Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 785824, 6 pages
http://dx.doi.org/10.1155/2013/785824

Research Article

A Novel Optimization-Based Approach for Content-Based Image Retrieval

Manyu Xiao,[1] Jianghu Lu,[1,2] and Gongnan Xie[3]

[1] *Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China*
[2] *Computer Science and Technology, Beihang University, Beijing 100191, China*
[3] *Engineering Simulation and Aerospace Computing (ESAC), Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China*

Correspondence should be addressed to Manyu Xiao; manyuxiao@gmail.com

Received 4 October 2013; Accepted 5 December 2013

Academic Editor: Feng Gao

Content-based image retrieval is nowadays one of the possible and promising solutions to manage image databases effectively. However, with the large number of images, there still exists a great discrepancy between the users' expectations (accuracy and efficiency) and the real performance in image retrieval. In this work, new optimization strategies are proposed on vocabulary tree building, retrieval, and matching methods. More precisely, a new clustering strategy combining classification and conventional K-Means method is firstly redefined. Then a new matching technique is built to eliminate the error caused by large-scaled scale-invariant feature transform (SIFT). Additionally, a new unit mechanism is proposed to reduce the cost of indexing time. Finally, the numerical results show that excellent performances are obtained in both accuracy and efficiency based on the proposed improvements for image retrieval.

1. Introduction

Nowadays, content-based image retrieval (CBIR) has more and more applications and constitutes one of the core problems in computer vision. Its features were thoroughly discussed by Smeulders et al. [1]. One of the most popular methods that yield results of content-based image retrieval is based on visual contents of an image. The visual features of images, such as color [2], texture [3], and shape features [4] have been extensively explored to represent and index image contents, resulting in a collection of research prototypes and commercial systems [5, 6]. Therefore, the performance of a CBIR system mainly depends on the particular image representation and similarity matching function employed [7]. Due to the rapid development and improvement of the internet, image capture devices and computer hardware cause the problem of storage and manipulation of images [8]. That is the reason that the relevant techniques developed by Google Inc. and Baidu Inc. did not perform adequately. The main limitation occurs on either retrieval accuracy or real-time or sometimes both. In order to overcome this limitation, in this work a novel optimization-based approach for content-based image retrieval is proposed. The conventional procedure of image retrieval is firstly introduced, as shown in Figure 1. It can be divided into three parts: the vocabulary tree building, the storage of test images, and their retrieval. The descriptions of these three parts are briefly discussed as follows.

1.1. Building Vocabulary Tree. In state-of-art techniques, a tree structure is usually built to store the image database. In other words, a training set is needed to get a discriminative and representative tree. The training set is a group of images, which are first transformed into SIFT [9] (scale-invariant feature transform) 128-dimensional descriptor vectors. SIFT features are distinctive invariant features that are used to robustly describe and match digital image content among different views of a scene. While being invariant to scale and rotation, and robust to other image transforms, the SIFT feature description of an image is typically large and slow to be computed [10] (curse of dimensionality). After that, a vocabulary tree [11] must be built. The traditional classification of the large features of image databases is often carried out by the Hierarchical K-Means method (HKM). Assuming

