Expert Systems with Applications 37 (2010) 4984–4998

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Region-based image retrieval system with heuristic pre-clustering relevance feedback

Wan-Ting Su, Ju-Chin Chen, Jenn-Jier James Lien *

Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 70101, Taiwan, ROC

ARTICLE INFO

Keywords:
Relevance feedback
Region-based image retrieval
Content-based image retrieval
Group Biased Discriminant Analysis
Attention center
Integrated region matching

ABSTRACT

Relevance feedback (RF) and region-based image retrieval (RBIR) are two widely used methods to enhance the performance of content-based image retrieval (CBIR) systems. In this paper, these two methods are combined to have the promising result of CBIR. Rather than using a single positive feedback group, the proposed approach embeds RF in the RBIR scheme using multiple positive and negative groups. To guide users in grouping the positive feedbacks, the proposed system provides an objectively heuristic pre-clustering result automatically. Referring to these guiding clusters, the users can then easily and subjectively re-group the positive feedbacks in accordance with his/her particular interests. A region-weighting scheme reflecting the process of human visual perception is proposed to enhance the weighting importance assigned to the region whose pixels are closer to the attention center. Finally, a modified Group Biased Discriminant Analysis (GBDA) is developed and applied to the similarity measure between images constructed on the basis of the region-based relevance feedbacks.

1. Introduction

Content-based image retrieval (CBIR) (Carson, Belongie, Greenspan, & Malik, 2002; Ceccarelli, Musacchia, & Petrosino, 2006; Chen & Wang, 2002; Jing, Li, Zhang, & Zhang, 2004; Li, Wang, & Wiederhold, 2000; Lin, Kao, Yang, & Wang, 2006; Ma, Zhou, Chelberg, & Celenk, 2004; Ma & Manjunath, 1997; Mezaris, Kompatsiaris, & Strintzis, 2004; Nakazato & Huang, 2002; Natsev, Rastogi, & Shim, 1999; Niblack et al., 1993; Pentland, Picard, & Sclaroff, 1994; Rui, Huang, Ortega, & Mehrotra, 1998; Rui & Huang, 2000; Su, Zhang, Li, & Ma, 2003; Su & Lien, 2006; Sun & Ozawa, 2003; Wang, Zha, & Cipolla, 2005; Wood, Campbell, & Thomas, 1998; Yoo, Jung, Jang, & Na, 2002; Yoshizawa & Schweitzer, 2004; Zhou & Huang, 2001) is a technique used for extracting similar images from an image database. The most challenging aspect of CBIR involves the gap between high-level semantic concepts and low-level image features. In general, two approaches are commonly employed to reduce, or at least to bridge, this gap.

The first approach involves extracting the region-based features in order to reflect the focus of the user's perception. Compared to global image feature retrieval schemes (Niblack et al., 1993; Pentland et al., 1994; Stricker & Orengo, 1995), region-based image retrieval systems (Carson et al., 2002; Chen & Wang, 2002; Jing, Li, Zhang, & Zhang, 2002; Jing et al., 2004; Li et al., 2000; Ma & Manjunath, 1997; Mezaris et al., 2004; Natsev et al., 1999; Sun & Ozawa, 2003; Wood et al., 1998; Yoo et al., 2002) apply an image segmentation approach to decompose an image into several regions. This technique more accurately mimics the processes involved in the human visual system. The performance of a region-based image retrieval system is fundamentally dependent on the method used to compare the two images, i.e. the performance is determined by the definition of similarity which is applied when performing the image similarity measurement. Some early region-based image retrieval systems, e.g. those of (Ma & Manjunath, 1997) and Blobworld (Carson et al., 2002), compared images on the basis of individual region-to-region similarities. Netra (Ma & Manjunath, 1997) uses localized region information to index images and includes an optional manual region-pruning step. During retrieval, the user is provided with the segmented regions of an image and is required to specify the expected number of regions. Meanwhile, Blobworld (Carson et al., 2002) models the joint distribution of the image features using a Gaussian mixture model and applies the Expectation–Maximization (EM) algorithm to estimate the parameters of the model. When performing a query, the user is required to select the desired region of the query image and the corresponding features to be used when evaluating similarity. This querying system provides users with a significant control over the retrieval process. However, automatic and semantically precise image segmentation is still an unresolved problem, as discussed in Li et al. (2000). For example, an image segmentation algorithm may decompose the image of one penguin into a single region (i.e. the entire penguin), but the image of another penguin into two regions (e.g. the head and the body), as shown in Fig. 1. Due to the

