

INTERACTIVE REGION-BAESD IMAGE RETRIEVAL

Shih-Huan Tseng and Chuech-Yu Li and Chiou-Ting Hsu

Department of Computer Science

National Tsing Hua University, Hsinchu, Taiwan

Abstract

This paper proposes an interactive region-based image retrieval system. Initially, we use color clustering by K-means algorithm and region labeling to segment an image into regions. Several geometric invariant features, such as dominant color, color histogram, moment invariants, and co-occurrence texture features, are extracted to index each region. Then, we describe each image as a combination of feature vectors of the segmented regions.

To measure the distance between images, we define a hierarchical distance function as a liner combination of region features. The retrieved results can be refined via interactive relevance feedback. To learn the “ideal” query regions that the users really want, we derive the weighting parameters of distance measurement using optimized learning technique. A series of experiments have been performed to demonstrate the effectiveness and performance of our work.

1. Introduction

Since a great deal of images becomes available through networks and high-capacity storages, an image retrieval scheme is indispensable for efficient indexing, browsing, and retrieval.

Traditionally, we annotate and query images by keywords. The keyword-based annotation takes a lot of time, cost, and effort to process the voluminous images. In addition, due to the rich contents in images and the subjectivity of human perception, different people often give different annotations to the same content. Thus, users cannot retrieve the desired image if the queried keyword is not properly annotated. Hence, instead of indexing and querying by keywords, content-based image retrieval (CBIR) aims to retrieve images according to their own content. Most CBIR techniques represent each image as a combination of low-level features. Once users submit an example image for query, the retrieval system automatically ranks and displays the retrieved results in the order of similarity.

Much of the previous works [18-22, 32] treats an image as a single entity and describes it by global features. However, for a complex image containing multiple objects, global features cannot capture local information (e.g. shape) and spatial relationship between objects. Representing an image as a combination of multiple objects is thus advantageous than processing an image as a single entity. Nevertheless, since current segmentation techniques cannot determine meaningful objects from arbitrary images, regions with homogeneous features are usually employed to describe the image objects. Such region representation also support partial query to a specific part of an image.

Relevance feedback [11-13, 21, 23, 26, 27] is an interactive process to reduce the gap between semantic concepts and low-level features from users' feedback. In each iteration of

feedback, the users rate retrieved images according to their preferences. The retrieval system then updates the distance measurement or the probability structure from the modified rating. The retrieved results will gradually converge to the desired images or move to other search paths based on users' feedback. In [12-13], an optimization-based learning technique is used to dynamically adjust the weights of different features. In [23], the weights of each segmented region from relevant images and irrelevant images are automatically determined through a formulated linear system. In [21], PicHunter employed Bayes's learning technique to predict the target image.

In this paper, we propose an interactive region-based image retrieval system. Figure 1 illustrates our proposed flowchart. Images are initially segmented into regions by K-means color clustering and region labeling algorithm. Next, we extract low-level features to describe each region. We measure the distance between query image and database images by a hierarchical distance function. At each feedback step, users may identify some of the retrieved images as relevant. Then, the retrieval system will interactively update the distance measurements and return the refined results. Experiments show that query by representing an image as regions performs better than representing an image as a single entity.

The rest of this paper is organized as follows. Sec. 2 describes our region segmentation algorithm by color clustering and labeling technique. Sec. 3 presents the feature extraction and indexing for each segmented region. Sec. 4 discusses our learning algorithm based on the

optimization learning technique. The experiments of two query types are compared and discussed in Sec. 5. And Sec. 6 concludes our work.

2. Region Segmentation

Representing an image as a combination of multiple regions is advantageous than processing it as a single entity. In order to incorporate local features into image retrieval, we have to partition images into regions beforehand. With segmented regions, local features (e.g. shape, color, texture) can be more easily estimated and indexed, and such locally indexed features enable us to query more specific content of an image.

Since automatic segmentation of an image into semantically meaningful regions is still one of the most challenging problems, most existing works [15,17,23,24] defined regions as areas with homogeneous low-level features. In this work, we define a region as a neighboring area with color homogeneity. We first partition the color space into several clusters. Then, the labeling algorithm is carried out to connect pixels within the same color cluster into regions.

2.1. Color Clustering

Color is a widely used visual feature in region segmentation [1][4], and is commonly represented as a point in three-dimensional color space, such as RGB, HSV, and CIE $L^*u^*v^*$ [5][9]. In CIE $L^*u^*v^*$ color space, the perceptual distance between colors can be approximately measured by Euclidean distance. Hence, we apply K-means algorithm [1-4] to partition color components in $L^*u^*v^*$ space.

Since different images have various color distributions, the number of clusters varies with the complexity of the content. Here, we apply a splitting technique [10] to adjust the number of clusters. In each splitting step, if the average distortion of a cluster is larger than a threshold, we add a perturbation vector to this cluster to construct a new cluster center.

