2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

Efficient and User-Friendly Image Retrieval System Based on Query by Visual Keys

M. Serata, K. Sakuma, Z. Stejic, K. Kawamoto, H. Nobuhara, S. Yoshida, K. Hirota
Department of Computational Intelligence & Systems Science, Tokyo Institute of Technology
4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan
E-mail: {sera, sakuma, stejic, kawa, nobuhara, shin, hirota}@hrt.dis.titech.ac.jp

Abstract – **A new query method, called query by visual keys, is proposed to aim easy operation and efficient region-based image retrieval (RBIR). Visual keys are constructed from representative regions/subimages in a given image database, and the database is indexed with visual keys. A system on PC is presented, where text retrieval techniques are applied to the image retrieval with visual keys. Experimental results show that one retrieval is done within 4ms and that the proposed system achieves the comparable retrieval precision (with user-friendly operation and low computational cost) to conventional region based image retrieval systems**

I. INTRODUCTION

Region-Based Image Retrieval (RBIR), subclass of Content-Based Image Retrieval (CBIR), aims to improve the retrieval performance by decomposing each image into a set of regions/subimages and evaluating local similarity. The typical RBIR systems, e.g., SIMPLIcity [1] and Blobworld [2], achieve high retrieval performance by using detailed queries and computing complicated similarity. But a little skill is requested of users to indicate appropriate queries, and a high performance computer is necessary to realize a fast retrieval system for a huge scale image database and huge number of users. A new user-friendly query method and a simple image matching algorithm should be developed for RBIR to realize an intelligent everybody, everytime, everywhere available image retrieval system.

The concept of visual keys is proposed to provide easy retrieval operation for users and easy computation for retrieval engine. Visual keys are a set of subimages picked up from image database in advance, and are used as the hint/query given by users to the system in order to retrieve desired images. Visual keys used in the proposed image retrieval system play a similar role to keywords widely utilized in text retrieval systems. Those who are not familiar with digital images are also able to access the proposed image retrieval system easily by indicating visual keys. On the other hand, the indexing of image database using visual keys reduces the retrieval computational cost. Experiments are done with a subset of the COREL database on a Pentium 4 (1.6GHz) PC to confirm that the proposed system achieves the comparable retrieval precision with user-friendly operation and low computational cost to conventional region based image retrieval systems.

The concept of visual keys is proposed followed by how to construct the visual keys and how to assign indexes to visual keys in II. Section III presents the proposed image retrieval system based on visual keys. Experimental results with comparison evaluation are shown in IV.

II. VISUAL KEYS AND INDEXING

A. Concept of Visual Keys

In most of the conventional RBIR systems, e.g., SIMPLIcity [1] and Blobworld [2], the user has to choose an image or a collection of subimages as a query. In a huge scale image database, it is not easy for users to choose the query from all images or all subimages in the database.

A method is proposed to reduce the burden for the users by constructing a set of representative subimages, called *visual keys*, in advance from the given image database. To retrieve images, the users are supposed to choose the query from a set of visual keys.

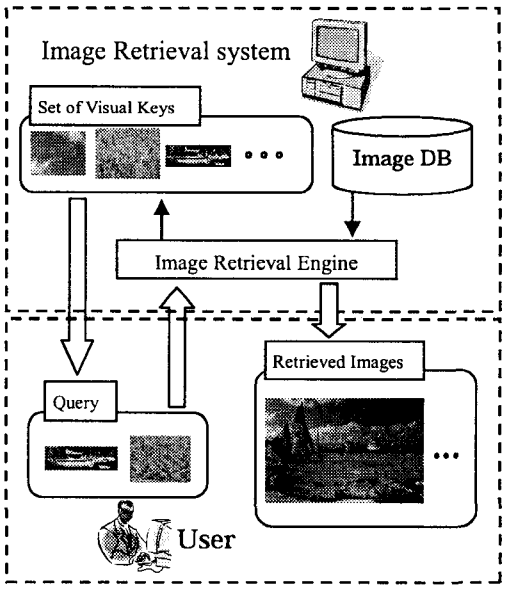

Fig.1: Image retrieval with visual keys

B. Visual Keys Construction

An algorithm how to construct visual keys is presented in the followings.

