Article

Hierarchical Image Retrieval by Multi-Feature Fusion

Xiaojun Lu, Jiaojuan Wang,Yingqi Hou, Mei Yang, Qi Wang* and Xiangde Zhang *

College of Sciences, Northeastern University, Shenyang 110819, China; luxiaojun@mail.neu.edu.cn(X.L.), 1600123@stu.neu.edu.cn(J.W.), 18640516280@163.com (Y.H.), 774218764@qq.com (M.Y.) ,
* Correspondence: wangqimath@mail.neu.edu.cn (Q.W.), zhangxiangde@mail.neu.edu.cn (X.Z.);
Tel.: +86-024-8368-7680

Abstract: Aiming at the problems that are poor generalization performance, low retrieval accuracy and large time consumption of existing content-based image retrieval system, the hierarchical image retrieval method based on multi feature fusion is proposed in this paper. The retrieval accuracy rates on Corel5K, UKbeach and Holidays are 68.23(Top 1), 3.73(N-S) and 88.20(mAp), respectively. The experimental results show that the method proposed in this paper can effectively improve the deficiency of single feature retrieval and save time significantly in the premise of a small amount of loss of accuracy.

Keywords: Hierarchical search; Image retrieval; Multi-feature fusion

1. Introduction

Large-scale image retrieval based on visual information is a key technology in many researches. In order to achieve the purpose of retrieving images, we need to describe the content of the image at first. There are a lot of methods to represent the image content, such as color, texture, and so on [1] [2]. From the perspective of image representation, the features of image content can be divided into two categories, local feature [3] and global feature [4]. For the occluded image retrieval task, the local feature has a better performance. However, it may return the image which is irrelevant to the query due to local feature only focus on part of the image. The global feature depicts the overall feature distribution of the image. But it has a large amount of redundant information, and retrieval tasks tend to return images that look similarly but do not correlate. Therefore, given a specific query, the retrieval system using a single feature is often difficult to meet the needs of users. To solve this problem, the retrieval system integrating multiple features has attracted more and more attention due to the characteristics of multiple complementary features. Multi-feature fusion can effectively improve the performance of the retrieval system has been confirmed [5], and there are many researches on multi-feature fusion has been done[6] [7]. At present, the popular method can be divided into two categories: one is fusion on the feature level; the other is rank aggregation. The former merges a variety of features into a new feature for retrieval while the latter fuses retrieval results of different features in the rank stage. The retrieval system based on two fusion methods can improve the retrieval precision. Ronald Fagin et al. [8] used a number of independent "voters" to sort the database images based on their similarity to the query, and then combined the rankings with some efficient fusion algorithms. Oriol Ramos et al. [9] used a non-Bayesian probabilistic framework to solve the problem of classifier combination, and got two linear combination rules to minimize the misclassification rate under certain constraints. Shaoting Zhang et al. [10] proposed a graph-based query fusion method, which re-ordered multiple retrieval sets by chain analysis on the fusion graph.

On the one hand, compared with the single feature retrieval, image retrieval based on

multi-feature fusion increases the retrieval time. On the other hand, many researches on image retrieval have been carried out on large-scale datasets, which may contain up to several million pictures, It is very time-consuming to search for the images we need from the massive images. On this issue, Jia Deng et al. [11] used a hierarchical relationship, which is most suitable for large-scale problems. At the same time, a new hash scheme is proposed, which reduces the computational cost efficiently. Kevin Lin [12] utilized the hidden layer learning binary code to represent the potential concept of the image, and used the hierarchical deep search, which improved the retrieval performance to a great extent.

Inspired by the above work, we present a hierarchical image retrieval system based on multi-feature fusion. Given a query, firstly we extract the CNN feature, and pre-retrieve with the binary CNN feature. Then, the Color, Lbp and GIST features are fused for second retrieval to obtain the search results. Further, we apply our approach to several datasets, and demonstrate its effectiveness and scalability.