To reduce the computation cost, the $L^*u^*v^*$ space is first quantized into 256 bins ($4 \times 8 \times 8$). The color histogram of an image \mathcal{X} is represented by $\mathbf{h} = [h_1, h_2, \dots, h_{256}]^T$, where h_i ($1 \leq i \leq 256$) indicates the number of pixels in \mathcal{X} with quantized color component $\mathbf{c}_i = [L_i, u_i, v_i]^T$. Let $\mathcal{Z}^t = \{z_1^t, z_2^t, \dots, z_m^t\}$ denote the set of m clusters at the t -th ($t \geq 0$) iteration stage, and \mathbf{y}_i^t be the center of cluster z_i^t . The initial cluster center is set to be $\mathbf{y}_1^1 = (\sum_{i=1}^{256} h_i \mathbf{c}_i) / (\sum_{i=1}^{256} h_i)$. At each step, we calculate the average distortion σ_i^t of each cluster z_i^t and the total average distortion σ_{mean}^t of all clusters by:

$$\sigma_i^t = \frac{\sum_{\mathbf{c}_j \in z_i^t} \|\mathbf{c}_j - \mathbf{y}_i^t\|^2}{|z_i^t|}, \text{ and} \quad (1)$$

$$\sigma_{\text{mean}}^t = \frac{\sum_{i=1}^m \sum_{\mathbf{c}_j \in z_i^t} \|\mathbf{c}_j - \mathbf{y}_i^t\|^2}{\sum_{i=1}^m |z_i^t|}. \quad (2)$$

If the stopping condition is not satisfied, we apply the splitting technique to increase the number of clusters. For each cluster z_i^t , if $\sigma_i^t > \sigma_{\text{mean}}^t$, we construct a new cluster center by $\mathbf{y}' = \mathbf{y}_i^t + [\sigma_L, \sigma_u, \sigma_v]^T$, where σ_L , σ_u , and σ_v are standard deviations of z_i^t in L^* , u^* , and v^* components, respectively. Then, we iteratively update the clustering result and cluster center by nearest neighbor rule.

The splitting and clustering processes are stopped when the average total distortion σ_{mean}^t comes close enough to $\sigma_{\text{mean}}^{t-1}$. However, since the average total distortion decreases as the number of cluster increases, a proper threshold is difficult to define for an optimal number of clusters. Thus, we further include the cluster-validity method [3] to refine the above procedure. We calculate the cluster separation measure $\rho(n)$ as follows:

$$\rho(n) = \frac{1}{n} \sum_{i=1}^n \max_{1 \leq j \leq n \wedge i \neq j} \left(\frac{\sigma_i + \sigma_j}{\mu_{ij}} \right), \quad n \geq 2, \quad (3)$$

where $\sigma_i = \frac{\sum_{\mathbf{c}_k \in \mathcal{Z}_i^t} \|\mathbf{c}_k - \mathbf{y}_i^t\|_1}{|\mathcal{Z}_i^t|}$, $\mu_{ij} = \|\mathbf{y}_i^t - \mathbf{y}_j^t\|_2$, and n is the number of clusters. The smallest $\rho(n)$ indicates best separation and the corresponding n is the optimal number of clusters. For example, in Fig. 2, the optimal segmentation occurs at the first valley of $\rho(n)$ curve, and Fig. 2(b) shows the segmentation results.

2.2. Region Labeling

Since color clustering ignores spatial information of the partitioned colors, we use the labeling algorithm [6] with eight-connectivity to build connected regions. Instead of applying region merging to process fragmental regions, small regions with area less than 1% of the whole image will be dropped. We observed that small regions are usually insignificant, and reducing the number of regions is much preferable for the indexing step.

3. Feature Extraction and Indexing

To index each segmented region, we have to find out several representative features to describe

the region. Many visual features have been widely used in image retrieval, such as colors, textures, and shapes.

3.1. Color

Because color is insensitive to background complication and independent of image size and orientation, color feature becomes one of the most widely used visual features in image retrieval. Several color feature representations have been applied in image retrieval, like color histograms [5], color moment [14], and color set [5]. Here, we extract the dominant color and the color histogram to index a region.

Dominant color is measured as the mean color of a region. We represent the dominant color as three components in $L^*u^*v^*$ space.

Color histogram counts the number of pixels of a region into appropriate histogram bins. The description by color histograms is invariant to translation, rotation about the image axes, small off-axis rotations, scale changes and partial occlusion [6]. Similarly, we quantize the $L^*u^*v^*$ color space into 256 color histogram bins ($4 \times 8 \times 8$).

3.2. Shape

In this work, we measure the moment invariants to represent the shape features of each region.

The moment invariants $\varphi_1 \sim \varphi_7$ [5, 7, 8] are invariant to translation, rotation and scaling of shapes. Because the dynamic range of $\varphi_1 \sim \varphi_7$ is very large, it is much convenient to deal with $Sgn(\varphi_i)\log|\varphi_i|$ rather than φ_i , where $Sgn(\varphi_i)$ retains the sign of φ_i .

3.3. Texture

Texture feature contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. The co-occurrence matrix [6] is a statistical texture description used to measure the repeated occurrence of several configurations within a region. However, the co-occurrence matrix itself is rarely used for similarity comparison [6]. Instead, several numeric features computed from co-occurrence matrix are used to represent the texture feature in a more compact form. Therefore, we measure five standard features [6], includes energy, entropy, contrast, homogeneity, and correlation, which are derived from a normalized co-occurrence matrix to index the texture feature of each region.

