

Analysis on Digital Image Re-Ranking Algorithm based on Multi-feature Fusion

Zhao Huanping^{1,*}, Liu Yan² and Shi Huojie³

¹ College of Computer and Information Engineering, Nanyang Institute of Technology, Nanyang 473004, China

² College of Information Engineering, Zhengzhou Institute of Technology, Zhengzhou 450044, China

³ Department of Biological Systems Engineering, Washington State University, Pullman, MI 98101, United States

Received 19 December 2016; Accepted 5 April 2017

Abstract

Digital image re-ranking algorithms directly retrieve images. Without text messages or images in these algorithms, a semantic gap between the visual feature and high-level semantics of the image exist, which will affect the sequencing performance of retrieval results. In this study, a digital image re-ranking algorithm based on multi-feature fusion was proposed to eliminate the effects of semantic gap on image re-ranking. This algorithm explored the deep relationship between the visual feature and semantic attributes of digital images by combining the background visual feature and semantic attributes. The mapping dictionary between visual features and semantic attributes was also constructed from the perspective of transition probability. The relevant weight components of visual feature in each dimension were calculated based on the query of context between the visual features in the expanded set and the relevance analysis of semantic attributes. Subsequently, the relevant fraction calculation formula was then established from diversity and relevance, and the relevant fraction of each digital image was sequenced. Finally, the proposed algorithm was verified by the comparison analysis using visual rank algorithm and pseudo-relevance feedback (PRF) algorithm. Results demonstrated that the non-interpolating average precision (AP) of the proposed algorithm in public databases Holiday, Oxford 5k, and Paris are 0.8456, 0.5412 and 0.7075, which are 2.4%, 3.29%, and 3.61%, and 0.86%, 4.24% and 6.4% higher than those of visual rank algorithm and PRF algorithm, respectively. The AP of the proposed algorithm increased significantly. Research results showed that the elimination of semantic gap could improve digital image re-ranking to some extent. Related conclusions could provide technical support to network-based image retrieval.

Keywords: Image re-ranking, Steerable pyramid, Visual semantic attributes, Dictionary learning

1. Introduction

Given the rapid development of multimedia and internet technologies, as well as the occurrence of Web 2.0 and massive data sharing, multimedia data on the internet are constantly increasing. One of important multimedia is the digital image, which is highly appreciated by users. Both text-based and image feature-based search [1,2] cannot generate perfect results. Therefore, key attention should be paid to achieving search results more efficiently and accurately by combining them.

Digital image re-ranking technology [3] is developed under the abovementioned circumstance. This technology extracts visual and semantic features from preliminarily searched images and then constructs a re-ranking model. With respect to the development of existing digital image re-ranking technologies, most re-ranking technologies often fail to address the semantic gap between the background visual feature and high-level semantic attributes of images. Xie and Lu [4] constructed the similarity map of images based on background features. However, they did not consider the semantic analysis of the image and the users' searching habits, resulting in the low search precision ratio. Yang and

Li et al. [5] introduced a relevance feedback algorithm based on the R_{norm} re-ranking mechanism, which significantly improved the retrieval accuracy and stability. However, they neglected the high-level semantic features during feature extraction. In this study, the retrieval intention of the users was explored between the background visual feature and high-rise semantic attributes, which could eliminate the semantic gap to some extent, reduce consumption of computing resources in the re-ranking process effectively, and increase re-ranking efficiency.

2. State of the art

Existing research on digital image re-ranking algorithms mainly concerns two aspects of image retrieval, namely, relevance and diversity of retrieval results. Most of these algorithms explore relevant data of images by background visual features for image re-ranking. The algorithms are mainly divided into unsupervised and supervised learning re-ranking algorithms [9]. Clustering [10] is one of the most common unsupervised learning re-ranking algorithms. Wei and Zhao et al. [11] of Beijing Jiaotong University ranked initial results, which were returned by a certain level using multi-modal features through the clustering algorithm. Then, they inferred query with high relevance based on the collaboration of categories at all levels of different modals.

*E-mail address: spring1610@163.com

ISSN: 1791-2377 © 2017 Eastern Macedonia and Thrace Institute of Technology. All rights reserved.

Jing et al. [12] clustered the network text search and image search results, thereby obtaining some cluster names related with the query. Subsequently, they used the cluster names as the search terms and submitted them to an image search engine. The returned results were used as the re-ranking results of each cluster. However, the algorithm used depends on the number of clusters and priori knowledge. Tian and Yang et al. of Microsoft Research Asia [13] proposed the general framework of image re-ranking from the Bayes perspective. The key of this algorithm is constructing an appropriate image model. Supervised learning re-ranking [14–18] determines the sample training classifier from initial results and then accomplishes re-ranking with the classifier. Supervised learning re-ranking algorithms include the pseudo-relevance feedback (PRF) algorithm used by Yan et al. [19] of Carnegie Mellon University, which was also the conceptual test modal algorithm used by Kennedy and Chang et al. [20] of Columbia University. All of these algorithms couldn't ensure either sample quantity or sample quality.

The re-ranking algorithm based on the single modal feature of an image is unsatisfactory due to the noises in the text information of images, subjectivity, and the label ambiguity of image annotation, as well as the semantic gap between the image visual and semantic features. Most existing image re-ranking technologies neglect the semantic content of images and fail to solve the semantic gap. To solve the semantic gap, the current study explored the retrieval intention of the users using preliminary image retrieval results, recognized visual words related to the retrieval intention of the users, and calculated the similarity between each retrieval and queried image for the re-ranking of initial results. These processes offset the effects of semantic gap and improve the ranking of retrieval results.