2. Materials and Methods

2.1. Datasets

- **Corel-5K[13]** dataset contains 5,000 images that are divided into 50 categories, such as beach, bird, jewelry, sunset, etc. Each category has 100 images. Each image is taken as query in turn, and the other 4999 images are served as retrieval library. The precision of top-r images, namely the ratio of similar images in the returned images, is used as the evaluation standard of the retrieval system.
- **UKbench[14]** dataset is released with 2550 objects, and each object has 4 pictures taken from different visual angles. All the 10,200 images were served as query images. We use N-S score, the average number of similar images in Top-4, as a measurement of retrieval performance.
- **Holidays[15]** dataset includes 1491 personal holiday pictures that are composed of 500 categories. The first image in each category is used as the query, and the remaining images are relevant images. mAP is used to evaluate the retrieval performance.

2.2. Features

- **Color.** For each image, we compute 2,000-dim HSV histogram. (H, S, V are 20, 10, 10).
- **LBP.** For each image, we divide the detection window into 16 * 16 small cells and extract 256-dimensional lbp descriptors.
- **GIST.** Each image is resized to 128 * 128. We use 4 scales with the number of orientations (8, 8, 8, 8), and extract 512-dimensional GIST descriptors.
- **CNN.** Here we use VGG network pre-trained on Imagenet, which consists of 13 convolution layers, 5 pooling layers and 3 fully connected layers. For all the images in the dataset, the fc6 layer (the first fully connected layer) feature is obtained by forward operation. The output of this layer contains a wealth of image information, whose dimension is 4096 dimensions.

2.3. Our Method

In the image retrieval based on multi-feature fusion, we mainly focus on two aspects: one is how to determine the weight of each feature to improve the retrieval accuracy; the other is how to improve the retrieval efficiency.

Traditionally, the weight of feature has two ways to be determined, the global weight and the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 April 2017 doi:10.20944/preprints201704.0174.v1

adaptive weight. The former is an average value or is decided by experienced experts, which leads the retrieval system to have poor generalization performance and low retrieval performance for different retrieval images. The latter is derived from retrieval feedback based on this feature, which is better than the global weight. However, in the sum or product fusion, the distinction between good features and bad features is not obvious. If the weight of the bad features in the retrieval work is large, it will also reduce the retrieval performance to a certain extent.

The traditional image retrieval system based on multi-feature fusion is a single-layer image retrieval system. The retrieval process is shown in Figure 1. That is, the system receives a query image and extracts a plurality of features of query and images in the image database. Then multiple features are fused to get the comprehensive similarity measure. Finally, the system returns relevant images to the user according to the comprehensive similarity measure. At each retrieval, the retrieval system needs to extract the features of all images and calculate the similarity between query and all images in the image database, which may lead to low retrieval efficiency and poor adaptability to the retrieval task on the large-scale image database.

In order to overcome the shortcomings of the traditional global weight method, a hierarchical image retrieval system based on multi-feature is proposed. The basic framework of the retrieval system is shown in Figure 2. Next, the hierarchical retrieval and multi-feature fusion will be described in detail.

2.3.1. Hierarchical retrieval

As shown in Figure 2, in the hierarchical retrieval system proposed in this paper, pre-retrieval is performed based on the binary CNN feature. Then Color, Lbp, GIST and binary CNN features are fused to perform second search within the pre-search results.

- **Pre-retrieval.** In the pre-retrieval stage, the CNN feature of query image and images in image database are extracted respectively. The CNN feature from the middle layer has a high dimension, so it will take a long time to be carried out the similarity measure by using it.