4. Retrieval with Relevance Feedback

As aforementioned, we represent an image as a combination of regions, and describe each region by its low-level features. Next, we aim to find the user's desired image via relevance feedback by optimization-based learning process [12-13].

We start with the definition of our notations. An image with n_x regions is denoted by a set $\mathcal{X} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{n_x}\}$, where each region vector $\mathbf{r}_i = [\mathbf{f}_{i1}, \mathbf{f}_{i2}, \mathbf{f}_{i3}, \mathbf{f}_{i4}]^T$ consists of 4 feature vectors, which are dominant color, color histogram, moment invariants and texture features, respectively. Each feature vector $\mathbf{f}_{ij} = [f_{ij1}, f_{ij2}, \dots, f_{ijn_j}]^T$ contains n_{j} components, where n_{j} is 3, 256, 7, and 5 for $j = 1\sim 4$.

4.1. Feature Distance Measurement

As described in Sec. 3, the dominant color of a region is represented as a feature vector \mathbf{f}_1 , which consists of 3 components in $L^* u^* v^*$ space. The texture feature vector \mathbf{f}_3 comprises five numeric components derived from the co-occurrence matrix. Seven shape moment invariants make up the shape feature vector \mathbf{f}_4 . We define the distance measurements for these three feature vectors as a generalized Euclidean distance [12, 6, 16, 25]:

$$d(\mathbf{f}_i, \mathbf{q}_i) = (\mathbf{f}_i - \mathbf{q}_i)^T \mathbf{W}(\mathbf{f}_i - \mathbf{q}_i), \quad i \in [1,3,4], \quad (4)$$

where \mathbf{q}_i denote the feature vector of a query region $\mathbf{q} = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4]^T$.

Though the distance between two color histograms can also be measured by Euclidean distance, such measurements usually exhibit poor performance [5]. Thus, we adopt the idea arises from histogram intersection, and measure the distance of two color histograms as the reciprocal of the histogram intersection:

$$d(\mathbf{f}_2, \mathbf{q}_2) = \frac{\sum_{k=1}^n (\mathbf{f}_2, k)}{\sum_{k=1}^n \min((\mathbf{f}_2, k), (\mathbf{q}_2, k))}. \quad (5)$$

where k is the index of the histogram bin. If the denominator equals zero, then $d(\mathbf{f}_2, \mathbf{q}_2)$ is defined as a maximum value based on the implementation platform.

4.2. Relevance Feedback by a Single Region or an Image

We have already defined the distance measurement for each feature vector. The overall distance between region vectors is then defined as a linear combination of feature vectors [13]. In the following discussion, the symmetric matrix \mathbf{W}_j ($j=1,3,4$) denotes the weights of individual entries within feature vector \mathbf{f}_j , and u_j ($j=1\sim 4$) denotes weight of feature vector \mathbf{f}_j . The

distance function between a region $\mathbf{r} = [\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4]^T$ and the ideal query region $\mathbf{q} = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4]^T$ is then defined as:

$$d(\mathbf{r}, \mathbf{q}) = \sum_{j=1}^4 u_j d(\mathbf{f}_j, \mathbf{q}_j) = u_2 d(\mathbf{f}_2, \mathbf{q}_2) + \sum_{j=1,3,4} u_j (\mathbf{f}_j - \mathbf{q}_j)^T \mathbf{W}_j (\mathbf{f}_j - \mathbf{q}_j). \quad (6)$$

Let N be the number of relevant images or regions. The above distance function leads to the following optimization problem:

$$\text{minimize } \sum_{i=1}^N d(\mathbf{r}_i, \mathbf{q}), \text{ where} \quad (7)$$

$$d(\mathbf{r}_i, \mathbf{q}) = \sum_{j=1}^4 u_j d(\mathbf{f}_{ij}, \mathbf{q}_j) = u_2 d(\mathbf{f}_{i2}, \mathbf{q}_2) + \sum_{j=1}^4 u_j (\mathbf{f}_{ij} - \mathbf{q}_j)^T \mathbf{W}_j (\mathbf{f}_{ij} - \mathbf{q}_j),$$

subject to the constraint $\sum_{j=1}^4 \frac{1}{u_j} = 1$, and $\det(\mathbf{W}_j) = 1$, for $j = 1, 3, 4$.

