ICSP2006 Proceedings

Multilevel Indexing Structure for Object Based Image Retrieval

Shikui Wei, Yao Zhao, Zhenfeng Zhu
Institute of Information Science, Beijing Jiaotong University, Beijing, 100044, China
E-mail: shkwei@gmail.com

Abstract

Object based technique is a good way to grasp the user's query intentions felicitously and at the same time to make retrieval procedure flexible with user to avoid pseudo retrieval in content-based image retrieval (CBIR). But the main problem for such technique is its high computation complexity for searching objects among image dataset. Aiming at the problem mentioned above, a multilevel indexing structure (MIS) for object-based image retrieval is proposed in this paper. A three-level tree structure in the image domain is first constructed to reduce the object searching complexity. In order to further fasten the speed of scanning objects, i.e., to eliminate those unrelated images in the dataset rapidly, the clustering in the dataset domain is also conducted. The final experimental results show that the proposed MIS for image retrieval is much more efficient.

Keywords

CBIR, MIS, ISODATA, Object Searching

1. Introduction

With the popularity of computers and digital cameras, the digital images are cropping up everywhere now. Consequently, the retrieval techniques based on the image content are demanded in many application areas.

For traditional CBIR, the user's query need is generally modeled by a globally extracted image representation rather than object representation. Thus there usually exist a large number of pseudo results. In order to grasp the user's information need felicitously and at the same time to avoid pseudo retrieval in CBIR, the region-based retrieval technique has attracted much attention of researchers for CBIR. However, most region-based retrieval methods [1, 2] strongly rely on the results of image segmentation which is still an open problem in computer vision community. For the calculation of similarity between two objects, many systems [1,2] rank the retrieval results under area constraint, i.e. the big regions are more important than small regions, The assumption of area constraint is fairly feasible for the images in which the object covers a big patch of canvas but usually fails when the object region is small.

In order to overcome these limitations, the windowed object matching approach, which is commonly applied in the area of object detection, was introduced into CBIR [4,6,7,10,11]. The system reported by D. Horem et al [7] performed a windowed search over locations and scales for each image in the database, where a Bayesian approach is proposed to classify those object candidates as positive or negative. In [6], K.Vu et al. have made an effort on the matching of regions of interest with various scales, called "SamMatch". WALRUS [4] also decomposed an image into multi-regions by using sliding windows of various sizes, and conducts a clustering among them based on the proximity of their features.

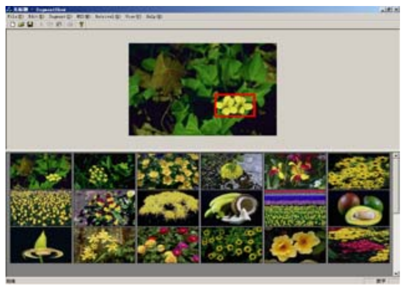

Fig.1. Retrieval results of a yellow flower in a database of 2000 images drawn from Corel set

Although the object retrieval method, based on window searching, partly solves the matching problem of different scales and avoids the tough segmentation of image. But the high computation cost for object searching is its main impediment. The focus of this paper is to introduce a novel multilevel indexing structure for object-based image retrieval that can speed up searching of objects in both the image domain and dataset domain. The object-based retrieval system developed by us is shown in Fig.1, in which the object outlined by the hand–drawn red rectangle box implicates the user's retrieval intention. For simplicity of the following illumination, we term the user indicated object in the query image as query object, and the objects corresponding to sub-images picked up from dataset images as candidate objects.

2. Object representation and similarity measurement

In CBIR, some traditionally utilized low-level visual features include color, texture and shape, etc. In order to construct an efficient object representation, the scalable color descriptor (SCD) proposed by the MPEG-7 Final Committee Draft (FCD) is adopted. SCD is a color histogram in HSV color space, which is encoded by a Haar transform. Its binary representation is scalable in terms of bin numbers. The flowchart of building SCD is illustrated in Fig. 2.

In our case, three subsets of the final coefficients are extracted, which are equivalent to the top 64,128, and 256 coefficients of Haar transform, respectively. A multi-resolution representation of object, finest, coarse and coarsest subset, is then obtained.

