

Image Retrieval Algorithm Based on Feature Fusion and Bidirectional Image Matching

Kaixuan Ji^{1,a}, Chuan Guo^{2,b}, Shengfu Zou^{3,c}, Yang Gao^{4,d}, and Hongwei Zhao^{5,e}

^{1,2,3,4,5}College of Software Engineering, Jilin University, Changchun 130012, China

^aemail:18844546502@163.com, ^bemail:15567740960@163.com

^cemail:1129930958@qq.com, ^demail:w2028953318@sina.com

^e Corresponding author: zhaohw@jlu.edu.cn

Keyword: Multiple feature fusion, Bidirectional matching, Color histogram, SIFT

Abstract. According to the low accuracy of single feature-based image retrieval algorithm, the paper proposes a strategy which fuses the SIFT and the color feature, to obtain a kind of relatively simple and accurate retrieval approach. To accelerate the retrieval process and improve the efficiency, the paper adopts the image feature-based hierarchical retrieval method. The RGB color feature is the first layers of retrieval feature, presented by the color histogram. SIFT is the second layer of retrieval feature. Meanwhile, for the SIFT retrieval process, by using the bidirectional matching strategy, the improvement of the image retrieval precision is expected. Experiment results show that both the retrieval rate and accuracy are efficiently improved.

1 Background

In recent years, the rapid development of network image information makes new image information ballooning exponentially and image itself also has the advantage of rich content, no language limitation and convenient communication, so it is widely used in real life. [1] As a result, research related to image processing has become a hot spot in the field of computer science. Image processing technology is quite extensive in biological applications and medical fields. It involves image preprocessing, feature extraction, feature matching and many other techniques.

Image feature is generally assorted into global features and local features. Global features involve color, texture, shape and other characteristics. It has strong robustness and is less affected by signal noise. But in a complex image environment, global feature descriptor can't distinguish between foreground and background and the lower mode can't be mapped to the similarity of human perception, which means that while our concerned object is affected by ambient occlusion and other factors, global features are likely to be destroyed. However local features can still stably exist. [2] Furthermore, for the local features, it typically includes SIFT, SURF, and DAISY and so on, besides it is more convincing for the description of a single target and it is much easier to identify the position of a target image. In a word, global features describe the overall characteristic information of a picture and the local features describe the partial characteristic information of the image.

Color feature and SIFT are two kinds of representative features of the image. Most existing algorithms focus on the analysis of a single kind of feature, while ignoring the association between different kinds of features in an image. Therefore, comprehensively considering the complementary and close correlation between color features and SIFT, synthesizing various parts to classify the image could contribute to a better understanding of the semantics of the image. Thus, this paper fuses the global features and local features, for searching a kind of simple and accurate retrieval strategy to establish a stable foundation for further image retrieval. In terms of global features, we choose color feature as the retrieval image feature, since that color correlates closely with the object and the scenes of the image. Compared to other global features color feature rely relatively weakly on the size, direction and visual angle of the picture, which lead

s to high robustness. [4]In terms of feature matching, the paper uses bidirectional matching strategy which adequately reduces the mismatching rate and improves the matching accuracy.[5]

2 Refined Image Fused-Feature Strategy and Bidirectional Matching Algorithm.

2.1 Implementation of Multi-Feature Fusion

In general, people will describe a picture via multiple angles, such as color, shape, texture and more. Therefore, in the Internet world which is full of abundant pictures, single kind of image feature has not been able to express the information of a picture precisely any more. On the contrast, multiple image features could adequately utilize the picture's information, by synthesizing the advantages of each kind of feature. As one kind of image global feature, the color feature is strongly related to the picture's background and scene. Hence, the color feature is steady and visual. However, color feature is not sensitive to the change of the size and direction of the picture area; as a result, color feature could not adequately express the image local characteristic, and also incapably express the information of the spatial distribution of the picture content. Compared to this kind of global feature, SIFT (Scale-invariant feature transform) could be used to express the local characteristic of a picture. Besides, SIFT is scale-invariant and also rotation-invariant, which means that though the angles, brightness or the located angles change, SIFT could still keep well performance on the image-feature extraction. Thus, we decide to fuse SIFT and color feature, to describe a picture from both global and local angle.