Note that, if there are no constraints for u_j and \mathbf{W}_j , this optimization problem will reduce to solution of all zeros. To solve this minimization problem, we use the Lagrange multipliers to reduce the constrained problem to an unconstrained one. Thus, the optimal solution for feature vector \mathbf{q}_j in ideal query region \mathbf{q} is just the weighted average of feature vectors in relevant regions [12]:

$$\mathbf{q}_j = \frac{\sum_{i=1}^N \mathbf{f}_{ij}}{N}, \quad j = 1 \sim 4. \quad (8)$$

The optimal solution for \mathbf{W}_j is derived as [12]:

$$\mathbf{W}_j = (\det(\mathbf{C}_j))^{1/K_j} \mathbf{C}_j^{-1}, \quad (9)$$

where $\mathbf{C}_j = [c_{rs}]_{K_j \times K_j}$ is the weighted covariance matrix of \mathbf{f}_j and

$$c_{rs} = \sum_{i=1}^N (\mathbf{f}_{ijr} - \mathbf{q}_{jr})^T (\mathbf{f}_{ijs} - \mathbf{q}_{js}). \quad (10)$$

And the optimal solution for u_j is [13]:

$$u_j = \sum_{l=1}^4 \sqrt{\frac{d_l}{d_j}}, \quad (11)$$

where $d_j = \sum_{i=1}^N d(\mathbf{f}_{ij}, \mathbf{q}_j)$. This solution indicates that if the total distance d_j for feature j is small, this feature should be assigned a higher weight.

4.3. Relevance Feedback by Multiple Regions

To measure the overall distance between an image $X = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{n_x}\}$ with n_x regions and the query image $Q = \{\mathbf{r}_1^q, \mathbf{r}_2^q, \dots, \mathbf{r}_{n_q}^q\}$ with n_q regions, we define the distance function as:

$$d(X, Q) = \sum_{a \in [1, N_Q]} w_a \min_{b \in [1, N_X]} d(\mathbf{r}_b, \mathbf{r}_a^q), \quad (12)$$

where $\mathbf{r}_b = [\mathbf{f}_{b1}, \mathbf{f}_{b2}, \mathbf{f}_{b3}, \mathbf{f}_{b4}]^T$, $\mathbf{r}_a^q = [\mathbf{q}_{a1}, \mathbf{q}_{a2}, \mathbf{q}_{a3}, \mathbf{q}_{a4}]^T$ and

$$d(\mathbf{r}_b, \mathbf{r}_a^q) = \sum_{j=1}^4 u_{aj} d(\mathbf{f}_{bj}, \mathbf{q}_{aj}) = u_{a2} d(\mathbf{f}_{b2}, \mathbf{q}_{a2}) + \sum_{j=1,3,4} u_{aj} (\mathbf{f}_{bj} - \mathbf{q}_{aj})^T \mathbf{W}_{aj} (\mathbf{f}_{bj} - \mathbf{q}_{aj}). \quad (13)$$

Let N be the number of the relevant images, the optimization problem is formulated as:

$$\begin{aligned} \text{minimize } & \sum_{i=1}^N d(X_i, Q) = \sum_{i=1}^N \sum_{a \in [1, N_Q]} w_a \min_{b \in [1, N_X]} d(\mathbf{r}_{ib}, \mathbf{r}_a^q), \text{ where} \\ & , \quad (14) \\ d(\mathbf{r}_{ib}, \mathbf{r}_a^q) &= \sum_{j=1}^4 u_{aj} d(\mathbf{f}_{ibj}, \mathbf{q}_{aj}) = u_{a2} d(\mathbf{f}_{ib2}, \mathbf{q}_{a2}) + \sum_{j=1,3,4} u_{aj} (\mathbf{f}_{ibj} - \mathbf{q}_{aj})^T \mathbf{W}_{aj} (\mathbf{f}_{ibj} - \mathbf{q}_{aj}) \end{aligned}$$

subject to the constraint $\sum_{a=1}^{N_Q} \frac{1}{w_a} = 1$, $\sum_{j=1}^4 \frac{1}{u_{aj}} = 1$, and $\det(\mathbf{W}_{aj}) = 1$ (for $j = 1, 3, 4$).

Since the regions within an image are non-overlapped, we model the overall distance as a linear combination of region distances. For a specific region a in the query image Q , the optimal solutions for \mathbf{q}_a , \mathbf{W}_{aj} and u_{aj} ($j=1, 2, \dots, n_f$) are the same as Sec. 4.2. However, we have to measure \mathbf{q}_a , \mathbf{W}_{aj} and u_{aj} for each region, the time complexity is much higher

than the case of single region. The optimal solution for $w_a (a = 1, 2, \dots, N_Q)$ is derived as:

$$w_a = \sum_{l=1}^{N_Q} \sqrt{\frac{D_l}{D_a}}, \quad (15)$$

where $D_a = \sum_{i=1}^N \min_{b \in \{1, N_X\}} d(\mathbf{r}_{ib}, \mathbf{r}_a^q)$.

5. Experimental Results and Analysis

The following experiments use Corel data set as our test data. We select 2000 images, including sunsets, skies and mountains, animals, fruits, foods, objects, etc., from Corel data set. Each category contains at least 100 images with size of 192×128 or 128×192 pixels. Many existing systems also used Corel data set to test their performance [13][18]. Some of the systems [19] use pre-selection categories. In our work, we don't classify the data set and just mix all these categories into a test database.

We use precision and recall to measure the retrieval performance. Precision P is determined as the number of retrieved relevant images over the number of the total retrieved images, while recall R is defined as the number of retrieved relevant images over the total number of relevant images. The meaning of "relevant" of retrieval results is highly dependent on users' subjectiveness.

