-------|
| TF-IDF                        | 0.325  | 0.367  | 0.345 |
| Weight-based TF-IDF            | 0.376  | 0.405  | 0.39  |
| Integration of topical features| 0.402  | 0.420  | 0.411 |

Given all that, not only extraction of text-based image features but also utilization of statistical and topical features of words may improve the precision of extraction.

4.2 Content-based image search check experiment and results analysis

4.2.1 Experimental evaluation indexes. Precision-Recall criteria are utilized for evaluation of performances of the algorithm. Precision and Recall are calculated by:

\[
\text{Precision} = B / A \tag{23}
\]

\[
\text{Recall} = B / C \tag{24}
\]

Where: $A$ represents the number of returned images after searching; $B$ represents the number of images related to searching images in the search results; and $C$ represents the total number of images related to the search images in the search database.

4.2.2 Experimental design and summary. The experiments were carried out in Corel 1000 of the public data set. This image database is made up of 10 types of subsets; and 90 images were selected at random per subset to form the image set (including 900 images) and solve the optimized weight for various features by means of the genetic algorithm; moreover, the remaining 100 images as the measurement image set were used for searching and calculation of the mean precision of search results of the two experiments as follows.

Experiment 1: the multi-feature fusion scheme was used based on the weight of the same feature; namely, it is thought that the contribution of each type of features is the same for the image contents. For example, while searching Image 611.jpg, the search results are represented in Figure 6.

Experiment 2: the above genetic algorithm was utilized for calculation of the weight of each feature; moreover, the multi-feature fusion scheme was utilized based on various weights. For example, while searching Image 611.jpg and using this algorithm, the search results are represented in Figure 7.
The evaluation indexes are listed in Table 3.

| Type       | Experiment 1 | Experiment 2 |
|------------|--------------|--------------|
| Africa     | 0.35         | 0.47         |
| Beach      | 0.26         | 0.31         |
| Buildings  | 0.25         | 0.38         |
| Bus        | 0.51         | 0.65         |
| Dinosaurs  | 0.36         | 0.51         |
| Elephants  | 0.34         | 0.48         |
| Flowers    | 0.37         | 0.42         |
| Horses     | 0.51         | 0.58         |
| Mountains  | 0.16         | 0.22         |
| Food       | 0.32         | 0.42         |

The above search results indicate that the degree of similarity of search results is higher and the images are positioned centrally in the front while the genetic algorithm is applied. Table 3 indicates that the algorithm utilized here leads remarkable improvement of Precision and Recall. Experiments were carried out to the public data set to gain that the mean precisions for the both algorithms are 0.343 and 0.444, respectively. Thus, the multi-feature fusion algorithm shall be feasible and effective based on calculation of weights of various features by means of the genetic algorithm.

4.3 Hybrid image search scheme. The hybrid search scheme schematically shown in Figure 8 will be described in detail.
1 – Crawler; 2 - html parser; 3 – Database; 4- Extraction of content features of images; 5 - Extraction of topical features of words; 6 - Extraction of the candidate set of keywords by means of the weight-based TF/IDF; 7 – Calculation of the weight per feature by means of the genetic algorithm; 8 - Extraction of keywords by integrating topical and statistical features; 9 – Establishment of the image indexes library; 10 - Establishment of the inverted indexes of keywords; 11 – Display of results; 12 – Matching; 13 – Inputting keywords; 14 – Inputting images; 15 - User Inquiry

Figure 8. Flow diagram of the hybrid search scheme

As for the text-based image search, the AHP-based weight analysis method is applied for analysis of context information of images. On the other hand, the genetic algorithm is utilized as for the content-based image search to gain the new solution of weights of visual features of images; finally, more precise image search scheme may be achieved. Based on the hierarchy search thought, preliminary search is first carried out by means of keywords and the secondary fined search is then performed by means of Search By Image from the primary searched result set.

4.4 Hybrid image search scheme

4.4.1 Experimental data set and evaluation indexes. The experiment data set is the same as that in Section 4.1 and the indexes for evaluation of performances of the algorithm are the same as those mentioned in Section 4.2.

4.4.2 Experiment settings and results analysis.

Experiment preprocessing:
- Establishment of the image library
- Establishment of the image indexes

An indexer sub-thread is established for each image folder; at the same time, while the start method of Indexer is called in the thread, an Indexer file will be established for each image in the folder for searching. The indexer sub-thread is under control of the Thread Pool in the outermost layer of the program.
- Calculation of weights of features of images

Prior to operation of the genetic algorithm, a group of similar image sets may be necessarily selected by people as the search standards. Our image set met such requirement during the experimental period. Some images in the image set are shown in Figure 9.

Figure 9. Some images in the image set

After selecting the similar image set, a group of source images shall be selected as inputs to start the program for solution of weights of features of images based on the genetic algorithm.