Section 3 introduces the method for extracting the visual feature and semantic attributes of digital images, the design of visual semantic dictionary, the relevance analysis design of visual semantic terms, and the context relevance design of visual words. Section 4 presents the experimental design and result analysis. Section 5 presents the conclusions.

3. Methodology

3.1 Feature extraction

Image features can be divided into background visual features and high-level features [21]. Selecting an appropriate feature extraction method helps in increase the understanding of image content and narrowing of semantic gaps.

3.1.1 Visual feature extraction

(1) Color feature extraction

Color is crucial in expressing the content of images. The color histogram is the most common way of expressing color features because of its intrinsic translation scaling and rotation invariant. The HSV color space corresponds to one conical subset in the coordinate system, where circumference H, radius S, and high V denote color, saturation, and color brightness, respectively. To calculate the HSV color histogram, the conical space is divided into 256 small color spaces and images are quantified in the HSV color space using the 16:4:4 non-uniform quantizing scheme. To reduce the color difference at the quantization border, two adjacent values of color histogram along the H direction

are added, which are a three-dimensional quantization color histogram that represents the color feature of images.

(2) Texture feature extraction

In comparison with color and shape features, texture feature represents shape distribution. In Ref. [22], image retrieval was accomplished by textures using the Gaussian pyramid; however, texture direction was not used completely in the decomposition of the pyramid. A steerable pyramid [23] is a multi-scale and multi-direction image decomposition method with automatic conversion. This method has translation and rotation invariants without spectrum aliasing. Therefore, the steerable pyramid can extract multi-directional texture features in images well.

The decomposition and reconstruction system structures of steerable pyramid in the frequency domain are shown in Fig. 1. The left portion is the decomposition of the steerable pyramid, whereas the right portion is the reconstruction process. $H_0(\omega)$ is the non-directional high-pass filter; $L_0(\omega)$ is a low-pass filter; $L_1(\omega)$ is a narrow-band low-pass filter; $B_k(\omega)$, where $k = 0, 1, \dots, n$, is a sub-band filter along $n + 1$ directions; $\uparrow 2$ is down-sampling; and $\downarrow 2$ is up-sampling.

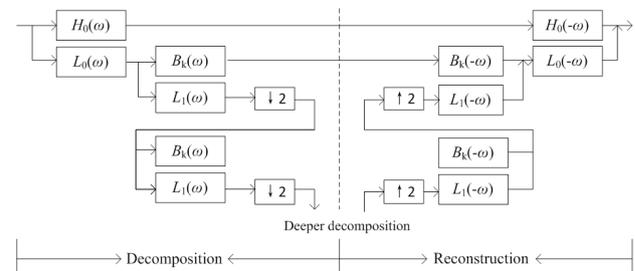


Fig. 1. Decomposition and reconstruction structures of the steerable pyramid

To provide rotation invariant to the algorithm, the principal direction in image texture is identified by the histogram along the gradient direction and is used as the 0° direction for feature extraction and matching, thereby realizing directional normalization. The feature extraction steps are as follows.

First, steerable pyramid decomposition is used to obtain multi-directional and multi-scale sub-band images. The level number of steerable pyramid decomposition and the quantity of direction can be set according to the required number. The key to steerable pyramid decomposition of images is the filter design. In the frequency domain corresponding to Fourier transform, definitions of the first low-pass and high-pass filters in the polar coordinates of frequency domain are as follows:

$$L_0(r) = \frac{L(r/2)}{2} \tag{1}$$

$$H_0(r) = H\left(\frac{r}{2}\right) \tag{2}$$

where r and θ are the coordinates in the polar coordinate system of the frequency domain. The trigonometric function means that the transfer functions of low-pass filter and directional bandpass filter are calculated as follows:

$$L(r) = \begin{cases} 2 & r \leq \frac{\pi}{4} \\ 2 \cos\left(\frac{\pi}{2} \log_2\left(\frac{4r}{\pi}\right)\right) & \frac{\pi}{4} < r < \frac{\pi}{2} \\ 0 & r \geq \frac{\pi}{2} \end{cases} \quad (3)$$

$$B_k(r, \theta) = H(r)G_k(\theta) \quad K \in [0, k-1] \quad (4)$$

where $B_k(r, \theta)$ is the bandpass filter along the k th direction. The high-pass filter is defined as follows:

$$H(r) = \begin{cases} 1 & r \geq \frac{\pi}{2} \\ \cos\left(\frac{\pi}{2} \log_2\left(\frac{2r}{\pi}\right)\right) & \frac{\pi}{4} < r < \frac{\pi}{2} \\ 0 & r \leq \frac{\pi}{4} \end{cases} \quad (5)$$

$$H(r) = \begin{cases} 1 & r \geq \frac{\pi}{2} \\ \cos\left(\frac{\pi}{2} \log_2\left(\frac{2r}{\pi}\right)\right) & \frac{\pi}{4} < r < \frac{\pi}{2} \\ 0 & r \leq \frac{\pi}{4} \end{cases}$$

$$G_k(\theta) = \begin{cases} \alpha_k \left(\cos\left(\theta - \frac{\pi k}{K}\right)\right)^{K-1} & \left|\theta - \frac{\pi k}{K}\right| < \frac{\pi}{2} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where

$$\alpha_k = 2^{(K-1)} \left(\frac{(K-1)!}{\sqrt{K[2(K-1)]!}}\right) \alpha_k = 2^{(K-1)} \left(\frac{(K-1)!}{\sqrt{K[2(K-1)]!}}\right).$$

Second, the binarization process is conducted to each sub-band image, which is obtained by pyramid decomposition. The binarization threshold is determined by the maximum value of the sub-band image, in which points that are higher than the threshold are set to 1, and those smaller or equal to the threshold are set to 0.