Fig. 3 shows three test images from Food Objects, Sunset and Evening Skies, and Museum Duck Decoys directories. We choose 50, 150, and 100 images from these three directories as relevant images, respectively. The experiments are performed on two query types: query by an image and query by multiple regions. The former treats the whole query image as a single

region, and employs the technique discussed in Sec. 4.2 to adjust the weights. The latter represents the image as a combination of multiple regions, and employs the technique discussed in Sec. 4.3 to adjust the weights.

5.1. Query by an image

Fig. 4 shows the retrieved results. The retrieved images in Fig. 4(a) contain relevant and irrelevant ones. Though the irrelevant ones are different to the query image subjectively, the low-level features are quite similar to that of the query. After we select some relevant images for updating the weighting matrix, the retrieved images of the first feedback are shown in Fig. 4(b). We can see that the weights of features, corresponding to the positions of slider bars, changed. Also, the system ranks the retrieved images according to the newly weighted distance measurement. If we continue to select relevant images from Fig. 4(b), the retrieved results are shown in Fig. 4(c). Here the weights of features and ranks of retrieved images changed again.

Three precision-recall curves of test images from Fig. 3 are shown in Fig. 5, where *no RF* indicates no relevant feedback is performed, and *1st RF* and *2nd RF* indicate retrieval with first and second feedback. From Fig. 5, we observed that relevance feedback did improve the performance. However, when the recall rate is higher than 0.4, the performance degrades very fast. We found that when the relevant images are few, if one of the dissimilar images with similar low-level features is highly ranked, the precision rate will decrease quite fast for high recall rate. In addition, the second feedback does not always perform better than the first one.

We observe that user usually select obvious relevant images at first feedback, but may feel difficult to select more relevant images at second feedback. Thus the learning of new query region and feature weights highly depends on users' selection.

5.2. Query by multiple regions

Fig. 6 shows the results of query by multiple regions. Note that the result in Fig. 6(a) is better than Fig. 4(a), because local features of regions are now combined to adjust the distance measurement. Fig. 6(b) and (c) show the result after the first and the second feedback.

The overall performance in Fig. 7 is better than that in Fig. 5, because such query type takes similar regions into consideration. Fig. 7(a) performs best because the relevant images contain obvious objects and noticeable shape features. The relevant images in Fig. 7(b) are natural images and have no similar local features. For example, the shape of suns may be covered by cloud, sea or other objects, and the color of some sunset images lays on red, yellow, or orange.

Overall, we observed that once the regions in an image are not distinct and well segmented, the extracted features represent poor local information about this image, and the learning process cannot derive a good query region and weights of features based on these ill-segmented regions.

6. Conclusion

We have developed an interactive region-based image retrieval system with relevance feedback.

We perform color clustering by K-means technique and follow with labeling algorithm to segment an image into regions. To describe each region, we extract dominant color and color histogram as color features, moment invariants as shape feature and co-occurrence matrix as texture feature. Combine the above features as a hierarchical feature vectors, we applied the optimized learning technique [12] [13] to derive the optimized distance function via relevance feedback.

From our experiments, the performance of query by multiple regions is better than that of query by an image, because local features are much easily extracted from segmented regions. Nevertheless, the performance of query by multiple regions highly depends on the segmentation results, since good shape features can only be extracted from well-segmented regions.

References

- [1] Kompatsiaris, M.G. Srinivas, "Spatiotemporal segmentation and tracking of object for visualization of videocoferece image sequences," *IEEE Trans. CSVT*, vol. 7, no. 1, pp. 1388-1402, Dec. 2000.
- [2] S.Z Selim, M.A. Ismail, "K-means-type algorithm," *IEEE Trans. PAMI*, vol. 6, pp. 81-87, Jan. 1984.
- [3] A. Hanjalic, H. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis," *IEEE Trans. CSVT*, vol. 9, no. 8, pp. 1280-1289, Dec. 1999.
- [4] T. Uchiyama, A. Arbib, "Color image segmentation using competitive learning," *IEEE Trans. PAMI*, vol. 16, no. 12, pp. 1197-1206, Dec. 1994.
- [5] A. Del Bimbo, *Visual information retrieval*, Morgan Kaufmann, 1999.