* Corresponding author. Tel: +886 6 2757575x62540; fax: +886 62747076.
E-mail address: jjlien@csie.ncku.edu.tw (Jenn-Jier James Lien).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.12.015

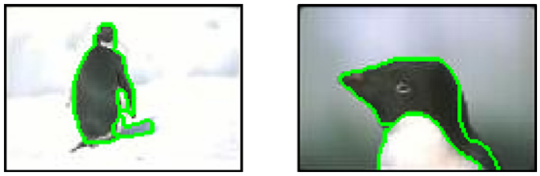


Fig. 1. Segmentation results of two images of a penguin.

difficulty in segmenting accurately and completely, it is insufficient for the user simply to choose the target query region, particularly when the image lacks distinctive objects or scenes.

To ensure robustness against inaccurate segmentation, several image-to-image similarity measurements, which combine the information from all regions of the two images, have been proposed (Chen & Wang, 2002; Li et al., 2000; Sun & Ozawa, 2003). For example, the integrated region matching (IRM) algorithm (Li et al., 2000) performs a region-based retrieval process in which images are segmented by their color, texture and shape features using the *k*-means algorithm, and a region of one image is then matched to several regions of the second image. In the IRM approach, the similarity between two images is expressed using the weighted sum of distances. In Sun and Ozawa (2003), Sun and Ozawa proposed a semantic-meaningful approach in which a pixel clustering algorithm was applied for image segmentation in the Low-Low-subband (LL-subband) of the image wavelet transformation and the feature vector of the segmented region was hierarchically extracted from all of the wavelet subbands. Finally, the similarity measure was determined by the weighted sum of distance and the weights of the feature components of the feature vector were tuned semantically. This study adopts the IRM approach to reduce the uncertainty of the region segmentation and to improve the retrieval performance.

The second approach taken to reduce the gap between the high-level semantic concepts and the low-level image features involves the use of relevance feedback. This approach employs an online learning scheme to improve the retrieval extraction performance by applying positive and negative samples according to the user's subjective perception (Jing et al., 2004; Mezaris et al., 2004; Niblack et al., 1993; Rui et al., 1998; Rui & Huang, 2000; Su et al., 2003; Wood et al., 1998; Wu, Tian, & Huang, 2000; Yoshizawa & Schweitzer, 2004; Zhou & Huang, 2001). A variety of relevance feedback methods have been proposed based on different learning mechanisms. In IDQS (Wood et al., 1998), for example, a query is initiated by selecting the region of interest from a key image. After an initial retrieval attempt, feedback is provided in the form of an acceptance or rejection of the retrieved images. Subsequently, the Learning Vector Quantization (LVQ) algorithm is used to cluster the selected regions of the feedbacks. Images with regions close to the positive cluster centroids are returned and then reclassified by the user. Rui and Huang (2000) proposed an optimal learning approach utilizing an optimization formulation process, which updated the weights by minimizing the distances between the query and all of the relevant feedbacks retrieved. However, only positive feedbacks were used in this approach. The discriminant-EM algorithm (Wu et al., 2000) formulates image retrieval as a transductive learning problem in a probabilistic framework in which both the labeled and the unlabeled data are used for training. In the Bayesian approach (Su et al., 2003), positive feedbacks are adopted to estimate a Gaussian distribution, which represents the desired query image, while any images near the negative feedbacks are penalized by increasing their distances to the query. Recently, Nakazato and Huang proposed a novel approach, referred to as Query-by-Groups (Nakazato & Huang, 2002), in which the user was provided with a mechanism to specify his/her interests in terms of multiple positive and negative image groups. For example, the system enabled the user to create two positive groups by separating single flowers from a bouquet of flowers, as shown in Fig. 2. To guide users in grouping the position feedbacks, the present study develops an objectively heuristic pre-clustering method. Referring to these guiding clusters, the users can then easily and subjectively re-group the positive feedbacks in accordance with his/her particular interest.