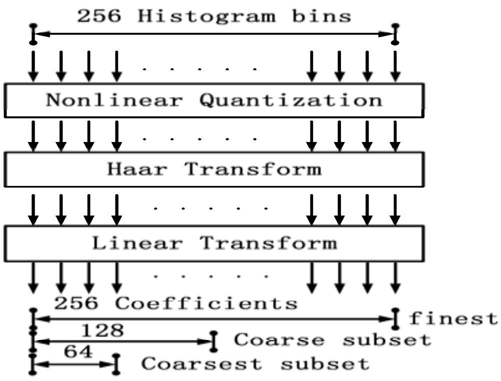

Fig.2. Flow chart of SCD extraction

The core experiments in MPEG-7 have proven that a good retrieval result can be achieved by using the SCD and L_1 norm measurement [8]. Assume that signatures of object i and j are represented by vectors $V_i = \{e_{i1}, e_{i2}, \cdots e_{in}\}$ and $V_j = \{e_{j1}, e_{j2}, \cdots e_{jn}\}$ respectively, where n denotes the top n coefficients of Haar transform, thus the similarity between object i and j can be defined as:

$$d_n(V_i, V_j) = \left\|V_i - V_j\right\| = \sum_{k=1}^{n} \left|e_{ik} - e_{jk}\right| \qquad (1)$$

where n =64,128,256, denoting the number of Haar coefficients in three representation subset respectively.

3. Multilevel index structure

In object-based image retrieval, the retrieval procedure can be performed in two domains, i.e. dataset domain corresponding to scanning objects across images in the dataset, and image domain corresponding to object searching among all candidate objects in one image. Thus, in order to improve the retrieval efficiency, a multilevel indexing structure (MIS) should be constructed in both image and dataset domains.

3.1. Constructing index structure in image domain

As a large number of objects are generated by window scanning over varied locations and scales in a single image, it is time consuming to search directly similar objects in the image. So it is necessary to exploit some strategies to reduce the search cost.

As mentioned in section 2, we can obtain multi-resolution representations for each candidate object embedded in a single image under the framework of SCD. Based on such multi-resolution representation, a tree-like indexing structure in image domain can be constructed, and the constructing process of indexing structure is shown in Fig.3. In each level, corresponding to resolution of object representation, the clustering is conducted. In this way, all objects within a single image are grouped grossly into several clusters (Nodes) by using the coarsest representation (the most prominent features) of object. An indexing tree is then constructed by clustering objects of parent note using finer representation of object. Here an improved ISODATA algorithm proposed by us in [3], which can update merging parameter θ_c and splitting threshold θ_s dynamically for each circle of iteration, is applied to such clustering. The number of clusters in a single image depends mainly on the image content and initial value of parameter set beforehand. The initial value, here, is set to be 4. The final number of clusters then varies in 2~8.

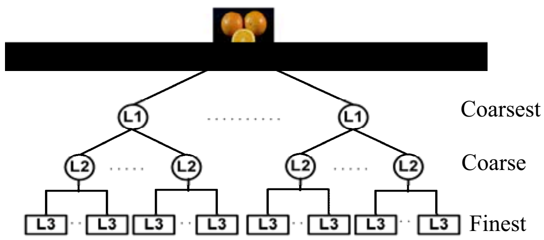

Fig.3. Constructing of index structure in image domain

For L_1 and L_2 node levels, each node stores object ID and the clustering center corresponding to the cluster, as well as the pointer to the next level nodes and location of image in the database. As a matter of fact, L_3 node level only contains finest features of objects. Endeavoring to avoid arising of false alarm, the final determination that whether the candidate objects is really similar with the query object is indeed given in the third node level, i.e., the finest L_3 node level.

3.2. Constructing structure in dataset domain

For a large size of image dataset, however, it's not enough to only narrow search range in the image domain. Hence, it's necessary to eliminate those unrelated images in the database before object searching within a single image is triggered. So a further clustering for L_1 nodes in the dataset domain should be conducted by using the improved ISODATA algorithm with 40 as the initial value, and accordingly a complete multilevel indexing structure is built and illustrated in Fig.4. As shown in Fig.4, the structure of L1 node and its child trees is kept, but L1 nodes within the same image may be assigned to different C nodes after clustering in the whole object candidates of dataset. Each C-node includes the center of the corresponding cluster and pointers to the L_1-nodes.