For the k th sub-band image I_{sk} , the binarization threshold T_{sk} is obtained from:

$$T_{sk} = \alpha \times \max(I_{sk}) \quad (7)$$

where α can be determined by experiment, and the sub-band binary image I_{swk} is obtained after binarization.

Finally, the row and column projections of I_{swk} are calculated to obtain the projection vector. The row and column projections of I_{swk} represent the features of corresponding scale and directional sub-band image, as shown as follows:

$$P_{rk} = \left(\frac{\sum_{j=1}^c I_{swk}(1, j)}{c}, \frac{\sum_{j=1}^c I_{swk}(2, j)}{c}, \dots, \frac{\sum_{j=1}^c I_{swk}(i, j)}{c}, \dots, \frac{\sum_{j=1}^c I_{swk}(r, j)}{c} \right) \quad (8)$$

$$P_{ck} = \left(\frac{\sum_{j=1}^c I_{swk}(i, 1)}{r}, \frac{\sum_{j=1}^c I_{swk}(i, 2)}{r}, \dots, \frac{\sum_{j=1}^c I_{swk}(i, j)}{r}, \dots, \frac{\sum_{j=1}^c I_{swk}(i, c)}{r} \right) \quad (9)$$

where c and r are the column and row numbers of the sub-band image, respectively.

3.1.2 Semantic attributes of image

Images with similar visual features may not have similar semantic contents due to the semantic gap between the background visual features and high-level semantic attributes. Therefore, an expression method is required to depict the semantic content of the image [24].

(1) Construction of image training set

For the semantic attributes of images, 2,659 semantic contexts are defined. The semantic labeled image set is expressed as $T = \{1, 2, \dots, C\}$, which is acquired by using each semantic label c as one text query and submitting it to the search engine, thereby obtaining the positive sample set of the training image for the corresponding semantic label.

(2) Predictor learning

Different image features reflect different aspects of image information. To use comprehensive image information in training the predictor of corresponding semantic attributes and extracting multiple visual features of the image, the multicore learning algorithm is used to accomplish the training of corresponding semantic attribute prediction.

Image training set is expressed as $(x_i, y_i) i = 1, 2, \dots, N$, and multiple visual features of the image are expressed as $\{f_1, f_2, \dots, f_m\}$. The m visual feature (f_m) corresponds to the kernel function k_m . For a given image x , the kernel function response value of f_m can be expressed as

$$K_m(x) = [k_m(x, x_1), k_m(x, x_2), \dots, k_m(x, x_N)]^T \quad (10)$$

where $K_m(x, x_i)$ is the distance between x and x_i in the kernel space. Each visual feature uses the Gaussian function.

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (11)$$

Multiple image features are integrated by the fusion coefficient β_m . The distance value between x and x' in the kernel space can be expressed as

$$K^*(x, x') = \sum_{m=1}^F \beta_m k_m(x, x') \quad (12)$$

Considering the complexity coefficient of the algorithm, a constraint $\beta_m > 0$ is added, and $\sum_m \beta_m = 1$. The objective function for the optimization during the semantic classifier training of semantic label c can be expressed as

$$\min_{\alpha, \beta, b} \frac{1}{2} \sum_{m=1}^F \beta_m \alpha_c^T K_m \alpha_c + C \sum_{i=1}^N L(y_i, b_c + \sum_{m=1}^F \beta_m K_m(x)^T \alpha_c) \quad (13)$$

$s \cdot t \sum_{m=1}^F \beta_m = 1, \beta_m \geq 0$
 $m = 1, \dots, F$

where $L(y, t) = \max(0, 1 - yt)$ is the final loss function of training data prediction.

After the training learning of the model optimal parameters is accomplished, the prediction function of c can be rewritten as

$$F_c(x) = \sum_{m=1}^F \beta_m (K_m(x)^T \alpha_c + b_c) \quad (14)$$

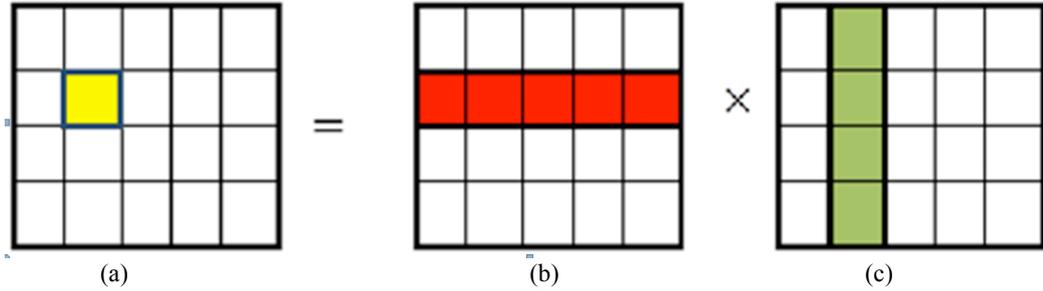


Fig. 2. Dictionary learning reconstruction structure. (a) Distribution matrix of visual words B . (b) Visual semantic mapping of dictionary D . (c) Probability distribution matrix of semantic attributes A