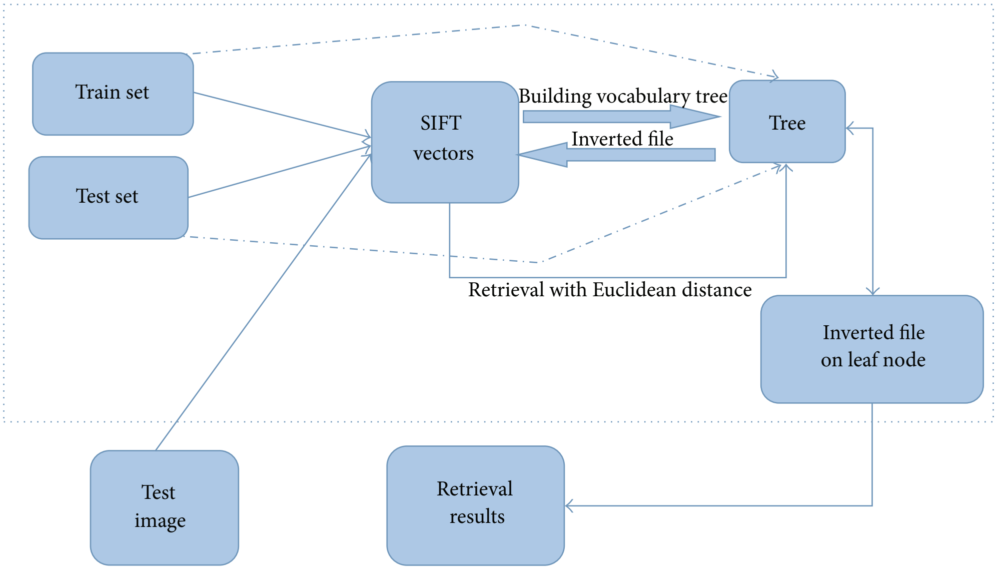

Figure 1: Content-based image retrieval procedure.

a priori N the number of branches and H the height of the tree, the all descriptors can be clustered into N parts, and N cluster centers can be obtained from N nodes at the first level, and then each part eventually reclustered into N new subparts to get all new N^2 nodes. This process is repeated until converged to the defined height and obtention of a complete tree.

1.2. Storage of Test Image and Database Information-Inverted File. Once a well-organized data structure is built, the image database can be stored. From the image database, the SIFT descriptors must also be extracted [12, 13]. For each image in the database, all its descriptors undergo the same classification as follows: comparison with the children nodes of the root node by Euclidean distance and then selection of the nearest one as new root node. This process has to be repeated until the leaf node is reached. In order to match and retrieve the image at the next search, a shortest path to the leaf node has to be constructed. This means an inverted file [11, 14, 15] is needed to be built for identifying the relationship between the image database and the test image. When a descriptor of the test image reaches a leaf node, the relevant inverted-file will record the corresponding information of this descriptor.

After all descriptors of a test image are stored in the vocabulary tree, a weight of a leaf node is calculated based on TF-IDF strategy so as to test the effectiveness of different leaf nodes accurately by the following formula:

$$w_i = \log \frac{N}{\sum_{j=1}^{m} f\left(id_{ij}\right)},$$
$$f\left(id_{ij}\right) = \begin{cases} 1 & id_{ij} > 0, \\ 0 & id_{ij} = 0, \end{cases} \tag{1}$$

where N is the number of images in the database and m is the dimension of the vectors in the inverted-file. If there only exist several images descriptors obtained by $\sum_{i=1}^{m} f(id_{ij})$ and its value is small, this means this leaf node is discriminative. This means the leaf node is discriminative and has a good retrieval feature. Otherwise, this value would be much bigger.

1.3. Retrieval and Rank. After organizing the information of test images and database images, two vectors can be obtained for the test image and the database images as follows.

For the test image,

$$S = \left(w_1 \times n_1, w_2 \times n_2, \ldots, w_n \times n_n\right). \tag{2}$$

For the database images,

$$D_i = \left(w_1 \times l_{i1}, w_2 \times l_{i2}, \ldots, w_n \times l_{in}\right), \tag{3}$$

where w_i is the weight of leaf node and n_i the number of test image nodes reaching the ith leaf node, while l_{ij} is the number of the ith database image nodes found in the jth leaf node during building image base.

Now, the ranking results can be obtained by the following formula:

$$h\left(S, D_i\right) = \left\| \frac{S}{\|S\|} - \frac{D_i}{\|D_i\|} \right\|. \tag{4}$$

$\| \cdot \|$ usually refers to L_1*-norm* or L_2*-norm.*

With the usual approach, accuracy issues often occur. First, as numbers of the test images increase, more noises and clutters are brought into the information database, which undoubtedly results in decreasing the retrieval accuracy. Secondly, more information in the image database leads to more time required to search the similar images from the database, which usually cannot satisfy the real-time demands. Finally, after dozens of trial-and-error tests, it is found that the norms of calculating the match degree cannot remove the magnitude of different image SIFT numbers, which reveals a loss

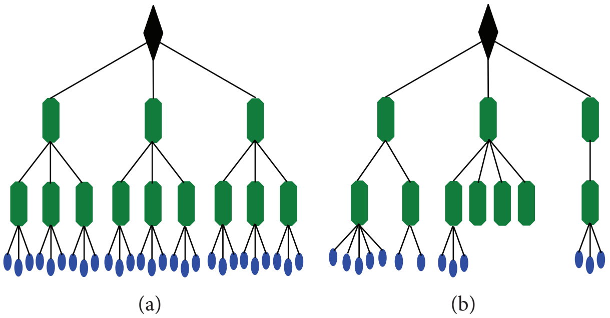

Figure 2: Traditional vocabulary tree (a) and new vocabulary tree (b).

of accuracy. That is what motivates to propose the improvements detailed hereafter.