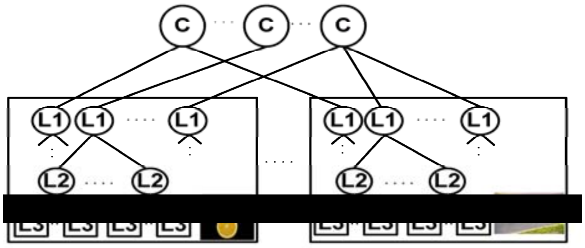

Fig.4. Complete multilevel index structure (MIS)

Given the specified query object by user in the query image, three subsets of the signatures are first extracted. Then a threshold ε_1 selected experientially in C-node level will be set to establish the determination whether the further searching is implemented in its child trees. If the distance between the query object and one C-node exceeds ε_1, then this C-node will be eliminated without taking any further measure, i.e., all sub-nodes associated with this C-node can be pruned. Otherwise all its child trees are searched. However, the searching process of child trees is carried out in the same manner, and the thresholds in each level of child tree are also given experientially according to resolution of representation.

Generally speaking, the candidate objects in dataset domain are usually diverse, thus with the introduction of the indexing structure in dataset domain more unrelated candidate objects can be eliminated from the candidate set without need to give further verification. So it is not hard to understand that the retrieval efficiency can be improved significantly.

4. Ranking based on similarity

As mentioned above, the task of object based image retrieval is to find all the images that contain similar semantic objects named by the user, thus the matching between images can be converted into similarity measure among objects with sub-image representation. Let's denote $S(Q,D)$ as the matching score between the query image Q and the database image D, and we have

$$S(Q,D) = \frac{1}{\min_j \{d(V_{Q,q}, V_{D,j})\}} \tag{2}$$

where $V_{Q,q}$ denotes the feature vector of the query object, $V_{D,j}$ denotes the feature vector of the j^{th} candidate object contained. When one query session was initiated, the top n ranked images based on $S(Q,D)$ will be returned.

5. Experiments and analysis

For evaluating the performance of our indexing algorithm, four image datasets drawn from Corel image database are tested. For simplicity, we name the four datasets by Lib1, Lib2, Lib3, and Lib4. For Lib4, it contains 20 objects categories, and each category includes 100 images. Lib1, Lib2 and Lib3 are the subsets of Lib4, which include 250, 500, 1000 images, respectively. All these images are stored in JPEG format, with size either 256×384 or 384×256.

In order to generate the candidate objects for matching with query object, a size-fixed 64×64 pixels sliding window at a spacing of 16×16 pixels is applied to all images in the image dataset. In addition, for consideration of the scale variation of object, down sampling on the original images is put up with 1/4 sampling ratio. Thus, a total of 460 candidate objects can be generated for each original image. The evaluation for an object based image retrieval system takes on mainly in two aspects of retrieval performance and efficiency. To show the retrieval efficiency of the proposed MIS for object based image retrieval, the comparison of MIS with NI and NII is developed. Here NI indicates that no any indexing structure is involved in carrying out the matching between the query object and candidate objects, i.e. the matching between them is conducted sequentially; and NII only constructs indexing structure in image level.

The tests are carried out on a Pentium(R) 4 CPU 2.66GHZ with 256MB of RAM. The average response time of NI, NII and MIS on the four testing datasets is shown in Table 1. As shown in Fig.5, although NII improves the efficiency significantly, the ratio of NI/NII keeps approximately invariant with the increasing size of dataset. In contrast to the ratio of NI/NII, NI/MIS keeps approximately linearly increasement relation with the size of dataset, which indicates that the proposed MIS shows its advantage greatly over NI and NII for image retrieval with large size of dataset. Note that for each dataset the average respond time in Table 1 is approximately over 75 query sessions.