The objective function of visual semantic mapping dictionary learning reconstruction is

$$\min_D \frac{1}{2} \|B_{d \times k} - D_{d \times n} A_{n \times k}\|_2 \quad (16)$$

$s \cdot t \cdot \sum_i D_{i,j}^2 \leq 1, \forall 1, 2, \dots, n$

D , which minimizes the objective function, is searched using the Lagrange dual method. The Lagrange operator is expressed as

$$L(D, \vec{\lambda}) = \text{trace}((B - DA)^T (B - DA)) + \sum_{j=1}^d \lambda_j (\sum_{i=1}^d D_{i,j}^2 - 1) \quad (17)$$

Each $\lambda_j \geq 0$ is a dual variable, and the Lagrange dual operator can be expressed as

$$F(\lambda) = \min L(D, \vec{\lambda}) = \text{trace}(B^T B - BA^T (AA^T + \text{diag}(\vec{\lambda}))^{-1} (BA^T)^T - \text{diag}(\vec{\lambda})) \quad (18)$$

$F(\lambda)$ gradient and Hessian can be calculated as

The prediction function can be viewed as a series of linear combination of predicted values of the prediction function. Each sub-prediction function can be expressed as

$$f_m(x) = K_m(x)^T \alpha_c + b_c \quad (15)$$

Each semantic corresponds to a group of optimal parameters. Parameter optimization is solved using the LPBoost method.

The prediction fraction of each image (x) by each semantic classifier is used as the membership fraction of this semantic label. Thus, membership fractions of each semantic labels can be connected into eigenvectors, that is, semantic attributes of the image x .

3.2 Re-Ranking structure and method

3.2.1 Production of semantic dictionary

Visual semantic mapping dictionary can be obtained by dictionary learning reconstruction (Fig. 2).

$$\frac{\partial F(\vec{\lambda})}{\partial \vec{\lambda}} = \|BA^T (AA^T + \text{diag}(\vec{\lambda}))^{-1} e_i\|^2 - 1 \quad (19)$$

$$\frac{\partial^2 (F(\vec{\lambda}))}{\partial \lambda_i \lambda_j} = -2((AA^T) + \text{diag}(\vec{\lambda}))^{-1} (BA^T)^T BA^T (AA^T + \text{diag}(\vec{\lambda}))^{-1} ((AA^T + \text{diag}(\vec{\lambda}))^{-1})_{i,j} \quad (20)$$

The Lagrange parameter λ can be obtained by calculating the maximum of $F(\lambda)$ of $L(D, \vec{\lambda})$. In $L(D, \vec{\lambda})$, the optimal visual mapping dictionary D is obtained using the gradient descent method, which can be expressed as

$$D^T = (AA^T + \text{diag}(\lambda))^{-1} (BA^T)^T \quad (21)$$

3.2.2 Relevance analysis of visual semantic words

During image re-ranking, the desired visual words of the users are searched and reflected by the mapping dictionary. The importance of these visual words should be considered in similarity calculation, thereby obtaining the re-ranking results that satisfy the retrieval intention of the users. The calculation of query semantic relevant weights of visual words is shown in Fig. 3.

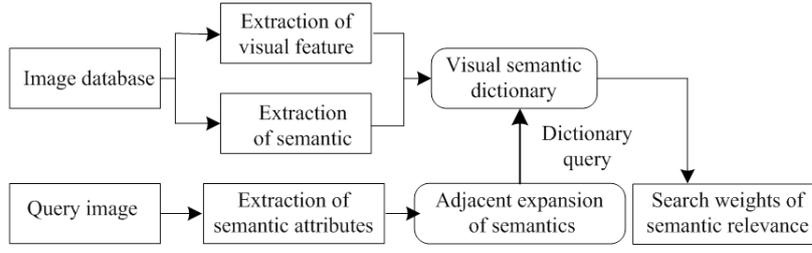


Fig. 3. Query visual semantic mapping dictionary

Definition 1: If two images (I_i and I_j) meet $I_i \in N_k(I_j) \cap I_j \in N_k(I_i)$, then they are called a pair of neighbor.

This study constructs the query expansion set of images by neighboring images. The specific steps are as follows.

First, the primary m images in the initial ranking are used as the objective set. Each image in the objective set is viewed as one node, and the similarity between images is viewed as the weight of connection side between nodes. All images in the objective set establish a K-NN graph.

Second, the neighboring image of the queried image is searched from the K-NN graph to form the expansion set of query images.

The semantic attribute of each image in the expansion set uses the sigmoid function for mapping to obtain the probability distribution of semantic attributes. The probability distribution (Q_{avg}) of semantic attributes, which are related with query and can reflect the retrieval intention of the users, can be obtained as follows:

$$Q_{avg} = \frac{1}{num} \sum_{k=1}^{num} f(A_k) \quad (22)$$

where num is the size of the query expansion set, A_k is the semantic attribute of the k th image, and f is the corresponding sigmoid function.