The paper is organized as follows: three improvements are described in Section 2. Then, the image retrieval application is tested in Section 3, followed by the discussion and conclusion in Section 4.

2. Three Improvements

2.1. Improving the Vocabulary Tree Building. From the process described above, it is known that the height and the branch number of the tree are both predefined, namely, a complete tree (see Figure 2(a)). After building each clustering, all the descriptors will be divided equivalently into several parts. Due to differences of distances, there may be different numbers of descriptors in different parts, and there may even happen that a part only includes a few descriptors. On the other hand, it might also occur that the biggest distance is already small in a certain part, but due to the limit of pre-defined level and branch, this part has to be divided continuously. These are not affordable.

In practical applications, the quantities of information in different test sets are different, and different trees are therefore needed. When the tree need not even be a complete tree, the conventional method certainly leads to some errors. In order to reduce or even eliminate these errors, the conventional K-Means processes and classification are combined to make sure that the height and branch number of vocabulary tree are defined automatically.

The proposed technique called Hierarchical Classification method (HCM) is done with two thresholds: one is for the number of descriptors in a part and the other for the distance inside a part. These two thresholds can determine when the clustering operations terminate; thus we will not know how many levels the tree has and will not know how many children nodes a parent node owns. The structure of two different trees can be shown as follows, respectively (Figure 2(b) improved tree): this new model provides not only improvements of efficiency but also in precision.

2.2. Shorter Time Spent on Retrieval. In previous works, a classification was often obtained by Euclidean distance of the children nodes of the root node, not the information of root node directly in the left of Figure 3. Obviously, it took much time for calculation. This distance was not bigger than the sum of the distances between descriptor-root node and the distance obtained from root node-relevant child node. It can be written as follows:

$$\|P-Q\| \leq \|P-R\| + \|R-Q\|, \tag{5}$$

where P denotes a descriptor, Q is one child node of root node, and R is root node.

As the distances between root node and its children nodes are all calculated in advance and reserved in the root node position, the proposed clustering technique consists in finding the next children tree by using only the first term of (5). The second part is used for next clustering rank shown in Figure 3(b).

Define M as the height of vocabulary tree, and K is the number of child nodes at each rank and T is computational time for each clustering. The total time with the traditional classification to find the nearest leaf node is about $(M-1)\times K\times T$, while the improved method only takes a time of $M\times T$; the later consumes about only $M/((M-1)\times K)$ of time needed before, which significantly reduces the retrieval time.

2.3. Improvement on Scoring Mechanism. In the traditional method, $\|\cdot\|$ usually is L_1*-norm* or L_2*-norm*. However, for lots of practical experiments, it is found that the results are not very reasonable because of the different memory requests. When the image is more complex, it is more possible to match with retrieval image as the large number of features (large number of eigenvectors of dimension 128). On the contrary, it is of poor effect for simple images. The third improvement will concentrate on dealing with this problem. The following *unit norm* is proposed to eliminate the weight of test image and great different database image memory:

$$h(S, D_i) = \left\| \frac{S}{N} - \frac{D_i}{N_i} \right\|, \tag{6}$$

where N stands for the number of test image vectors and N_i the number of the ith database image. This *unit norm* can make sure that different images stay at the same level; a more reasonable score will hence be produced. This new *unit norm* is called *unit norm*.