- [6] L.G Shapiro, G.C. Stockman, *Computer vision*, Prentice Hall, 2001.
- [7] M. Sonka, V. Hlavac, R. Boyle, *Image processing, analysis, and machine vision*, 2nd edition, Brook/Cole, 1999.
- [8] R.G. Gonzalez, R.E. Woods, *Digital image processing*, Addison Wesley, 1993
- [9] Color Space, <http://cs.fit.edu/wds/classes/cse5255/cse5255/davis/index.html>.
- [10] Y. Linde, A. Buzo, and R.M Gray, "An algorithm for vector quantization design," *IEEE Trans on Communications*, COM-28"84-95, January 1980.
- [11] Y. Rui and T.S. Huang, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Trans. CSVT*, vol. 8, no. 5, pp. 644-655, Sep. 1998.
- [12] Y. Isikawa, R. Subramanya, and C. Faloutsos, "MindReader: querying database through multiple examples," *Proceeding of 24th VLDB Conference*, 1998.
- [13] Y. Rui, T.S. Huang, "Optimizing learning in image retrieval," *CVPR'00*, , Jun. 2000.
- [14] Y. Rui, T.S. Huang, and S.F. Chang, "Image retrieval: current techniques, promising directions, and open issues," *Journal of Visual Communication and Image Representation* 10, pp.39-62, Jan 1999.
- [15] C.S. Fuh, S.W Cho, and K. Essig, "Hierarchical color image region segmentation for content-based image retrieval system," *IEEE Trans. IP*, vol. 9, no 1, pp. 156-162, Jan. 2000.
- [16] J.S. Payne, L. Hepplewhite, and T.J Stonham, "Perceptually based metrics for the evaluation of textural image retrieval methods," *Proc. IEEE International Conference on Multimedia Computing and Systems*, pp 793-797, vol. 2, 1999.
- [17] D. Zhong, S.F. Chang, "An integrated approach for content-based video object segmentation and Retrieval," *IEEE Trans. CSVT*, vol. 9, no. 8, pp. 1259-1268, Dec. 1999.
- [18] H.W. Yoo, et al., "Visual information retrieval system via contented-based approach," *Pattern Recognition*, vol. 35, no. 3, pp. 749-769, March 2002.
- [19] A. Mojsilovic et al., "Matching and retrieval based on the vocabulary and grammar of

- color patterns,” *IEEE Trans. IP*, vol. 9, no 1, pp. 38-54, Jan. 2000.
- [20] X.S. Zhou, Thomas S. Huang, “Edge-based structural features for content-based image retrieval”, *Pattern Recognition Letters*, pp. 457-468, 2001.
- [21] I.J. Cox et al., “The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments,” *IEEE Trans. IP*, vol. 9, no 1, pp. 20-37, Jan. 2000.
- [22] A. Vailaya, Mario A.T. Figueiredo, A.K. Jain, and H.J. Zhang, “Image Classification for content-based indexing,” *IEEE Trans. IP*, vol. 10, no 1, pp. 117-129, Jan. 2001.
- [23] J.W. Hsieh et al., “Region-based image retrieval,” *Proc. ICIP*, 2000.
- [24] H. Greco and P. Lambert, “Image retrieval by partial queries,” *Proc. ICIP*, 2001.
- [25] M.K. Hu, “Visual Pattern Recognition by Moment Invariant,” *IRE Trans. Information Theory*, Vol. 8, No. 2, pp. 179-187, 1962.
- [26] T.S Huang, X.S Zhou, “Image retrieval with relevance feedback: from heuristic weight adjustment to optimal Learning methods,” *Proc. ICIP*, 2001.
- [27] J. Yoon, N. Jayant, “ Relevance feedback for semantics based image retrieval,” *Proc. ICIP*, 2001.
- [28] B.S. Manjunath et al., “Color and texture descriptors,” *IEEE Trans. CSVT*, vol. 11, no. 6, pp. 703-715, Jun. 2001.
- [29] M. Bober, “MPEG-7 visual shape descriptors,” *IEEE Trans. CSVT*, vol. 11, no. 6, pp. 716-719, Jun. 2001.
- [30] R. Milanese, “A rotation, translation, and scale-invariant approach to content-based image retrieval,” *Journal of Visual Communication and Image Representation*, pp. 186-196, 1999.
- [31] Y. Rui, T.S. Huang, and S. Mehrotra, “Content-based image retrieval with relevance feedback in MARS,” *Proc. ICIP*, 1997.
- [32] X. Wan and C.-C. J. Kuo, “A new approach to image retrieval with hierarchical color clustering,” *IEEE Trans. CSVT*, vol.8, no. 5, pp. 628-643, Sep. 1998.

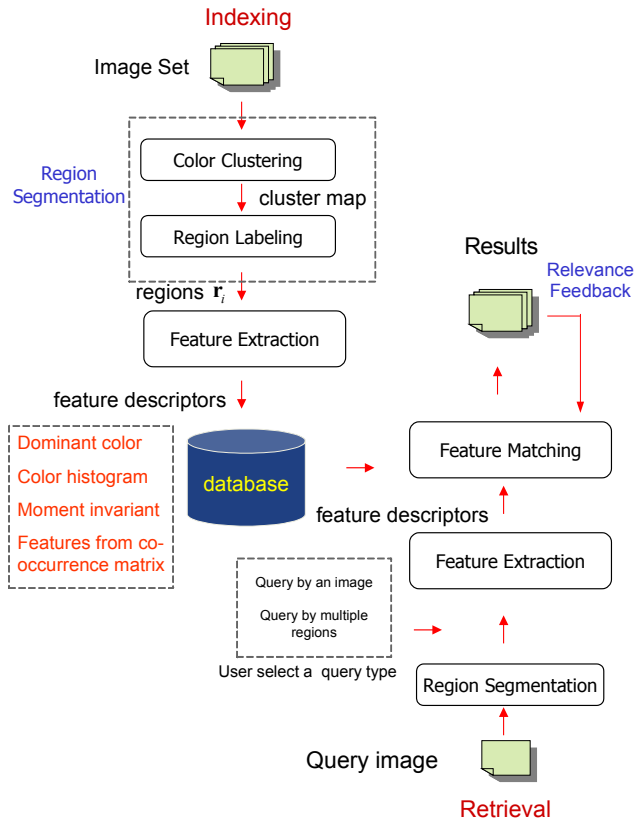


Figure 1. Flowchart of our proposed retrieval system.