Although many relevance feedback methods using global features have been developed, these methods have only rarely been applied to the RBIR system. Relevance feedback mechanisms based on support vector machines (SVMs) were proposed in Jing et al. (2004) and Mezaris et al. (2004). However, these mechanisms considered only a single positive feedback group. The retrieval system proposed in the current study aims to integrate region-based image retrieval and relevance feedback using multiple positive and negative groups. The feedback algorithm is designed according to the characteristics of the region-based representation. Furthermore, a region-weighting scheme, which mimics the process of human visual perception, is also proposed.

The remainder of this paper is organized as follows. Section 2 presents a brief overview of the proposed system. Section 3 elaborates on the basic elements of the proposed region-based image retrieval system, namely image segmentation, region representation and image similarity measurement. Section 4 develops a relevance feedback process using Group Biased Discriminant Analysis (GBDA) and a heuristic pre-clustering method. Section 5 describes the experimental results obtained in evaluating the performance of the proposed approach. Finally, Section 6 provides some brief conclusions.

2. System overview

Fig. 3 presents a flowchart of the proposed system. As shown, the system comprises two major modules: an offline module to perform region-based image retrieval and an online module to carry out heuristic pre-clustering reference feedback based on GBDA. During the offline preliminary preparations, the features of the segmented regions and the region weights of all of the images in the database are automatically extracted. During the online process, when a query image is supplied by the user, all of the images are sorted according to their similarities to the query image. If the user is dissatisfied with the retrieval results, he or she can specify the feedbacks to use in refining the results in the next iteration. In addition, the proposed system provides heuristic pre-clusters to guide the user in manually grouping the positive feedbacks.

Fig. 4 shows the user interface of the proposed system. The big left block is the retrieved image presentation area, in which 'o', 'x' and '?' denote positive, negative and "don't care" feedbacks, respectively. The top-right block is the feedback group presentation area, in which the number displayed in the positive group list indicates the positive group to which the currently displayed images belong. In the proposed system, the user is provided with

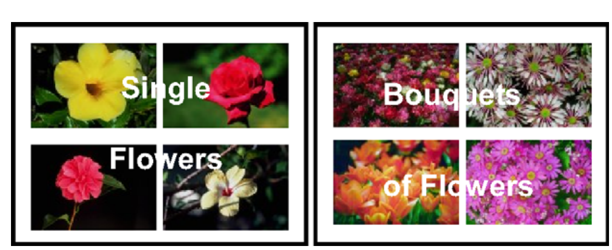


Fig. 2. Two groups: single flowers and bouquets of flowers.