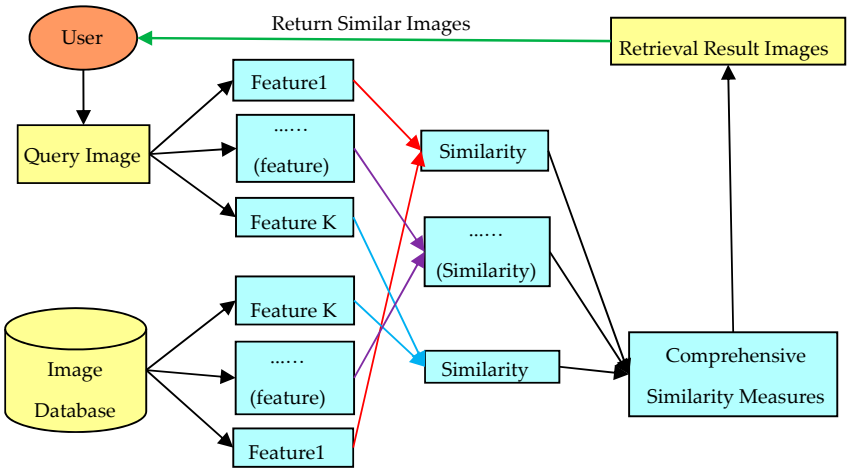


Figure 1. Traditional image retrieval system framework based on multi-feature fusion

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 April 2017 doi:10.20944/preprints201704.0174.v1

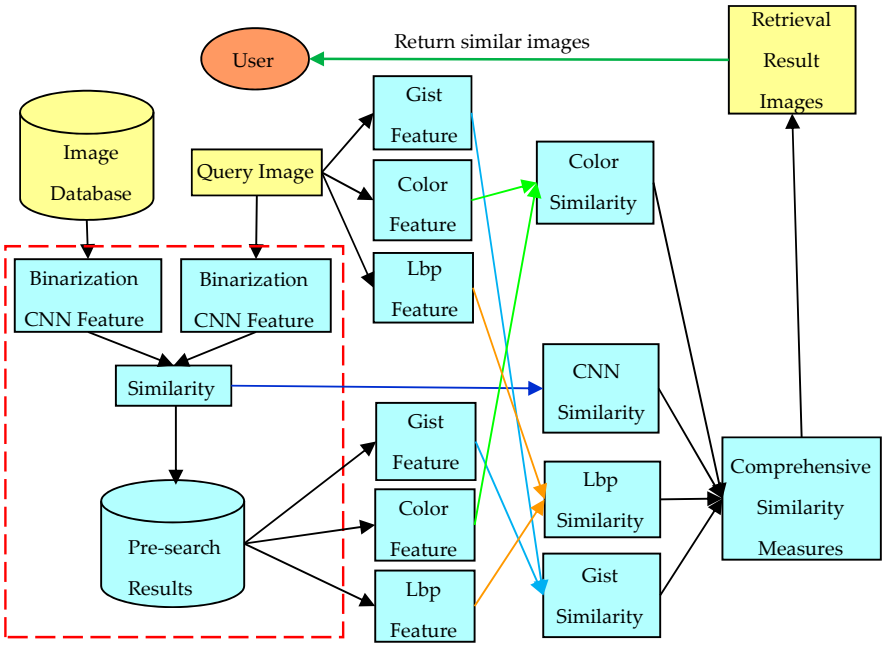


Figure 2. The proposed hierarchical retrieval system framework. The red box represents the pre-retrieval process and the blue box represents the secondary retrieval process.

According to the characteristics of Hamming distance operation, this paper binarizes 4096-dimensional CNN feature as follows:

For each bit of featureF_0, we output binary codes F by:

$$F^j = \begin{cases} 1 & F_0^j \geq ave(F_0) \\ 0 & F_0^j \leq ave(F_0) \end{cases} \qquad j \in \{1, 2, \ldots, n\} \tag{1}$$

Here, $ave(F_0)$ is the mean of feature F_0 , n is the feature dimension 4096.

Then, in the light of feature F, the similarity score between query and images in image database is calculated. We output top-N images as the result of pre-retrieval according to the similarity score.

- **Secondary-retrieval.** In the secondary-retrieval stage, we adjust the similarity measure based on CNN according to the similarity score vector and the retrieval results in the pre-retrieval stage. Next we extract lbp, color, GIST features of query image and images in the pre-retrieval results, and compute similarity measure based on each feature. Then we obtain retrieval performance of each feature by searching similar images. We determine the weight of each feature in accordance with it, and calculate the comprehensive similarity measure. Finally, the retrieval results are obtained according to the comprehensive similarity measure.