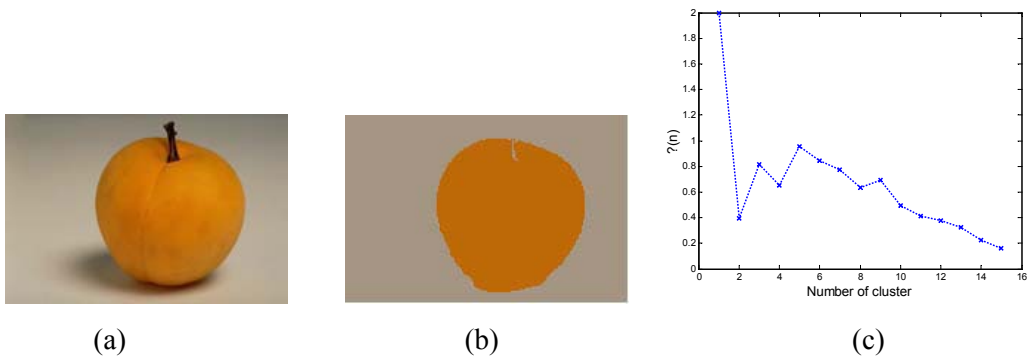


Figure 2. Experiment on color clustering using color-validity method: (a) original image, (b) segmentation result, (c) cluster separation $\rho(n)$ curve.

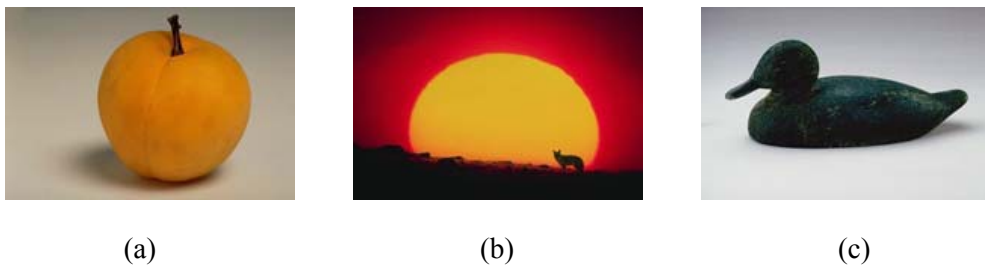


Figure 3. Test images.

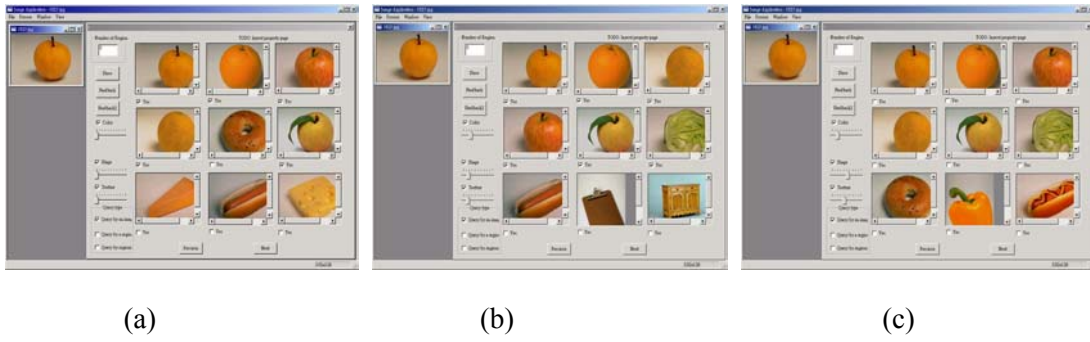
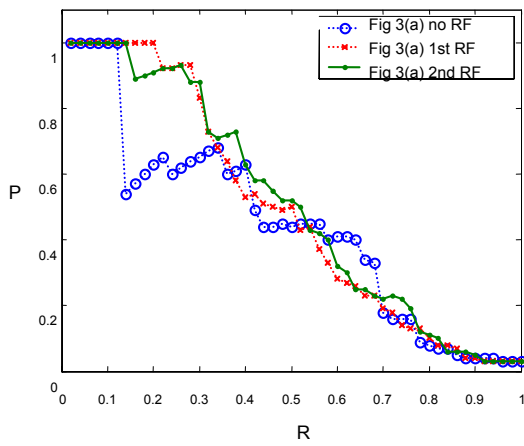
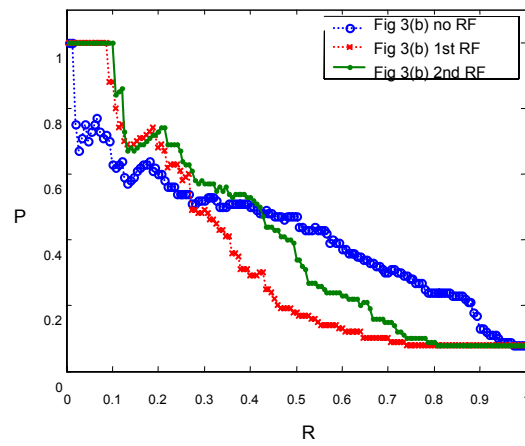