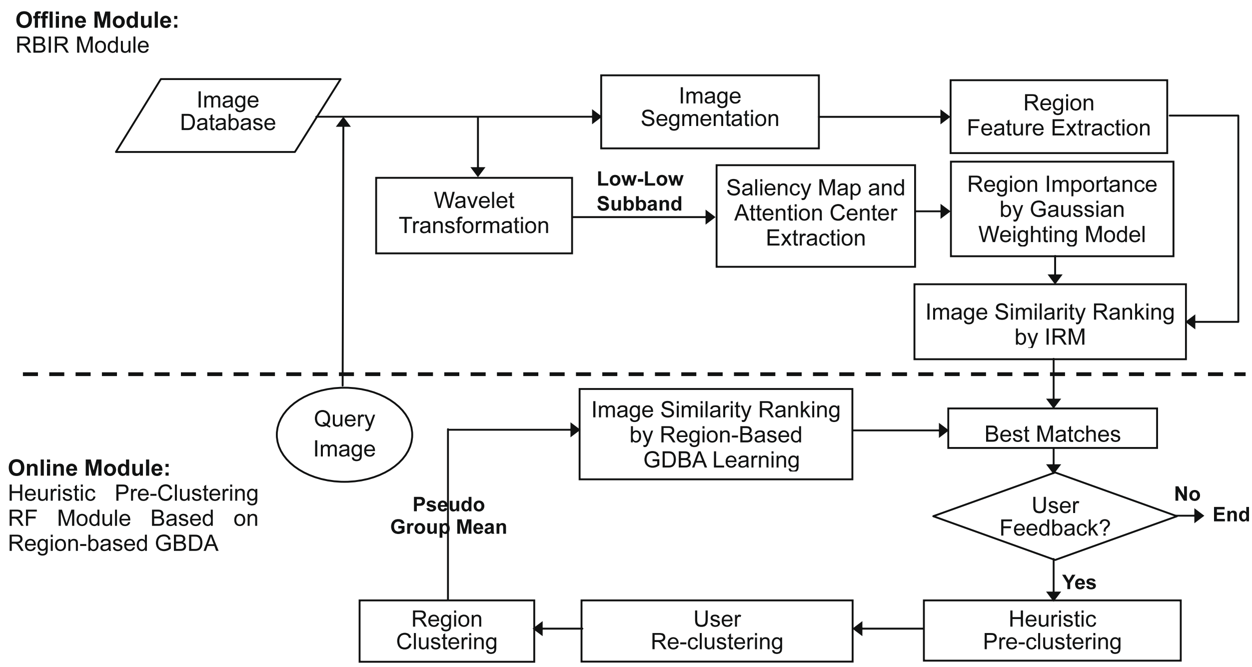


Fig. 3. Flowchart of the proposed system.

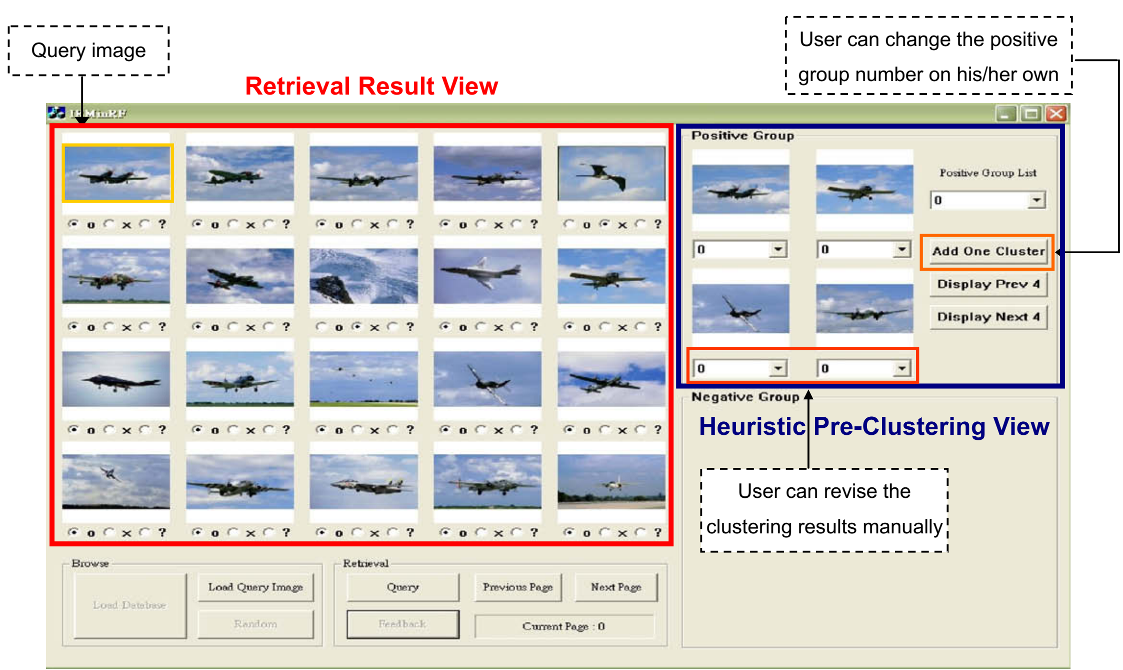


Fig. 4. System interface.

a mechanism to change the group number below the image in order to express his or her interest manually.