i) Extracting subimages;

All subimages are extracted from all images in the image database by applying the segmentation method used in the Blobworld system [2] (http://elib.cs.berkeley.edu/photos/blobworld/).

ii) Extracting features from the subimages;

A feature vector of each subimage is defined as a 25-D vector, whose components are:

(a) Color: Color moments

The subimage is represented in HSV space. The histogram of the subimage in each of Hue, Saturation, and Value is calculated. Three color moments, i.e., mean, standard deviation, and skewness, for each of the three histograms are normalized in [0,1], and are put as a part of 3x3=9 dimensional feature subvector.

(b) Texture: Texture neighborhood

The subimage is represented as a gray scale image. An eight dimensional feature subvector is generated for the gray scale image as follows. If the value of an indicated pixel is lower than that of each neighboring pixel, then the directional counter prepared is incremented. This is done for all inner pixels of the subimage. The values of the directional counter are divided by the number of inner pixels (of the subimage). The normalized eight values of the directional counter constitute an eight dimensional feature subvector (of the subimage).

(c) Shape: Sobel edge detector

The subimage is represented in HSI space, and only S and I planes are used. The directional counter is initialized. Sobel operators of eight directions are applied, and the maximum value and its direction are memorized for each inner pixel of S and I planes. The maximum values (of the same inner pixels of S and I planes) are compared and the bigger value related direction is counted by incrementing the directional counter, where if the maximum value is less than 35%/15% of the greatest maximum values in calculated S/I planes, then the counting operation is omitted. The counted eight values of the directional counter are normalized such that the whole sum is equal to 1, and they are assigned as an eight dimensional feature subvector of the given subimage.

iii) Clustering;

Subimages are classified by applying a hierarchical clustering method to 25-D feature space. The difference between clusters C_i and C_j measured by

$$d(C_i,C_j)=\frac{|C_i||C_j|}{|C_i|+|C_j|}\|v_i-v_j\|, \quad (1)$$

where $|C_i|$ and v_i are the number of feature vectors in C_i and the centroid of C_i, respectively.

The number of clusters is the same as that of visual keys and is given through the optimization process of the image retrieval (cf. IV.C).

iv) Selecting visual keys;

For each cluster, a visual key is obtained by assigning the subimage of the feature vector closest to the centroid of the cluster. The set of all visual keys is denoted by $\mathbf{VK}=\{vk_i\}$.

C. Indexing

If visual keys are compared to keywords widely used in text retrieval, the target image database is naturally indexed by applying the inverted file indexing with TFIDF (term frequency - inverse document frequency) [3] to the visual keys that are extracted from the database in advance. In the same way as text retrieval, it is expected that the inverted file index is able to perform retrieval in short time, and that the TFIDF indicates the significance measure of each visual key.

An index of an image in the target image database is constructed as follows.

i) Extracting subimages;

A subimages set $\mathbf{S}(I_j)$ is extracted from each image I_j in the target image database **DB** by segmentation, in the same way as mentioned in II.B, where the subimages in $\mathbf{S}(I_j)$ are supposed to be mutually disjoint.

ii) Representation by visual keys;

The obtained subimages set $\mathbf{S}(I_j)=\{s_{jk}\}$ from the image I_j (in the target image database) is represented by the set of corresponding neighboring visual keys,

$$\mathbf{VK}(I_j)=\left\{vk_{k\min}\middle|vk_{k\min}=\arg\min_{vk_i\in\mathbf{VK}} d(s_{jk},vk_i), s_{jk}\in\mathbf{S}(I_j)\right\}, \quad (2)$$

where $d(s_{jk}, vk_i)$ indicates the Euclidean distance between 25-D feature vectors of s_{jk} and vk_i.

iii) Constructing indexes;

For each visual key vk_i, the index is constructed as

$$index(vk_i)=\{(I_j,d(s_{jk},vk_i))|I_j\in\mathbf{DB}, vk_i\in\mathbf{VK}(I_j), \quad (3)$$
$$s_{jk}: \text{corresponding subimage of } vk_i \ (\text{in } I_j)\}.$$

The frequency of appearance for each visual key in the image database may be changed from one to another. To characterize the importance of each visual key by frequency of appearance, a text retrieval related technology [3], called TFIDF, is used. The TF (Term Frequency) and IDF (Inverse Document Frequency) are used as local weight and global weight for each visual key, respectively.