The hierarchical retrieval system proposed in this paper improves the retrieval efficiency in two aspects. Firstly, we binarize the high-dimensional CNN feature and calculate the Hamming distance to improve the retrieval efficiency. Secondly, we fuse multiple features to make accurate search within the pre-retrieval results, which reduces the retrieval range and improves the retrieval efficiency to a certain extent.

2.3.2. Multi-feature fusion

Specifically, K features are fused, q is query image, p is a database image. The proposed fusion method is as follows.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 April 2017 doi:10.20944/preprints201704.0174.v1

We normalize the features:

$$F_i(p)=\frac{1}{\sum_{j=1}^{n}p^i(j)}(p^i(1),p^i(2),.....,p^i(n)) \quad (\mathrm{i}\in\{1,2,.....,K\}) \tag{2}$$

$$F_i(q)=\frac{1}{\sum_{j=1}^{n}q^i(j)}(q^i(1),q^i(2),.....,q^i(n)) \quad (\mathrm{i}\in\{1,2,.....,K\}) \tag{3}$$

Here， $p^i(j)$ is the j-th component of the i-th feature of the database image p. $q^i(j)$ is the j-th component of the i-th feature of the query image q. n is the dimension of feature.

We calculate the distance between q and p and normalize it:

$$d^i(k)=d^i(q,p_k)=\sum_{j=1}^{n}\left|q^i(j)-p_k{}^i(j)\right| \qquad k\in\{1,2,.....,m\},\mathrm{i}\in\{1,2,.....,K\} \tag{4}$$

$$D_i(q)=\frac{1}{\sum_{k=1}^{m}d^i(k)}(d^i(1),d^i(2),.....,d^i(m)) \quad (\mathrm{i}\in\{1,2,.....,K\}) \tag{5}$$

Here, $D_i(q)$ is the similarity vector between the query image q and the all database images, which is calculated based on i-th feature. m is the total number of images in the image database.

We calculate the comprehensive measure by fusing multiple features:

$$sim(q)=\sum_{i=1}^{K_1}\tilde{w}_q^{(i)}D_i(q)+\sum_{i=K_1+1}^{K}w_q^{(i)}D_i(q) \tag{6}$$

Here $w_q^{(i)}$, $\tilde{w}_q^{(i)}$, K_1 are the weight of a good feature, the weight of a bad feature, and the number of good features respectively, when the query image is q.

- **Good features and bad features**. $ac_i(q)$ is retrieval performance, which is obtained by $D_i(q)$,

  Good features and bad features are defined as follows：

$$\begin{aligned} & ac\_mean=\frac{\sum_{i=1}^{K}ac_i(q)}{K} \\ if \quad & ac_i(q)>ac\_mean \\ & feature_i\in\{good\_feature\} \\ else \quad & \\ & feature_i\in\{bad\_feature\} \end{aligned} \tag{7}$$

- **Weight Determination.** In order to make better use of good feature in getting the comprehensive measure, after getting the weights, the paper carries on a power calculation to the good weight, which can increase the difference of the good feature and the bad feature. The weights of features are shown as follows:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 April 2017 doi:10.20944/preprints201704.0174.v1

$$\begin{aligned} &if \quad feature_i \in \{good\_feature\} \\ &\quad \tilde{W}_q^i = \exp\left(\frac{ac_i(q)}{\sum_{i=1}^{K} ac_i(q)}\right) \\ &else \\ &\quad W_q^i = \frac{ac_i(q)}{\sum_{i=1}^{K} ac_i(q)} \end{aligned} \tag{8}$$

3. Results

3.1. The effectiveness of feature fusion

In this paper, we conduct retrieval experiments based on maximum fusion, multiplication fusion [16] and sum fusion. And we experiment with adaptive weight and average global weight. The sum fusion is what we use in this paper. The maximum fusion and multiplication fusion are shown as follows.