For feature fusion, there're mainly two kinds of method: one is hierarchy-based retrieval; the other method is weight assigning. In the hierarchy-based retrieval, the retrieval process is divided into several times (normally 2 or 3 times), one kind of feature is used to retrieve firstly, and then another kind of feature is used to retrieve pictures based on the result picture set of the precedent one. Weight assigning would distribute different weights for each kind of feature, and the distributing strategy also varies under different retrieval picture set. [3] This paper adopts the hierarchy-based strategy (Fig 1). First of all, we use the extracted color histogram to proceed the color feature matching, and obtain a specified number of pictures which corresponds the prior match-degree of color feature. And then, we proceed the SIFT matching on the pictures of precede color feature matching, and finally we get the result which satisfies the demanded match-degree of SIFT.

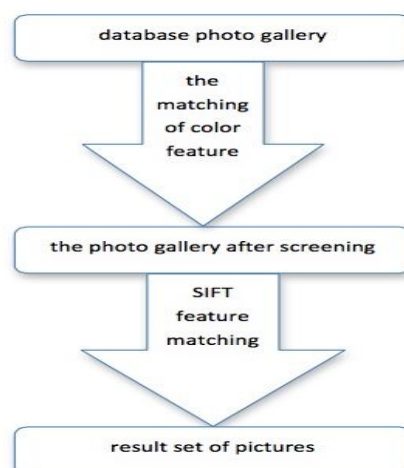


Fig.1. Image Feature-Based Hierarchical Retrieval Method

Because every picture could be extracted out thousands of SIFT points, plus, each point is represented by a 128-dimension vector, the amount of data in the whole database is tremendous. Through screening out part of picture gallery by the color feature, this design could indirectly decrease the amount of matched SIFT points. Furthermore, the retrieval time could be effectively reduced, and the retrieval efficiency could also be improved.

While implementing the design, we extract the color feature firstly. Then, according to the Euclidean distance of color feature, we calculate and sort the color feature similarity. Next, we extract SIFT from the picture, after forming the scale space, detecting the extreme point in the scale space, precisely locating the extreme point, specifying direction, we obtain the coordinates, angles and the corresponding 128-dimension vectors of each point as the feature descriptors. Finally, the descriptors are used to proceed the bidirectional matching, and then we output the result picture set by descending sort of the match-degree of SIFT.

2.2 Refine Image Feature Matching Algorithm

Generally, a characteristic picture could be extracted out thousands of SIFT descriptor. While such a great number of SIFT points are used to match, high accuracy could be hardly warranted. Always, the more SIFT points are involved in the matching process; the more mismatching pairs would appear. Mostly, we choose the Euclidean distance to measure the multiple feature matching-degree. Firstly, we screen out the matching pairs in the feature sets of two pictures, during which we calculate the nearest distance and the second nearest distance of each SIFT point in one picture to all of the SIFT points in another picture. If the rate of the two distance is smaller than the specified threshold value, then we think the SIFT point is matched with the SIFT points correspond to the nearest distance, and these two points are called a matching pair.

According to different requirement of matching accuracy, the set of threshold value is also different. Usually, the smaller the threshold is, the stronger the stability of matching result is, the precision of matching process would be improved as well. But on the other hand, the matching pairs would sharply diminish, which could also negatively influence the accuracy of matching-degree. In this paper, because the source pictures are selected randomly from Internet, and are not related to each other obviously, the value of threshold is 0.9.

However, during the experiment we have found the following two questions. Take the matching process of picture A and picture B for an example. One is that a number of feature point of picture A always successfully match with one or minor feature points of picture B in a small area. The other is that there is a great discrepancy between the amount of matching pairs of “A match B” and “B match A”, even dozens of times different sometimes. This discrepancy makes it hard to decide which one suitably represent the match degree.