They are defined as:

$$TF_{ij}=\begin{cases}0.5+0.5\dfrac{f_{ij}}{\max_k f_{kj}} & (f_{ij}>0)\\ 0 & (f_{ij}=0)\end{cases}, \quad (4)$$

$$IDF_i=\frac{1}{n_i}, \quad (5)$$

where f_{ij} is the percentile of the subimage corresponds to the visual key vk_i to an image I_j, and n_i (>0) is the number of the images (in the image database) that have the visual key vk_i.

III. Image Retrieval System with Visual Keys

A. Query represented by Visual Keys

In the proposed image retrieval system, a user is supposed to indicate several visual keys from a given set **VK** of visual keys as a query $\mathbf{q}=\{vk\}$, where each vk is treated as the hint by the system to retrieve the desired image of the user.

B. *Query Matching*

For a given query $\mathbf{q}=\{vk\}$, the proposed system selects the matched images $\mathbf{M}(\mathbf{q})$ as

$$\mathbf{M}(\mathbf{q})=\left\{I_j \middle| (I_j,d)\in index(vk), vk\in\mathbf{q}\right\}. \quad (6)$$

The matched images in $\mathbf{M}(\mathbf{q})$ are ranked according to the quality index given as

$$Q(I_j)=\sum_{(I_j,d)\in\bigcup_{vk_i\in\mathbf{q}} index(vk_i)} w_i TF_{ij} IDF_i (1-d), \quad \text{for } {}^{\forall}I_j\in\mathbf{M}(\mathbf{q}) \quad ,(7)$$

$$w_i=\begin{cases}1 & (vk_i\in\mathbf{q})\\ 0 & (vk_i\notin\mathbf{q})\end{cases}$$

where w_i is the weight for each visual key to apply relevance feedback in III.C and is initialized to 1 or 0. Then the higher quality index means the better candidate for the user. All images in $\mathbf{M}(\mathbf{q})$ are sorted from the higher quality index to the lower one. The obtained ordered image set is written by the same symbol as $\mathbf{M}(\mathbf{q})$ hereafter.

Most of the users may be satisfied with the presented result $\mathbf{M}(\mathbf{q})$. For those users, however, who do not satisfy, a method of modifying $\mathbf{M}(\mathbf{q})$ is prepared by the system based on the relevance feedback technology mentioned in III.C.

C *Relevance Feedback*

Relevance feedback used in text retrieval systems [4] is applied to the proposed system. The user is supposed to choose a desired image set $\mathbf{DM}(\mathbf{q})$ $(\subset\mathbf{M}(\mathbf{q}))$ from the ordered image set $\mathbf{M}(\mathbf{q})$ presented by the system. Then the system modifies the given query $\mathbf{q}=\{vk\}$ by the user and weights w_i's by applying the following algorithm to the desired images $\mathbf{DM}(\mathbf{q})$.

i) The query $\mathbf{q}$ is modified to a query $\mathbf{q}'$ as

$$\mathbf{q}'=\left\{vk \middle| vk\in\mathbf{q} \text{ or } \left(vk\in\mathbf{VK}(I_j), I_j\in\mathbf{DM}(\mathbf{q})\right)\right\}. \quad (8)$$

ii) The weight w_i of the visual key vk_i is also modified by using eq (4) to the weight w_i' as

$$w_i'=w_i+\sum_{I_j\in\mathbf{DM}(\mathbf{q})} TF_{ij}, \quad (9)$$

where f_{ij} in eq (4) is the percentile of the subimage (corresponded to the visual key vk_i) to the image I_j that belongs to $\mathbf{DM}(\mathbf{q})$.

iii) The proposed system retrieves again an image set $\mathbf{M}(\mathbf{q}')$ by using the modified query $\mathbf{q}'$ and weights $\{w_i'\}$.

iv) The new matched images $\mathbf{M}(\mathbf{q}')$ with quality index information are presented to the user. If the user requests the relevance feed back again, then go back to i). It is expected (and confirmed by the experiments) that the user will satisfy the result within a few repetitions.