The maximum fusion:

$$sim(q) = \arg\max_{D_i(q)} \{\max\{\tilde{W}_q^i \mid i = 1,2,...,K_1\}, \max\{W_q^i \mid i = K_1+1, K_1+2, ..., K\}\} \tag{9}$$

The multiplication fusion:

$$sim(q) = \prod_{i=1}^{K_1} \tilde{w}_q^{(i)} D_i(q) \times \prod_{i=K_1+1}^{K} w_q^{(i)} D_i(q) \tag{10}$$

The experimental results on the maximum fusion, multiplication fusion, and sum fusion are 86.15\87.74\88.20, which are shown in Figure 3(a). The experimental results on the adaptive weight and global average weight are shown in Figure 3(b). From the Figure 3(a) we can see that the sum fusion is much better than the other two. Furthermore, it can be seen from the Figure 3(b) that the adaptive weight proposed in this paper is of great performance. Table 1 shows the performance comparison between the single-feature retrieval and multiple-feature retrieval. As what can be seen from the table, the latter's performance is optimal. Experimental results show that the proposed adaptive multi-feature fusion method is effective and has good performance.

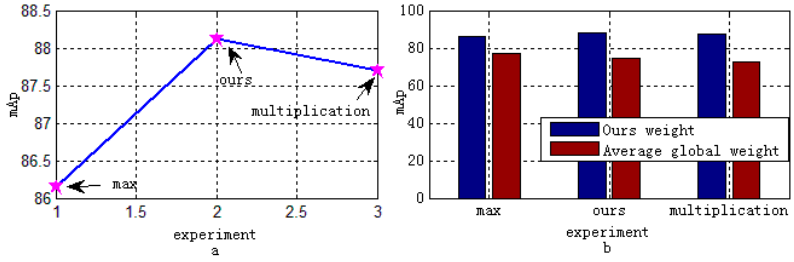


Figure 3. (**a**) A comparison of the retrieval performance based on the three fusion methods of maximum fusion, product fusion and sum fusion. (**b**) The performance comparison between the adaptive weight proposed in this paper and the global average weight.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 April 2017 doi:10.20944/preprints201704.0174.v1

Table 1. On the Holidays dataset, the comparison of search results based on single features and fusion features

Features	CNN	color	lbp	GIST	fusion
mAp	77.25	62.22	37.85	39.03	88.20

3.2. Retrieval performance evaluation

On the UKbeach, Holidays, Corel-5K datasets，we carries out experiments based on different fusion features. The retrieval results are shown in Figure 4. From the figure, we can see that on the Holidays dataset, fusing CNN, color, lbp, GIST features to search image achieves the best performance. On the UKbeach dataset, optimal image retrieval performance can be obtained by fusing CNN, color features. On the Corel-5K dataset, we get the best result by fusing CNN, color, lbp, GIST features.

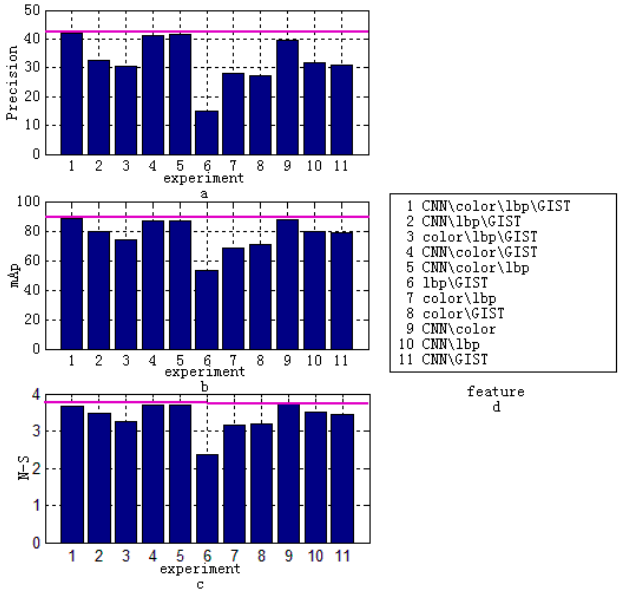


Figure 4. On the UKbeach, Holidays, Corel-5K datasets, the experiments based on different fusion features are compared. Figure (**a**) shows precision comparison retrieved on the Corel5K dataset, here the number of similar images returned is 30. Figure (**b**) shows the mAP comparison retrieved on the Holidays dataset,. Figure (**c**) shows the comparison of N-S values retrieved on the UKbeach dataset.