Important notation: when programming, there may be millions of pictures in the image resource, and there will be

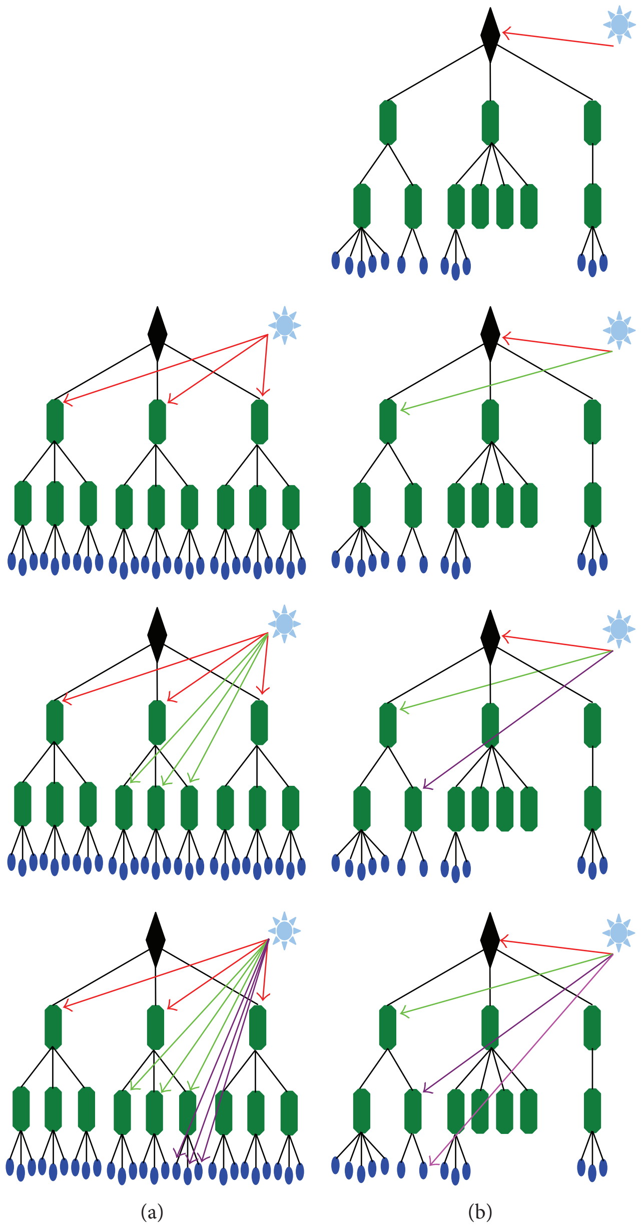

Figure 3: Traditional classification (a) and new clustering technique (b).

even more than 106 leaf nodes, while for indexing each image in the database, there will be thousands of dimensions equal to 0 in vectors. In order to save the memory space, assigning storage dynamically is proposed.

3. Test Examples

Ukbench image database contains 2550 groups of images, and every group includes 4 similar images. More precisely, these 4 images with a much similar characters are snapshots in a same image but in different illumination intensity and orientation. Analyzing the expectations of users, the following strategies are taken: indexing one image from the database images, if three of four similar images can be exhibited in the ranking 10 results, this image retrieval implementation is a successful process. The index frequency is calculated by

$$f_i = \frac{n_i}{4}, \quad n_i = 0, 1, 2, 3, 4, \tag{7}$$

where n_i is the number of similar images shown in ranking 10 images and f_i an index accuracy for ith test image.

The average of this accuracy is finally used to test the effectiveness of different *unit norm*. Different quantities of images retrieval are shown in Figure 4.

From Figure 4, it is observed that a more reasonable score h will be obtained with *unit norm* (that means matching degree is bigger between similar images and smaller between irrelevant ones). The index accuracy is thence much better with *unit norm* than with L_1-*norm* and L_2-*norm*.

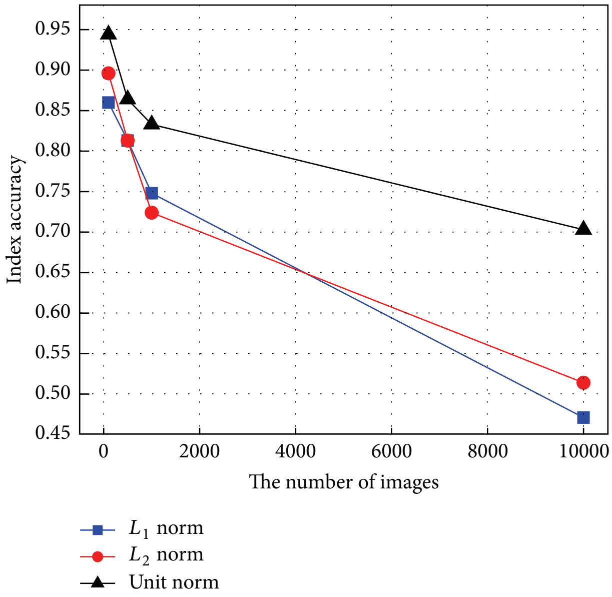

Figure 4: Comparison of three different *norms*.