IV. Rerieval Experiments Using Corel Image Data

The proposed system is constructed by JAVA language on a Pentium 4 (1.6GHz) PC (Windows operating system). The system is tested on a subset of the COREL database. For each image, the features and subimages mentioned in II.B are calculated and stored in advance.

A. *Experimental Conditions*

To evaluate effects of visual keys, the proposed image retrieval system is compared with the system based on the similarity of the feature vectors. In the compared system image retrieval is done based on the ranking obtained by the similarity related score

$$score(I_j,\mathbf{QS})=\sum_{qs\in\mathbf{QS}}\left(1-\min_{s_{jk}\in\mathbf{S}(I_j)} d(qs,s_{jk})\right), \quad (13)$$

where $\mathbf{QS}$ is a query that is a set of subimages, and the higher score means the better candidate for the user.

The system is evaluated for a subset (in total 1,000) of the COREL database, formed by 10 image categories, each containing 100 pictures. The categories are Africa, Beach, Buildings, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains, and Food. Subimages in total 9,000 are extracted from 1000 images based on the method mentioned in II B i). A retrieved image is considered correct if and only if it is in the same category as the query.

The followings are the summary of notations used in the performance evaluation. The *ID(i)* denotes the *Category ID* of image i ($1\le i\le 1{,}000$ since there are in total 1,000 images in the sub-database). For a query image i, $r(i,j)$ is the *rank* of image j (ranked position of image j in the retrieved images for query image i, it is an integer between 1 and 1,000). The precision $p(i)$ for query image i is defined by

$$p(i)=\frac{1}{100}\sum_{1\le j\le 1{,}000, r(i,j)\le 100, ID(j)=ID(i)} 1, \quad (10)$$

that is the percentile of images belonging to the category of image i in the first 100 retrieved images.

The overall average retrieval precision p for all images in the sub-database is defined as

$$p=\frac{1}{1{,}000}\sum_{i=1}^{1{,}000} p(i). \quad (11)$$

B. *Retrieval Speed & Easy Operation*

In the environment described in the beginning of IV, the necessary retrieving times (for one retrieval) on the proposed system and the conventional one are within 4ms and over 200ms, respectively. Both are allowable but the proposed system provides about fifty-times faster retrieval speed than that of the conventional one.

The query should be selected by the user from 9,000 subimages in the case of the conventional one, whereas the presented system requests the user to choose the query from 80 visual keys. The presented system realizes an easy operation for users.

C. *Effect of Visual Keys to Precision*

The overall average retrieval precision p is 0.377 in the case

of the conventional one with 9,000 subimages (cf. chain double-dashed line in Fig. 2). If the candidates of the query (by users) are restricted to visual keys without indexing (that are subset of 9,000 subimages), the precision p becomes better as the number of visual keys increases (cf. ⊟ line in Fig. 2). It is clear from Fig. 2 that the precision p is monotonically increasing and is saturated, where 90% of p=0.377 (in the case of 9,000 subimages) is achieved when the number of visual keys becomes 200. The presented system, i.e., query by visual key (with indexing), gives the precision p as shown by ⊖ line in Fig. 2. If the number of visual keys increases the precision p becomes better in the first part. Too small number of visual keys prohibits the user to select desired visual keys, resulted in poor retrieval precision. After the number of visual keys exceeds 80, the precision p decreases because the number of retrieved images **M(q)** decreases as the number of visual keys becomes bigger (cf. equations (6), (2), and (3)), resulted in poor retrieval precision (even though the user can select enough visual keys). It should be noted that the precision p achieves 90% of p=0.377 at the number of visual keys around 80. Hence the number of visual keys is set as the experimental optimum 80 hereafter.