3.3 The effectiveness of hierarchical retrieval

On the UKbeach dataset, we conduct experiments based on the direct retrieval and hierarchical retrieval respectively. Figure 5 shows that 100 images are randomly selected as query images. We compare the retrieval time based on binary CNN feature and CNN feature, from what we can see that image retrieval based on the binary CNN can achieve the effect of saving time. Then we set the number of pre-retrieved images to 8 and randomly select 100 images as query image. Figure 6 shows the comparison of the retrieval time and N-S score. Figure 7 shows the query images with different N_S value. Figure 8 shows the retrieval results obtained by the hierarchical retrieval and the direct retrieval.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 April 2017 doi:10.20944/preprints201704.0174.v1

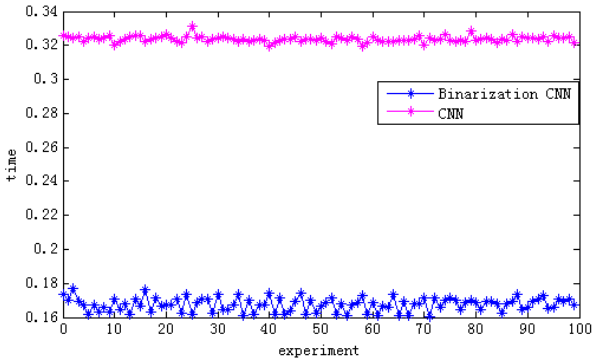


Figure 5. On the UKbeach dataset, 100 images were randomly selected as query images, the retrieval time of binary CNN feature is compared to the retrieval time of CNN feature.

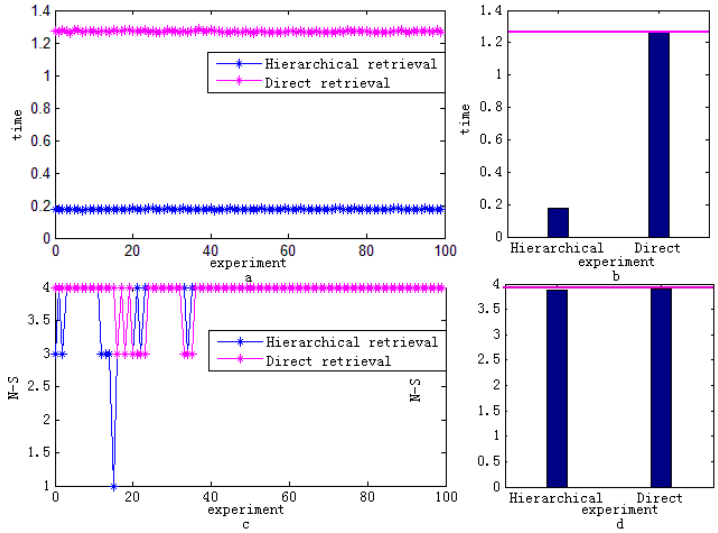


Figure 6. On UKbeach dataset, We randomly selected 100 images as the retrieval image, and set the number of pre-retrieved images to 8. Comparison of direct and hierarchical retrieval performance. (**a**)The comparison of retrieval time in each search. (**b**) The average time of 100 searches comparison. (**c**) The N-S value of each search comparison. (**d**) The average N-S values of 100 searches comparison.