The query semantic relevant weight vector of visual words can be calculated as

$$\omega_1 = D_{d \times n} \times Q_{avg_{n \times d}} \quad (23)$$

3.2.3 Context relevance analysis of visual words

If one visual word occurs in several images in the initial ranking results, it can reflect the retrieval intention of the users to some extent. Meanwhile, the other words that are highly correlated with this word deserve higher attention. In this study, the context relevant weight of each visual word is learned by the idea of random walk based on the context relationship of visual words. The context weight learning process of visual words is shown in Fig. 4.

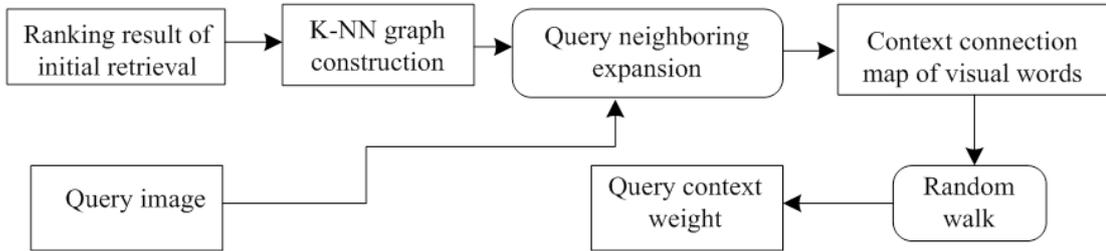


Fig. 4. Context learning of visual words

Vector $\omega_2 = (b_1, b_2, \dots, b_n)$ is used to represent the context significant weight of visual words. The context relationship among visual words is expressed by the matrix

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} & \dots & a_{1M} \\ \dots & & & & \\ a_{m1} & \dots & a_{mn} & \dots & a_{mM} \\ \dots & & & & \\ a_{M1} & \dots & a_{Mn} & \dots & a_{MM} \end{bmatrix} \quad (24)$$

Suppose ω_2^i is the result of the i th iteration, then,

$$\omega_2^i = \omega_2^{i-1} A \quad (25)$$

Since each iteration has some probability to return to itself from one visual word. Each iteration result can be rewritten as

$$\omega_2^i = \alpha * \omega_2^{i-1} A + (1 - \alpha) \pi \quad (26)$$

where $\pi = \omega_2^0 = (\frac{1}{n}, \dots, \frac{1}{n})$. ω_2^i will approach ω_2 infinitely.

If ω_2^i and ω_2^{i-1} , which are obtained from two iterations, have a small difference, then stop the iteration. The final iteration formula is

$$\omega_2^{i+1} = (1 - \beta) \pi + \alpha A^T \omega_2^i. \quad (27)$$

The context of visual words is measured by the conditional transition probability $p(v_i | v_j)$. The skip probability between visual words is expressed as

$$p(v_i | v_j) = \sum_{k=1}^K p(v_i | I_k) p(I_k | v_j) \quad (28)$$

3.2.4 Image re-ranking

The retrieval intention of the users is explored by the initial image retrieval results. Visual words, which are closely related with the retrieval intention of the users, are found, and the similarity between each retrieval ranking image and

query image is calculated. Key attention is paid to the matching degree of these important visual words. The algorithm framework is shown in Fig. 5.

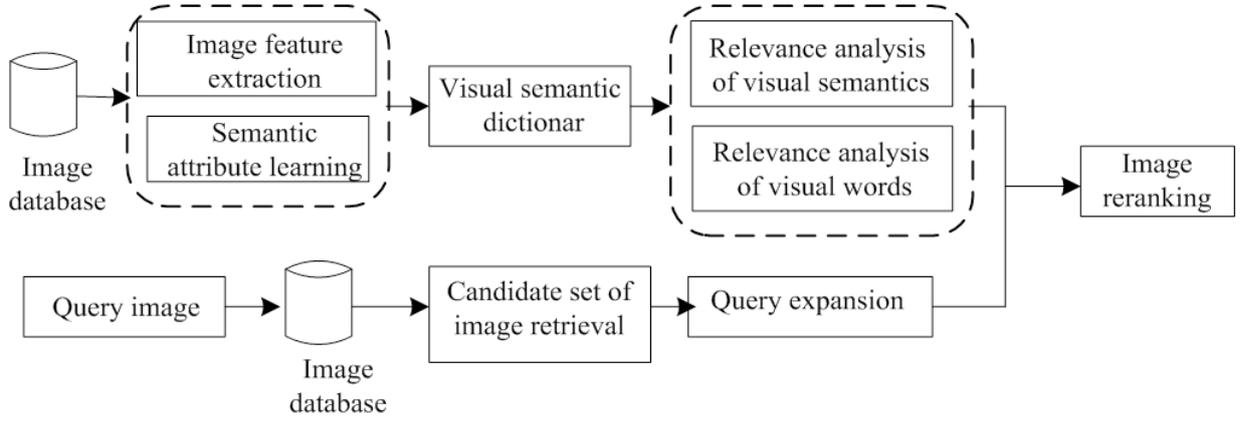


Fig. 5. Structure of image re-ranking algorithm

The importance of each visual word is considered completely to recalculate the similarity of each image in the initial retrieval candidate set and the query image. The relevance fraction of images is calculated for image re-ranking. The image ranking fraction can be calculated as

$$F(I) = sim(I, Q) = \frac{\sum p(v_k | I_i) * p(v_k | Q)}{\|X_i\|_2 * \|X_Q\|_2} f(R_k) \quad (29)$$

where $p(v_k | I_i)$ is the occurrence frequency of visual words in the corresponding images, $f(x) = e^x$ is a monotone non-decreasing smooth function, and $\|X_i\|_2 * \|X_Q\|_2$ is the corresponding normalization factor.