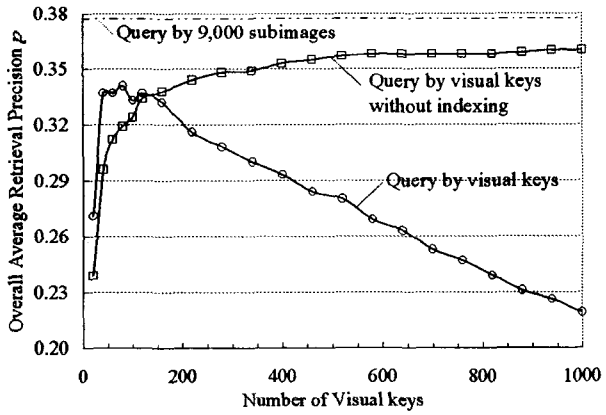

Fig.2: Overall average retrieval precision vs. number of visual keys.

D. Effect of Relevance Feedback to Precision

The experimental result of relevance feedback is shown in Fig. 3. The overall average retrieval precisions p's without relevance feedback in the cases of query by visual keys and query by 9,000 subimages are 0.337 and 0.377 (broken line and dashed line in Fig. 3), respectively. The result of relevance feedback with 9,000 subimages is shown by △ line in Fig. 3, where desired images are selected by user from top scored 20~100 images indicated by the system. The results of relevance feedback with visual keys are shown by ◇ line (one iteration), ✕ line (two iteration), and ✳ line (five iteration). The overall average retrieval precision p in the case of two iteration is better than that of one iteration but there is no significant improvement in the case of more than three iteration. The precision p of relevance feedback with visual keys is better than that of relevance feedback with 9,000 subimages when the number of selected images by user exceeds five.

It is concluded from the experiment that the relevance feedback with visual keys (2 iteration) using 5 selected images is mostly recommended.

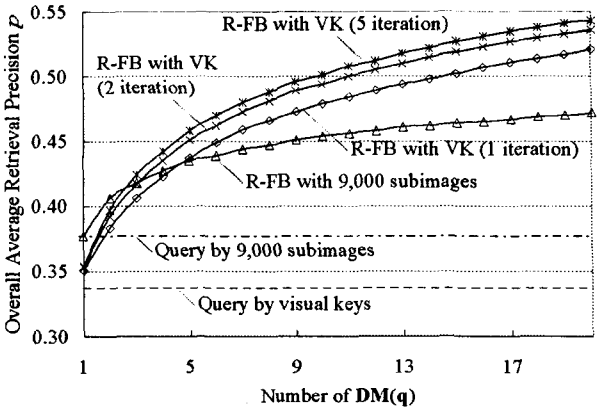

Fig.3: Overall average retrieval precision vs. number of selected images by user.

V. CONCLUSIONS

A new region based image retrieval system is presented based on the concept of visual keys. The proposed system is constructed by JAVA language on a Pentium 4 (1.6GHz) PC (Windows operating system), and is tested on a subset of the COREL database. The user should select the query from 9,000 subimages in the case of conventional system, whereas the presented system requests the user to choose the query from 80 visual keys. The presented system realizes an easy operation for users. On the other hand, the necessary retrieving times for one retrieval on the proposed system and the conventional system are within 4ms and over 200ms, respectively. Relevance feedback widely used in text retrieval systems is also applied by taking into account to the balance of the overall average retrieval precision, easy operation, and retrieval speed. Finally it is concluded that the relevance feedback with visual keys (2 iteration) using 5 selected images is mostly recommended.

The proposed system is able to retrieve images from general image database whose size is up to 100,000, i.e., most of the current general image databases are covered. In the field of image retrieval, there exist many human related issues and soft computing will be a useful tool to analyze and to step up the proposed system with visual keys.

REFERENCES

[1] Y. Chen, J. Z. Wang, A region-based fuzzy feature matching approach to content-based image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, No. 9, pp. 1252-1267, 2002

[2] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, Blobworld: A System for Region-Based Image Indexing and Retrieval, Proc. Int. Conf. on Visual Inf. Sys., pp.509-516, Amsterdam 1999.

[3] G. Salton, C. Buckley, Term Weighting Approaches in Automatic Text Retrieval, Information Processing and Management, Vol. 24, No. 5, pages 513-523, 1998.

[4] J. J. Rocchio, Relevance feedback in information retrieval, the SMART Retrieval System – Experiments in Automatic Document Processing, Salton, G. (Ed.), Pretice Hall, pp. 313-323, 1971.