Table 1: Improvements on discarding invalid descriptors.

Discard	Quantity			
	100	500	1000	10000
No	90.6%	82.3%	76.2%	53.9%
Yes	94.3%	86.3%	83.25%	70.2%

Table 2: Comprehensive comparison of new and traditional mechanism.

Method	Quantity			
	100	500	1000	10000
HKM	89.5%	84.4%	80.25%	62.3%
HCM	94.3%	86.3%	83.25%	70.2%

In building test image base, taking the strategy of discarding into action, much better performance is achieved as listed in Table 1.

Based on these two important improvements, the efficiencies of classification with the famous HKM method and HCM (the proposed method) are compared. The results on Table 2 show that the proposed improvements are more feasible and efficient.

4. Conclusions and Discussion

In this work, three improvements are proposed during the content-based image retrieval: strategy in image classification, mechanism to calculate the Euclidean distance of eigenvectors between source images and research image, and development of the inverse file. As a result, the index accuracy can be greatly enhanced. Furthermore, we can get a faster index procedure, which satisfies the real-time image retrieval quite well. In the point of theoretical view, the proposed technique takes only about one-sixth time of traditional method needed. In the future work, it is necessary to verify the efficiency of proposed improvement in a practical situation as the time is really very short in the above example.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work has been supported by the Basic Project Foundation of Northwestern Polytechnical University (Grants no. JC20120241) and by the National Natural Science Foundation of China (Grants no. 11302173).

References

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[2] R. O. Stehling, M. A. Nascimento, and A. X. Falcao, "On shapes of colors for content-based image retrieval," in *Proceedings of the ACM International Workshop on Multimedia Information Retrieval (ACM MIR '00)*, pp. 171–174, 2000.

[3] M. Flickner, H. Sawhney, W. Niblack et al., "Query by image and video content: the QBIC system," *Computer*, vol. 28, no. 9, pp. 23–32, 1995.

[4] D. S. Zhang and G. Lu, "Generic Fourier descriptors shape-based image retrieval," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '02)*, vol. 1, pp. 425–428, 2002.

[5] F. Jing, M. Li, H. Zhang, and B. Zhang, "An effective region-based image retrieval framework," in *Proceedings of the 10th ACM International Conference on Multimedia*, pp. 456–465, December 2002.

[6] A. Gupta and R. Jain, "Visual information retrieval," *Communications of the ACM*, vol. 40, no. 5, pp. 71–79, 1997.

[7] H. B. Kekre, S. D. Thepade, A. Athawale, A. Shah, P. Verlekar, and S. Shirke, "Energy compaction and image splitting for image retrieval using kekre transform over row and column feature vectors," *International Journal of Computer Science and Network Security*, vol. 10, no. 1, 2010.

[8] M. K. Mandal, F. Idris, and S. Panchanathan, "Critical evaluation of image and video indexing techniques in the compressed domain," *Image and Vision Computing*, vol. 17, no. 7, pp. 513–529, 1999.

[9] D. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[10] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 2161–2168, 2006.

[11] P. Indyk and R. Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," in *Proceedings of the 30th Annual ACM Symposium on Theory of Computing (STOC '98)*, pp. 604–613, 1998.

[12] J. Sivic and A. Zisserman, "Video google: a text retrieval approach to object matching in videos," in *Proceedings of the 9th*

IEEE International Conference on Computer Vision (ICCV '03), pp. 1470–1477, Nice, France, October 2003.

[13] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pp. 1000–1006, June 1997.

[14] J. Zobel, A. Moffat, and K. Ramamohanarao, "Inverted files versus signature files for text indexing," *ACM Transactions on Database Systems*, vol. 23, no. 4, pp. 453–490, 1998.

[15] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 26–33, June 2005.