To solve the problems, we make some refinements to the experiment process. According to intersection, if the feature point p in picture A has a matched feature point p' in picture B, naturally, the feature point p' should also has a matching point in picture A. Thus, based on the first matching, we detect the matching point in picture A with each matched point in picture B, if the matching point exist, the matching pair is correct, or it's a mismatching. What worth mentioning is that the threshold values could be different in these two matching processes. In this experiment, the threshold values in matching and converse matching are 0.9 and 0.95. The experiment result shows that bidirectional could efficiently get rid of the mismatching pairs.

Meanwhile, in order to settle the discrepancies between the amount of matching pairs of “picture A matches B” and “picture B match A”, we proceed two times of matching: “A match B” and “B match A”, which all adopt bidirectional matching. And the final amount of matching pairs is the weight sum of the result amount of these two times. After many times of testing, while the weight values are 0.5, 0.5, the test performs better.

3 Experiment Result and Analysis

First of all, we collect enough numbers of pictures of Tower, Aircraft, Cruises, Arch Bridge and Ferris wheel these five kinds of pictures, and extract the color features and SIFT points of these pictures. Then we storage all these image features in database. While testing, we will input a picture of Tower, Aircraft, Cruises, Arch Bridge and Ferris Wheel respectively, and will adopt five matching methods to match the pictures in the source picture, which consists of equal amounts of the five kinds of pictures we just mentioned.

Take the table 1 for an example, the picture gallery to be matched involves 200 pictures, among which the numbers of pictures of Tower, Air Craft, Cruises, Arch Bridge, and Ferris wheel are all 40. We output 60 pictures which possess the highest matching degree with the input picture. And then we count the correct matching pictures and calculate the recall ratio of every test. In the same way, table 2 has 300 pictures totally, and consists of 60 pictures of each kind, and output the 90 pictures which possess the highest matching degree with the input picture. At last, we make statistics of the number of correct matching pictures and the recall ratio.

3.1 Improve the Accuracy of Matching Algorithm

Some instructions of relevant symbols are as below:

Image recall ratio is the rate of the successfully retrieved relevant pictures among the total relevant pictures.

$$recall\ ratio = \frac{|\{\text{relevant pictures}\} \cap \{\text{retrieved pictures}\}|}{|\{\text{relevant pictures}\}|} \quad (1)$$

For example, in table 2, there are 29 pictures of Tower in the retrieved picture while the input picture is a tower and the matching method is mode B. We know there are 40 pictures of tower in the source pictures, so the recall ratio is $29/40 = 0.725$.

SIFT unid: the matching method is the traditional unidirectional matching, and the matching degree is the amount of matching pairs.

Mode A: While the matching pictures are picture A, B, symbol A->B expresses that picture A is used to match picture B, and |A->B| expresses the amount of the matching pairs of the A->B. In mode A, the matching degree is:

$$\frac{|A \rightarrow B| + |B \rightarrow A|}{|\{\text{SIFT of A}\}| + |\{\text{SIFT of B}\}|} \quad (2)$$

Mode B: Mode B adopts bidirectional matching strategy. For the matching pairs of A->B, we use the matched SIFT points in picture B to match the SIFT points in picture A, if we could still find the matching pairs, we think this is a correct matching pair. In mode B, the matching degree could be formulated by |A->B->A|.

Mode C: The combination of mode A and mode B. The matching degree is:

$$\frac{|A \rightarrow B \rightarrow A| + |B \rightarrow A \rightarrow B|}{|\{\text{SIFT of A}\}| + |\{\text{SIFT of B}\}|} \quad (3)$$

Mode D: Mode D attempt to screen out 80% of the total source pictures according to the match degree of color feature firstly. And then we use the mode C to match the input picture with the 80% of picture that is screened out before.