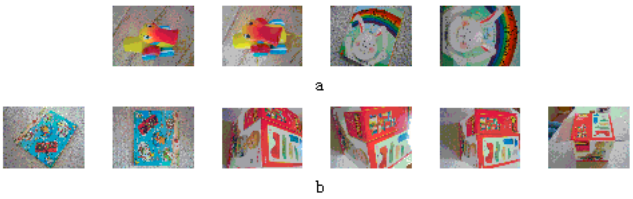
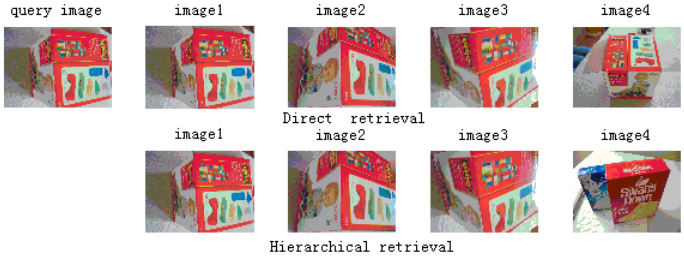


Figure 7. Query images with different N-S value. (**a**) Query images whose N-S value of the hierarchical retrieval is greater than the direct retrieval. (**b**) Query images whose N-S value of the hierarchical retrieval is less than the direct retrieval.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 April 2017 doi:10.20944/preprints201704.0174.v1

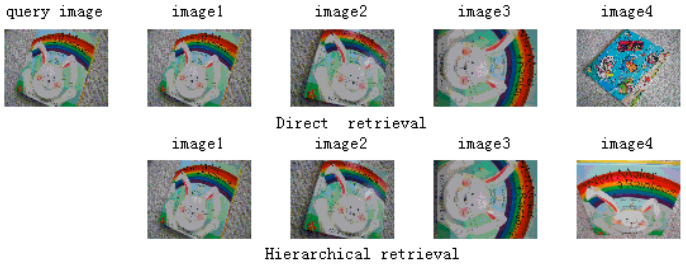


Figure 8. The search results. The first image in the upper left corner is query image, and the remaining images are similar images. In accordance with the similarity from large to small, we arrange them from left to right. The first row is search results based on the direct retrieval. The second row is search results based on the hierarchical retrieval.

3.4. Comparison

Table 2 shows the results of this paper compared with other papers. On the Corel-5K dataset, when the number of returned similar images is 1, the precision of our method is improved by 14% compared with the method in [13]. On the UKbeach dataset, the N-S value of our method is 3.73, which is 0.04 lower than [13], but increased by about 0.18 compared with [17] and [22], 0.31 compared with [20], and 0.08 compared with [19]. On the Holidays dataset, the mAp of image retrieval based on our method is 88.20%, which is about 3.5% higher than [13], [20] and [22]. Compared with [18], [19] and [21], it increased by about 8%. Compared with [17], increased by 8.9%.

Table2. Comparison with other methods with post processing in image retrieval level.

Model	Corel5K(top1)	UKbeach	Holidays
Ours	**68.23**	3.73	**88.20**
Graph-density[13]	54.62	**3.77**	84.64
Global-CNN[17]	- -	3.56	79.3
MOP-CNN+PCA+Whitening[18]	- -	- -	80.18
【19】		3.65	80.2
Bag-8 + PCA[20]	- -	3.43	84.7
【21】	- -	3.60	80.86
【22】	- -	3.55	84.8

4. Discussion

In this paper, we propose a hierarchical search method based on adaptive multi-feature fusion. Given a query image, firstly, we get the pre-retrieval results according to binary CNN feature. Then, we extract the color, lbp, GIST features of query image and images in the pre-retrieval results and search image based one every single feature respectively. After getting the accuracy of each retrieval, we adjust their weights and gain the comprehensive measure. Finally, the retrieval results are obtained according to the comprehensive measure. The proposed method has better efficiency than traditional direct retrieval, and multi-feature fusion can effectively improve the accuracy. In the comparison of the algorithms which have achieved good performance in image retrieval, we find that our method has a good performance and generalization ability for different image databases.

In the future work, we will consider using our method in unsupervised image retrieval work. With the increase of the number of network images, there are many difficulties in manual marking.

The traditional unsupervised image retrieval system has low precision and weak generalization ability. It is significant to consider the application of our method in unsupervised image retrieval.

Acknowledgement：This research is partially supported by National Natural Science Foundation of China (Grant No.31301086).