4. Results analysis and discussion

4.1 Experimental design

The validity of the proposed algorithm is evaluated by an experiment based on three public image databases, namely Holiday, Oxford 5k, and Paris. The average precision (AP) ratio of each relevant sample is used as the result for evaluation.

The AP calculation formula of the first K images is

$$AP@K = \frac{1}{Z} \sum_{i=1}^K precision(i) * recall(i) \quad (30)$$

where $precision(i)$ is the precision ratio of the first i images, and $recall(i)$ is the recall ratio of the first i images. The calculation formulas are

$$precision(i) = \frac{true_num(i)}{i} \quad (31)$$

$$recall(i) = \frac{trun_num(i)}{relevent(k)} \quad (32)$$

where $trun_num(i)$ is the number of occurrence of the relevant samples in the first i images, and $relevent(k)$ is the number of relevant samples in the first k images.

The AP values of several groups are obtained through retrieval ranking of the query images, and the average AP value is used as the final evaluation parameter of the experiment.

$$MAP = \frac{1}{num} \sum_k AP(k) \quad (33)$$

where num is the total retrieval time.

The comparative experiments of this algorithm are introduced as follows.

a) Different ranking strategies for comparison in the experiment:

Baseline1: Ranking results obtained from the similarity fraction are directly calculated by Bow features.

Baseline2: The similarity fraction ranking is directly calculated by semantic attributes.

Baseline3: The 8192-dimensional Bow features and 2659-dimensional semantic attributes are connected in a series. Principal component analysis algorithm is used to reduce the dimension to 1,000, and then the similarity fraction ranking is directly calculated.

b) Existing image re-ranking algorithms serve as a reference in the experiment.

Visual rank algorithm uses each image in the initial ranking result as one node, constructs an image model based on image similarity, and estimates the relevance fraction of each image based on the random walk model. Details are introduced in Ref. [25].

PRF algorithm predicts the image relevance in the initial retrieval results based on its support vector machine classification learning training by the PRF from the perspective of image classification and uses predicted fraction as the relevance fraction of each ranking image. Details are introduced in Ref. [26].

4.2 Experimental results and analysis

The MAP comparison results of the first 200 searched images in the three databases are listed in Tables 1 and 2.

Table 1. Average AP of different ranking strategies on three databases

Dataset	Baseline1	Baseline2	Baseline3	Proposed algorithm
Holiday	0.8253	0.8109	0.8107	0.8456
Oxford	0.5154	0.4651	0.4708	0.5412
Paris	0.6571	0.5703	0.5652	0.7075

Table 2. Average AP of existing algorithms on three databases

Dataset	Ref. [25]	Ref. [26]	Proposed algorithm
Holiday	0.8216	0.8127	0.8456
Oxford	0.5051	0.5326	0.5412
Paris	0.6651	0.6435	0.7075

The comparison of $MAP@K$ curves between the proposed algorithm and different image re-ranking algorithms is shown in Fig. 6.