Table 1 average response time of NI, NII and MIS

	Lib1	Lib2	Lib3	Lib4
NI(s)	19. 93	44. 87	93. 12	234.23
NII(s)	4.51	11.39	22.60	55.37
MIS(s)	0. 689	1. 047	1. 252	1. 855

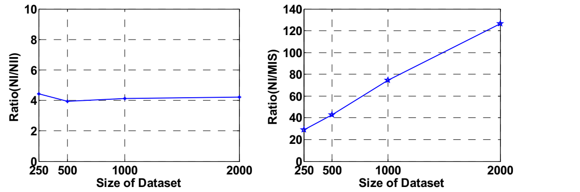

Fig.5. Cost ratio of NI/MIS and NI/NII with different size of database

Finally, the retrieval performance of the proposed MIS for object-based image retrieval is then tested on the dataset Lib4. Firstly seven categories of images, including Lion, Horse, Flower, Orange, Car, Leopard, and Racing car, and further ten images for each category are randomly selected to form the query dataset. Here the retrieval precision with different scopes is considered as the performance measure and the average retrieval precision for each of seven categories of images is illustrated in Fig.6.

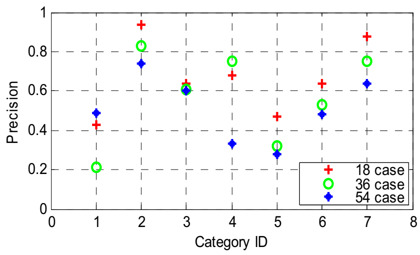

Fig.6. Average retrieval precision with different scopes

5. Conclusions

In order to grasp the user's information needs felicitously and at the same time to make retrieval procedure flexible with user to avoid pseudo retrieval in CBIR, object based technique is a good way. To overcome the problem of its high computation complexity for object searching among image dataset, a multilevel indexing structure (MIS) for object-based image retrieval is proposed in this paper. Under the framework of SCD, a three level tree-like structure in the image domain is first constructed to reduce the object searching complexity. In order to further fasten the object scanning i.e. to eliminate those unrelated images in the dataset domain rapidly, the clustering in the dataset domain is also conducted. The final experimental results show that the proposed MIS for image retrieval is much efficient.

Acknowledgements

This work was supported in part by NSFC(No. 90604032, No. 60373028), SRFDP， NCET and Specialized Research Foundation of BJTU.

Reference

[1] J.Z. Wang, J. Li, and G. Wiederhold, "SMPLICITY: semantics-sensitive integrated matching for picture libraries," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 9, pp. 947-963, Sep. 2001.

[2] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 8, pp. 1026-1038, Aug. 2002.

[3] Shikui Wei, Yao Zhao， Zhenfeng Zhu, "Meaningful Regions Segmentation in CBIR", Proc. IEEE Int'l WorkShop VLSI Design and Video Technology, pp.253-256, May 28-30 2005

[4] A. Natsev, R. Rastogi, K. Shim, "WALRUS: A similarity retrieval algorithm for image databases", IEEE Trans. Knowledge and Data Engineering, vol. 16, no. 3, pp. 301-316, March 2004.

[5] Z. Yujin, G. Yongying, and L. Yun, "Object-Based Techniques for Image Retrieval", In: Multimedia Systems and Content-based Image Retrieval, Idea Group, Inc, Chapter7 (156-181), 2004

[6] K. Vu, K.A. Hua,, N. Jiang , "Improving image retrieval effectiveness in query-by-example environment", Proc. ACM Symposium Applied computing , pp. 774 – 781, Mar. 2003

[7] D. Hoiem, H. Sukthankar, H. Schneiderman, and Huston, L, "Object-based image retrieval using the statistical structure of images", Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, vol. 2, pp 490- 497, 27 June-2 July 2004.

[8] B.S. Manjunath, J.-R Ohm, V.V. Vasudevan, "Color and texture descriptors", IEEE Trans. Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 703-715, June 2001.

[9] A.D. Bimbo, P. Pala, " Shape indexing by multi-scale representation", J. Image and Vision Computing, vol. 17, no. 3-4, pp. 245-261, Mar 19991

[10] A. Kimura, T. Kawanishi, and K. Kashino, "Acceleration of similarity-based partial image retrieval using multistage vector quantization", Proc. IEEE Int'l Conf. Pattern Recognition, vol. 2, pp. 993-996, Aug 23-26 2004.

[11] Y-S. Moon, K-Y Whang, and W-S Han, "General Match: A subsequence matching method in time-series databases based on generalized windows", Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 382-393, June 3-6, 2002.

[12] Text of ISO/IEC 15 938-3 Multimedia Content Description Interface—Part 3: Visual. Final Committee Draft, ISO/IEC/JTC1/SC29/ WG11, Doc. N4062, Mar. 2001.