3. Region-based image retrieval

The global features-based retrieval systems can not have satisfactory retrieval results because they do not have adequate ability to capture important properties of objects (Niblack et al., 1993; Pentland et al., 1994; Stricker & Orengo, 1995). Instead, several retrieval systems, as introduced in Section 1, are proposed based on the region-based representation, which is closer to the human visual perception, in order to extract important region features (Carson et al., 2002; Chen & Wang, 2002, Jing et al., 2002, Jing et al., 2004, Li et al., 2000; Ma & Manjunath, 1997; Mezaris et al., 2004; Natsev et al., 1999; Sun & Ozawa, 2003; Wood et al., 1998; Yoo et al., 2002). Thus, in the design of our region-based image retrieval module (off-line module) in Fig. 3, image segmentation is applied to decompose an image into several regions. In other words, each image can be represented by a set of regions, which contain low-level features and are assigned by importance weights. Subsequently, both the low-level features and the importance weights are used to measure the similarity between two images.

3.1. Image segmentation

The segmentation algorithm used in the present study is based on the local homogeneity analysis presented in (Jing et al., 2002). The basic principle involved in defining the homogeneity of a pattern is to integrate the directional intensity changes of the surrounding pixels, which are located within a local window. The process is that assuming the location of a pixel is (x,y) and its intensity is $I(x,y)$. Let P be the pattern to compute homogeneity. Currently, we consider P to be a square window of width $2N+1$, $N \in \{1, 2, \ldots, N\}$. Let $c = (x_c, y_c)$ be the center of the pattern with the intensity $I(x_c, y_c)$. Each pixel $p_i = (x_i, y_i), 1 \leqslant i \leqslant (2N+1)^2$ in P corresponds to a vector $cp_i = (x_i - x_c, y_i - y_c)$. Based on cp_i, we construct a new vector f_i:

$$f_i = (I(x_i, y_i) - I(x_c, y_c)) \cdot \frac{cp_{\mathrm{i}}}{\|cp_{\mathrm{i}}\|} \tag{1}$$

Let f be the sum of all vectors defined in P, i.e.,

$$f = \sum_{\mathrm{i}=1}^{(2N+1)^2} f_{\mathrm{i}} \tag{2}$$

Based on above preparations, the measure H is defined as the norm of f, that is,

$$H = \|f\| \tag{3}$$

Thus, applying above process to the original image will yield an Homogeneity image, known as the H-image, which is a gray-scale image whose pixel values correspond to homogeneity values. High and low pixel values in the H-image indicate potential region boundaries and region interiors, respectively. Finally, regions are merged using an agglomerative algorithm based on their color histogram similarities to avoid over-segmentation.

Fig. 5 shows some example patterns and their corresponding H values. Symbols "?" and "×" indicate pixel intensities 0 and 1, respectively. The distance between the closest 4-connective neighboring pixels is d ($d > 0$) and the distance between the closest diagonal neighboring pixels is $\sqrt{2}d$. The solid line denotes f_i, while the dashed line denotes f. For the f_i and f with zero norms, no lines are drawn. In addition, some typical segmentation results are shown in Fig. 6.

Having located the boundaries between segmented regions, it is found that pixels in the boundaries cannot be assigned unambiguously to any particular region. In order to generate a more accurate definition of each region, the boundary pixels are deleted when performing feature extraction. Deleting the ambiguous boundary pixel information has the effect of lessening the uncertainty caused by image segmentation in the region-based image retrieval module.