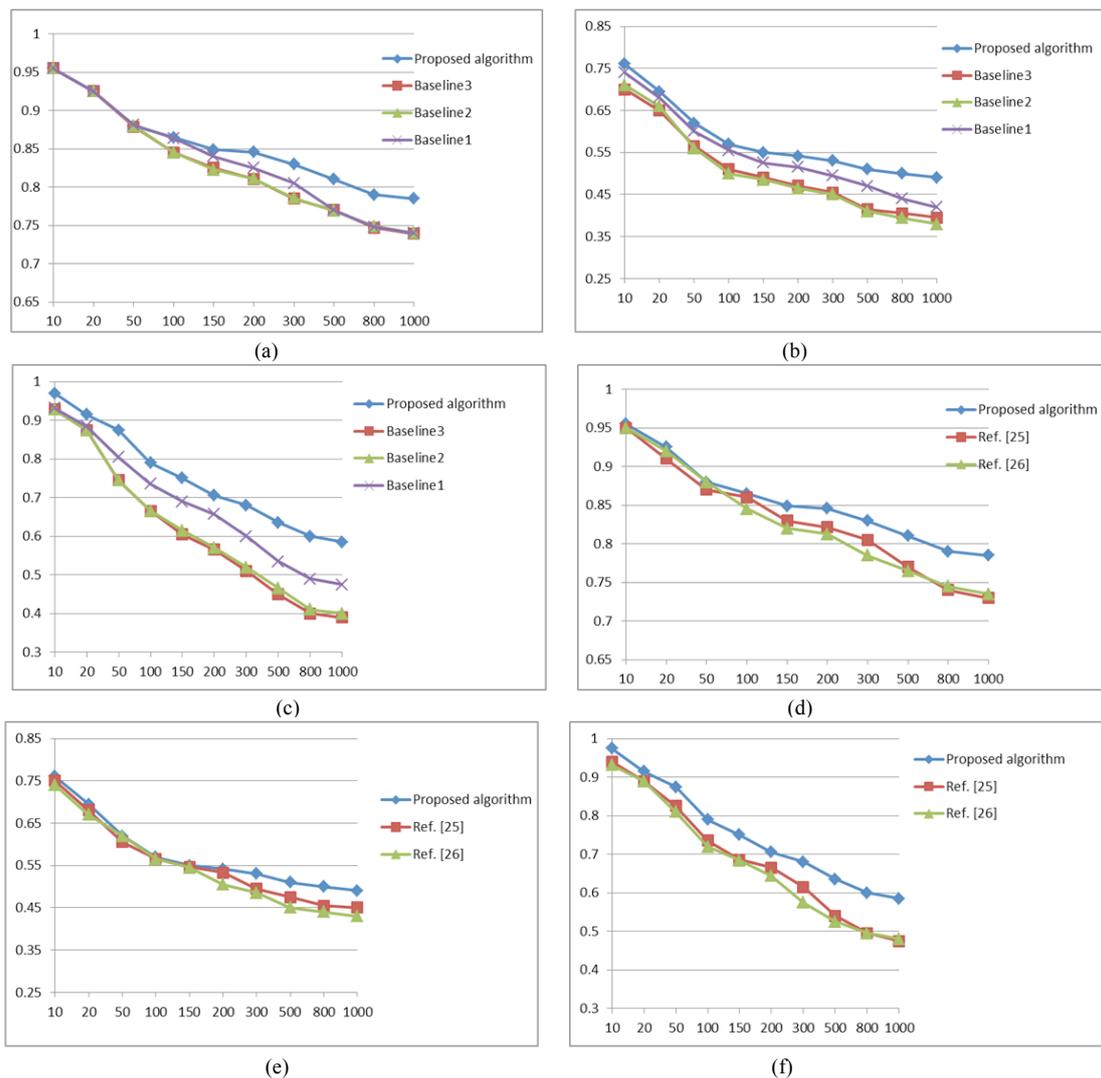


Fig. 6. Average AP curves of the proposed algorithm and different re-ranking algorithms. (a)Holiday. (b) Oxford 5K. (c)Paris.(d)Holiday. (e)Oxford 5K. (f) Paris

As shown in Fig. 6, the proposed algorithm achieves a significantly high performance in retrieval ranking. The re-ranking results of the proposed algorithm are shown in Fig. 7. The left image is a query image used in retrieval. Each retrieval of the three rows in the right are initial retrieval

ranking, re-ranking results based on semantic attributes, and re-ranking results of the proposed algorithm. The red frame represents irrelevant images with the retrieval intention of the users. Evidently, the proposed algorithm is superior to other algorithms in image re-ranking performance.

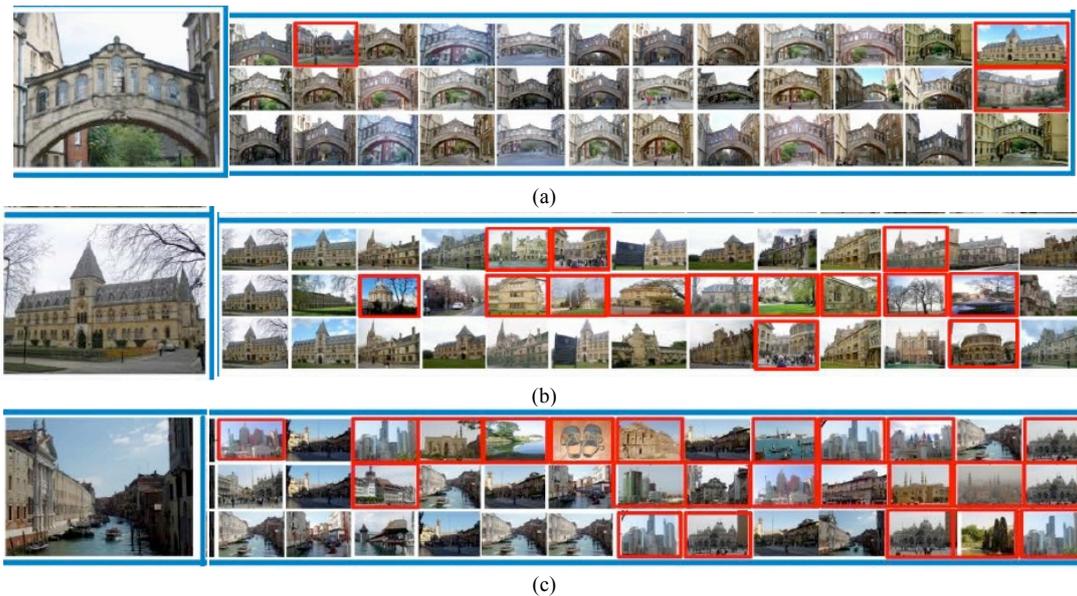


Fig. 7. Retrieval results of the proposed algorithm

5. Conclusions

Digital images with similar visual feature may not have similar semantic contents. To eliminate the effects of the semantic gap between background visual feature and high-level semantic attributes in digital images on image re-ranking, a digital image re-ranking algorithm based on multi-feature fusion is proposed by using Holiday, Oxford, and Paris as database samples. The proposed algorithm is used to explore the retrieval intention of the users and correct the effects of semantic gap on the relevance fraction of digital images. The conclusions drawn are as follows.

(1) The texture features of images are extracted by the steerable pyramid. The steerable pyramid has rotation and translation invariants and can independently decompose different scales and directions of sub-band information independently, in which the sub-band direction is adjustable.

(2) The visual semantic mapping dictionary can be used to search visual words that reflect the visual retrieval intention of the users. Important visual words occur frequently in the query expansion set, and the words that are closely related with these words are also important.