Weights are calculated by means of Eqs. (16) and (17) given in Section 3.2 for this experiment. All weights of features of images shall be written in the configuration file as the standard weights for searching images. At this point, all basic search work is OK; next, the image search is necessarily created and the search is ready for inputting an image.
After completion of java and html parser, 3 search scenes were established for measurement of performances of the hybrid algorithm, which are as follows:

Experiment 1: Using the text-based image search method

Experiment 2: Using the content-based image search method

Experiment 3: Hybrid image search method based on texts and contents: the addresses for searching keywords and images are input to the search bars at the same time and the multi-tiered search method is utilized for getting similar images.

The examples of sampling results for experiments are shown in Table 4.

### Table 4. Experimentally sampling presentation

| Keyword | Input image | Number of related images (C) | Number of search results (A) | Number of related images in the search results (B) | Recall (B/A) | Recall (B/C) |
|---------|-------------|------------------------------|------------------------------|---------------------------------|--------------|--------------|
| Bottle  | N/A         | 180                          | 196                          | 110                             | 0.561        | 0.611        |
| Bowl    | N/A         | 180                          | 192                          | 118                             | 0.615        | 0.656        |
| N/A     | 1-36.jpg    | 180                          | 138                          | 83                              | 0.601        | 0.461        |
| N/A     | 5-5.jpg     | 180                          | 149                          | 86                              | 0.577        | 0.478        |
| Pots    | 3-8.jpg     | 180                          | 158                          | 112                             | 0.709        | 0.622        |
| Flagons | 4-7.jpg     | 180                          | 148                          | 116                             | 0.784        | 0.644        |

Experiments 1, 2 and 3 were carried out for 50 times, respectively; and the Precision - Recall results are shown in Figure 10 (horizontal ordinate: Recall; and vertical axis: Precision). While Recall is 0, Precision is the mean precision while the first image and 50 search images in the same type have been searched; and while Recall is 20%, Precision is the mean precision of 50 search images while 20% of B (number of related images in the search results) images have been searched; similarly, while Recall is 100%, Precision is the mean precision while B (number of related images in the search results) images have been searched.

![Figure 10 Precision vs. Recall](image-url)
The above figures and tables indicate that the precision for the hybrid image search based texts and contents is higher than that for the text-based or content-based image search; moreover, the feasibility and effectiveness of the hybrid image search algorithm are verified; in addition, such case also indicates text and visual features of any image are able to reinforce each other and complementary.

5. Conclusions

The feasibility and effectiveness of the hybrid image search algorithm were verified here and the hybrid searching was performed based on improvement of performances of the text-based and content-based image search methods; and the primary work is described in detail as follows:
1. Extraction of features of images based on texts was put forward by integrating statistical and topical features to overcome the limitation (extraction of keywords only by means of statistical features of words).
2. Search By Image was put forward by means of the multi-feature fusion method to solve the imprecision for the single-feature content-based image search.
3. The genetic algorithm was used to determine the weights of features to clarify the contributions of various features to contents of any image.
4. The layered-searching method was applied to overcome the difference between the text-based and content-based image search methods; for example, “Semantic Gap” exits so that direct fusion shall be very difficult.

1. References

[1] Wang X J, Zhang L and Ma W Y Duplicate-Search-Based Image Annotation Using Web-Scale Data [J]
[2] Proceedings of the IEEE, 2012, 100(9):2705-2721
[3] Huang Peng 2008 Research on Web Image Retrieval based on the Fusion of Textual and Visual Information [D] (Zhejiang University)
[4] Qiu Zhaowen 2009 Research on Key Techniques of User-oriented Web Image Retrieval [D] (Harbin Institute of Technology)
[5] Xu Hongtao 2009 Study of Webimage Semantic Analysis and Automatic Labeling [D] (Fudan University)
[6] Zhong Yi, Deng Hui and Zhang Hui 2015 Basic Theory and comparison of Content-Based Image Retrieval [I] (China New Telecommunications) chapter 14 pp 106-106
[7] Eakins J and Graham M 1999 Content-based Image Retrieval [R] A Report to the JISC Technology Applications Programme
[8] Datta R, Li J and Wang J Z 2005 Content-based image search: approaches and trends of the new age [J] (Proceedings of MIR’05. 2005), chapter3 pp 36-43
[9] Yuan Yunli 2013 The Research on the Construction Engineering Project Risk Management based on Fuzzy Analytic Hierarchy Process (FAHP) [D] (Chongqing University)
[10] Tian L, Zheng D and Zhu C 2011 Research on image classification based on a combination of text and visual features.[C]// Eighth International Conference on Fuzzy Systems and Knowledge Discovery (Shanghai, China.) pp 1869-1873
[11] Wu Wei 2014 Image Classification and Retrieval Based on Semantic Distance and Features Combination [D] (Inner Mongolia University)
[12] Deng Xue, Li Jiaming, Zeng Haojian, et al 2012 Research on Computation Methods of AHP Wight Vector and Its Applications [J] (Mathematics in Practice and Theory) chapter 7 pp 93-100