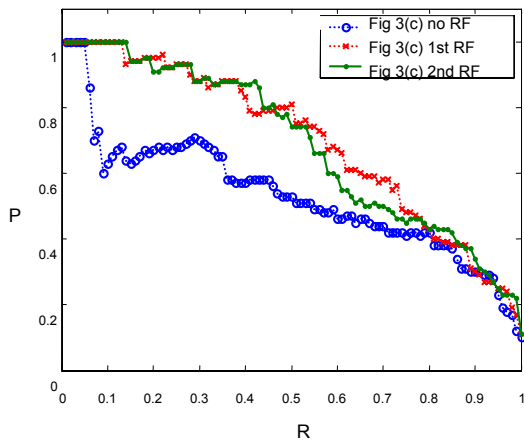
Figure 4. Result of query by an image: (a) no relevance feedback, (b) the first feedback, and (c) the second feedback.



(a)



(b)



(c)

Figure 5. P-R curves of query by an image for Fig. 3(a), (b) and (c).

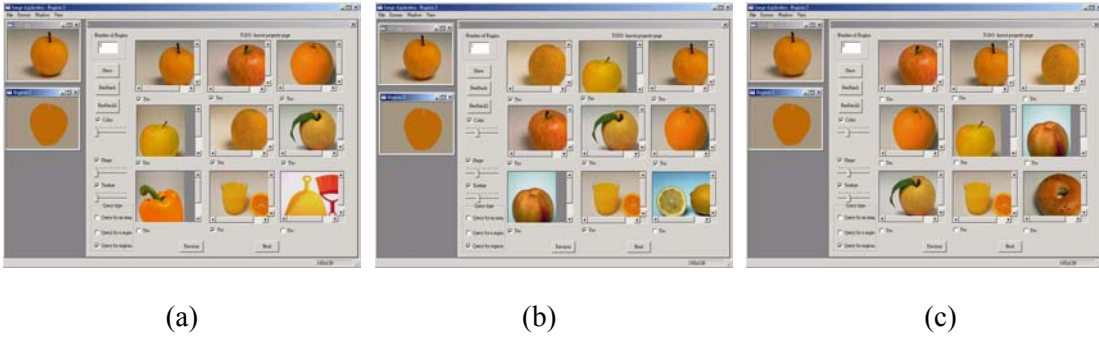
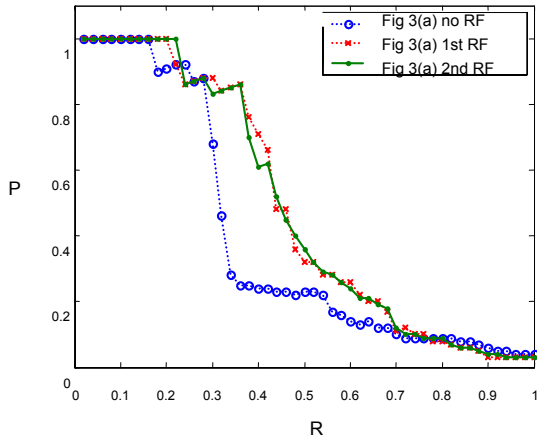
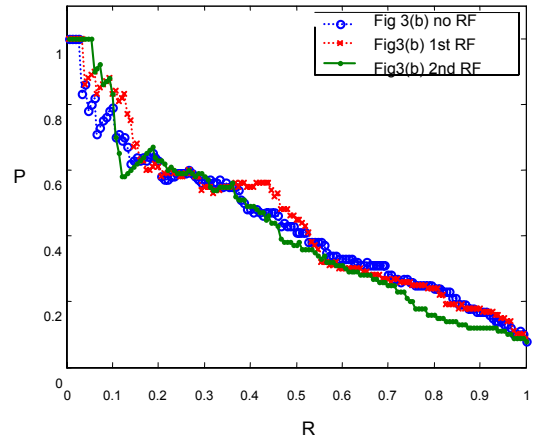


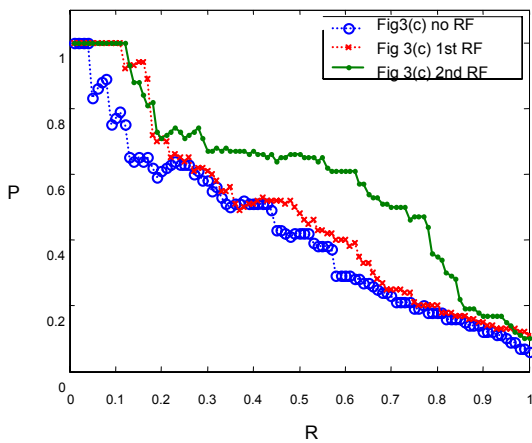
Figure 6. Result of query by multiple regions: (a) no relevance feedback, (b) the first feedback, and (c) the second feedback.



(a)



(b)



(c)

Figure 7. P-R curves of query by multiple regions for Fig. 3(a), (b) and (c).