(3) The relevance fraction of each image reflects the diversity and relevance of retrieval results. The weight vector of visual words reflects the visual intention of the users, which is not calculated by a single method but is obtained by linear integration after context learning and semantic relevance analysis of the visual words.

The weight vector of the visual words in the proposed algorithm is considered from diversity and relevance. This algorithm can accurately reflect the retrieval intention of the users, thereby providing convenient and accurate technological support for network-based image retrieval. However, the proposed algorithm has poor recognition accuracy to the semantic attributes of digital images. Another appropriate way to express the semantic attributes of digital images should be investigated.

Acknowledgements

The study was supported by the Science and Technology Research Project of Henan Province (142102210554).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence



References

1. Fan, J., Shen, Y., Yang, C., Zhou, N., "Structured max-margin learning for inter-related classifier training and multipliable image annotation". *IEEE transactions on Image Processing*, 20(3), 2011, pp.837- 854.
2. Yan, R., Huet, B., Sukthankar, R., "Large-scale multimedia retrieval and mining". *IEEE multimedia*, 18(1), 2011, pp.11-13.
3. Zhang, J., Qu, X.J., Jin, Zh., "Overview of content-based image and video search reranking". *Computer Engineering and Applications*, 47(29), 2011, pp.171-174.
4. Xie, H., Lu, Y.M., "Content-based image re-ranking technology in search engine". *Journal of Computer Applications*, 33(2), 2013, pp.460-462.
5. Yang, D.S., Li, M., Li, L., "New relevance feedback based on re-ranking mechanism in content-based image retrieval". *Computer Engineering and Applications*, 44(20), 2008, pp.155-158.
6. Lu, L. Zh., Liu, R.Y., Liu, N., "Remote sensing Image Retrieval Using Color and Texture Fused Features". *Journal of Image and Graphics*, 9(3), 2004, pp.328-333.
7. Qu, X. J., "Content-based Web Image Reranking". *Tianjin University*, 2011.
8. Li, H.F., "Research on Reranking Algorithms for Image Search". *Henan University*, 2015.
9. Y, Jing., Baluja, S., "VisualRank: applying page rank to large-scale image search". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11), 2008, pp.1877-1890.
10. Bai, C., Zou, W., Kpala, K., Ronsin, J., "Efficient colour texture image retrieval by combination of colour and texture features in wavelet domain". *Electronics Letters*, 48(23), 2012, pp.1463-1465.
11. Wei, S., Zhao, Y., Zhu, Z., Liu, N., "Multimodal fusion for video search reranking". *JEEE Transactions on Knowledge and Data Engineering*, 22(8), 2010, pp.1191 -1199.
12. Jing, F., Wang, C. H., Yao, Y. H., Deng, K. F., "Web image search results clustering". *International Conference on Multimedia*, Santa Barbara, CA, USA: ACM, 2006, pp.377-384.
13. Tian, X. M., Yang, L. J., Wang, J. D., Yang, Y. C. "Bayesian video search reranking". *International Conference on Multimedia*, Vancouver, Canada: ACM, 2008, pp.131-140.

14. Hwanjo, Y., "SVM Selective Sampling for Ranking with Application to Data Retrieval". *Proceeding of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 2005, pp.354- 363.
15. Donmez, P., Carbonell, J., "Active Sampling for Rank Learning via Optimizing the Area under the ROC Curve". *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, 5478, 2009, pp.78-89.
16. Wei, C., Ghahramani, Z., "Extensions of Gaussian Processes for Ranking: Semi-supervised and Active Learning". *Proceedings of the NIPS 2005 workshop on Learning to Rank*, 2005, pp.29-34.
17. Radlinski, F., Joachims, T., "Active Exploration for Learning Rankings from Clickthrough Data". *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007, pp.570- 579.
18. Tsechansky, M. S., Provost, F., "Active Learning for Class Probability Estimation and Ranking". *Proceeding of 17th International Joint Conference on Artificial Intelligence*, 54(2), 2001, pp.911-917.
19. Yan, R., Hauptmann, A. G., "A boosted reranking approach for video-retrieval". *International Conference on Image and Video Retrieval*, 2004, pp.60-69.
20. Kennedy, L. S., Chang, S. F., "A reranking approach for context-based concept fusion in video indexing and retrieval". *ACM International Conference on Image and Video retrieval*, 2007, pp.333- 340.
21. Chen, X. L., Zheng, Y., Yu, Ch. Ch., "Multi-feature image retrieval based on Tri-training". *Application Research of Computers*, 31(11), 2014, pp.3506-3509.
22. Ning, J. J., Zhou, H. Y., "Study on Image Retrieval methods with Texture on the Basis of Compression". *Computer Applications and Software*, 28(6), 2011, pp.254-256.
23. Chen, X. L., Zheng, Y., Yu, Ch. Ch., "Image Retrieval Based on Steerable Pyramid Binary Image Projection". *Computer Simulation*, 31(1), 2014, pp.438-442.
24. Wang, J. L., "Image Re-ranking Based on Query Dependent Semantic Relevance Analysis". *Xidian University*. 2014.
25. Hsu, W. H., Kennedy, L. S., Chang, S. F., "Video Search Re-ranking Through Random Walk over Document-level Context Graph". *In proceedings of ACM Multimedia*, 2007, pp. 971-980.
26. Yan ,R., Hauptmann, A., Jin, R., "Multimedia Search With Pseudo Relevance Feedback". *Image and video Retrieval Lecture Notes in Computer Science*, 2728, 2003, pp.238-247.