Author Contributions: X.Lu, Q.Wang, and X.Zhang conceived and designed the experiments; J.Wang and Y.Hou performed the experiments; J.Wang, Y.Hou, and M.Yang analyzed the data; J.Wang and and Y.Hou wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hiremath P S; Pujari J. Content Based Image Retrieval Using Color, Texture and Shape Features. *IEEE Computer Society*. **2007**,780-784. [CrossRef]
2. Tamura B H; Mori S, Yamawaki T. Texture features corresponding to visual perception. IEEE Trans. Syst. Man Cybern. **2010**,460-473. [CrossRef]
3. Lowe D G; Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*. **2004**, 60(2):91-110.[CrossRef]
4. Oliva B A; Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int J Comput Vision*. **2001**.,145–175.[CrossRef]
5. Zheng Y; Huang X; Feng S. An image matching algorithm based on combination of SIFT and the rotation invariant LBP. *Journal of Computer-Aided Design & Computer Graphics*. **2010**, 22(2):286-292. [PubMed]
6. Yu J; Qin Z; Wan T; et al. Feature integration analysis of bag-of-features model for image retrieval. *Neurocomputing*. **2013**, 120(10):355-364.[CrossRef]
7. Wang X; Han T X; Yan S. An HOG-LBP human detector with partial occlusion handling.*Computer Vision, 2009 IEEE 12th International Conference on. IEEE*. **2009**,32 - 39. [CrossRef]
8. Fagin; Ronald; Kumar; et al. Efficient similarity search and classification via rank aggregation. *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*. **2003**, 301-312.[CrossRef]
9. Terrades O R; Valveny E; Tabbone S. Optimal Classifier Fusion in a Non-Bayesian Probabilistic Framework. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. **2008**, 31(9):1630-1644. [CrossRef]
10. Zhang S; Yang M; Cour T; et al. Query specific fusion for image retrieval. *European Conference on Computer Vision. Springer-Verlag*, 2012,660-673. [CrossRef]
11. Jia D; Berg A C; Li F F. Hierarchical semantic indexing for large scale image retrieval. *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. **2011**, 32(14):785-792. [CrossRef]
12. Lin K; Yang H F; Hsiao J H; et al. Deep learning of binary hash codes for fast image retrieval. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. **2015**,27-35. [CrossRef]
13. Zhang S; Yang M; Cour T; et al. Query specific fusion for image retrieval. *Computer Vision–ECCV 2012. Springer Berlin Heidelberg*. **2012**:,660-673. [CrossRef]
14. Nister D; Stewenius H. Scalable recognition with a vocabulary tree. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE*. **2006**, 2: 2161-2168. [CrossRef]
15. Jégou H; Douze M; Schmid C. Hamming embedding and weak geometry consistency for large scale image search-extended version. **2008**, HAL Id : inria-00548651, version 1. [PubMed]
16. Zheng L; Wang S; Tian L; et al. Query-adaptive late fusion for image search and person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. **2015**, 1741-1750. [CrossRef]
17. Babenko A; Slesarev A; Chigorin A; et al. Neural codes for image retrieval. *European Conference on Computer Vision*. **2014**, 584-599. [CrossRef]
18. Gong Y; Wang L; Guo R; et al. Multi-scale orderless pooling of deep convolutional activation features. *European Conference on Computer Vision*. **2014**, 392-407. [CrossRef]
19. Babenko A; Lempitsky V. Aggregating local deep features for image retrieval. *Proceedings of the IEEE International Conference on Computer Vision*. **2015**, 1269-1277. [CrossRef]
20. Perronnin F; Larlus D. Fisher vectors meet neural networks: A hybrid classification architecture. *Proceedings of the IEEE conference on computer vision and pattern recognition*. **2015**, 3743-3752. [CrossRef]
21. Zhang S; Yang M; Wang X; et al. Semantic-aware co-indexing for image retrieval. *Proceedings of the IEEE International Conference on Computer Vision*. **2013**, 1673-1680.[CrossRef]

 doi:10.20944/preprints201704.0174.v1

22. Jégou H; Douze M; Schmid C. Improving bag-of-features for large scale image search. *International Journal of Computer Vision*. **2010**, 87(3): 316-336. [CrossRef]