Table 1. The Recall Ratios of Testing of 200 Pictures

Input Picture	Recall ratio (%)				
	SIFT Feature Unid	Mode A	Mode B	Mode C	Mode D
Tower	0.5333	0.7167	0.6000	0.6833	0.5667
Aircraft	0.3667	0.4000	0.4333	0.4833	0.2500
Cruises	0.3833	0.4333	0.5333	0.5833	0.5833
Arch bridge	0.4667	0.5000	0.5000	0.5667	0.5833
Ferris wheel	0.3000	0.4167	0.4833	0.4667	0.5167

Table 2. The Recall Ratios of Testing of 300 Pictures

Input Picture	Recall ratio (%)				
	SIFT Feature Unid	Mode A	Mode B	Mode C	Mode D
Tower	0.5750	0.6500	0.7250	0.7500	0.5500
Aircraft	0.3750	0.4500	0.5250	0.5250	0.3000
Cruises	0.3500	0.4000	0.5500	0.5750	0.6000
Arch bridge	0.4750	0.5250	0.6000	0.5250	0.6250
Ferris wheel	0.3000	0.4000	0.4500	0.4000	0.4250

According to table 1, 2, we can see that, compared to SIFT unidirectional matching, mode A, B, C all have improvement on the recall ratio in different degree. Ignoring minority exceptions, the comparison of the degrees of improvement are: C>B>A. Mode D combines color feature and SIFT. While the input picture is Arch Bridge, Cruises, or Ferris, the degree of improvement of mode D is approximately equal to mode C, or even higher in some circumstances. But while the input picture is Aircraft and Tower, the result of mode is some disappointing, almost the same to mode A.

After analyzing the data, we found that the low recall ratio is due to the versatile background style and the various color of the pictures of Tower and Aircraft. While screening the pictures, different types of pictures could be confused. However, compared to SIFT matching, color matching proceeds faster and occupy less memory, thus proper application of mode D could both improve the accuracy and accelerate the retrieval process in some degree.

3.2 Improve the Efficiency of Retrieval Algorithm

For the same input picture, we retrieve the pictures in the source pictures by using fused feature and SIFT feature respectively, and record the proceeding time of the two approaches while under the same running environment and the same source pictures which has 100, 200, 300, 400, 500 pictures step by step. The final output pictures are the 30% pictures of the whole pictures which possess the highest matching degree. The experiment only consider the process of extracting the features of input pictures and feature retrieval, ignoring the process of fetching data from Internet database, to get rid of the uncertain factors of Internet.

Table 3. The Retrieval Time of Image Feature Extraction and Matching

Retrieval methods	Retrieval time (s)				
	100	200	300	400	500
Color feature + SIFT feature	5.2420	7.5149	10.2253	12.9082	15.0730
SIFT feature (100%)	6.9070	10.63989	13.5285	16.5129	19.6348

Table 3 shows that, compared to the traditional SIFT feature retrieval; the approach of feature fusion saves approximately 30% of time. The refined algorithm improves the efficiency of image retrieval on a high degree, and fit the real-time retrieval better.

4. Conclusions

Through analyzing the experimental data, we can see that, in these various matching methods, the combination of bidirectional matching with improved matching metrics improve the recall ratio and the accuracy of image retrieval most. At the same time, compared to the traditional SIFT image retrieval, the combination of bidirectional matching and feature fusion made retrieval time reduced

by nearly thirty percent and generally improve recall ratio substantially. In the real-world, using the bidirectional matching strategy and the image fusion technology could satisfy the needs of real-time retrieval system better.

Acknowledgment

The corresponding author is Hongwei Zhao. The authors are grateful to the anonymous reviewers for their insightful comments which have certainly improved this paper. This work is supported by Plan for Scientific and Technology Development of Jilin Province (20140101184JC).

References

- [1] Kennedy L S, Naaman M. Generating diverse and representative image search results for landmarks[C]//WWW.2008:297-306.
- [2] H.Z.Wang, X.H.He, W.J.Zai. Image Retrieval Based on Combining Local and Global Features [J].Optics and Precision Engineering, Vol.1098-1104, (2008), p. 6
- [3] L.Chen. Image Retrieval Algorithm Research Base on Saliency Analysis and Multi-Feature Fusion[D]. University of Electronic Science and Technology(2013).
- [4] S.H.Guo, M.Yang. Research of Color-Based Image Retrieval [J]. Journal of Southwest University(Natural Science Edition), Vol. 127-133(2012), p. 34(1)
- [5] M.H.Qin. Current Research Situation and Trend of Color-Based Image Feature Retrieval [J]. China Computer and Communication, Vol. 175-175(2010), p. 3