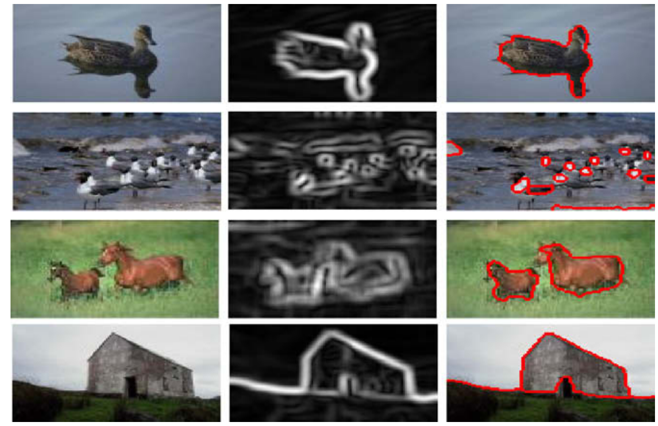


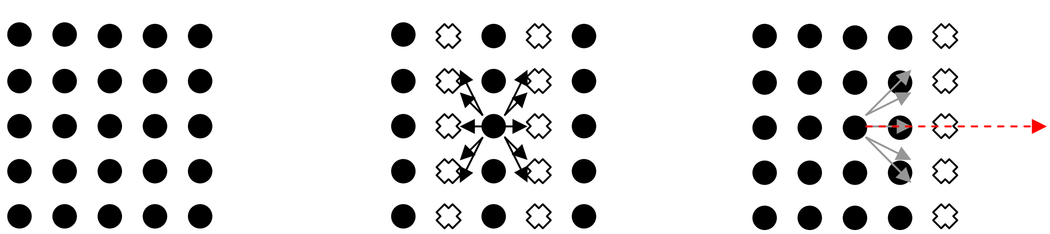
Fig. 6. Left: original image, middle: H-image, right: segmentation results.

3.2. Region feature extraction

In the current implementation, each region of an image in the database is characterized based on its extracted color and texture features. The color features have been analyzed and used in Chan and Chen (2004), Li et al. (2000), Stricker and Orengo (1995), Yoo, Jung, Seo, and Lee (2004) due to its simplicity and efficiency. Considering the color information, the first two moments, i.e. the mean μ and the standard deviation σ, from each channel of the HSV (hue, saturation, and value) color space are extracted in this proposed system. The texture features are extracted by using the wavelet-based approaches (Li et al., 2000; Smith & Chang, 1994; Sun & Ozawa, 2003; Wang, Li, & Wiederhold, 2001). To avoid the feature being dominated by just one of the feature spaces, the difference of dimension between the color and texture feature spaces is controlled to be as small as possible. Therefore, when considering the texture information, the texture feature is represented by the standard deviation of the wavelet coefficients in 4 pyramid-like de-correlated subbands (Smith & Chang, 1994). Eventually, the dimensionality of the visual feature spaces is 10 (6 dimensions in color feature space and 4 dimensions in texture feature space).

3.3. Region importance decision

In this study, two processes are involved in evaluating the region importance decision and then establishing the image similarity measure. First, an attempt is made to extract the attention center of the entire image, where this attention center corresponds to the region assigned particular importance in the process of human visual perception. Second, a Gaussian weighting model is proposed, which assigns higher weights to the pixels near the attention center and lower weights to the pixels which are more



(a) H = 0 (b) H = 0 (c) $H = (1+\sqrt{2}+\frac{4}{\sqrt{5}}) \times d$

Fig. 5. Example patterns and their corresponding H values.

remote. The region of importance is then established on the basis of the individual weights of all the pixels inside it.

3.3.1. Attention center extraction

In Ma and Zhang (2003), concluded that the color contrast in an image is the most important factor in determining the human visual perception of that image. Accordingly, they proposed an image attention analysis method based on the use of a contrast-based saliency map. In their approach, the contrast level in the saliency map was regarded as the image density and the attention center was represented by the centroid of the saliency map.

Wavelet transformation is widely applied in image processing since its properties of multi-resolution decomposition can be adapted to describe image features. To reduce the computational cost while preserving the basic image contents and features, contrast extraction is applied to the wavelet coefficient in the LL-subband, as shown in Fig. 7b. Subsequently, the image contrast is applied in the LUV color space. The contrast value $C_{i,j}$ of pixel p at image location (i,j) is defined as (Ma & Zhang, 2003):

$$C_{i,j} = \sum_{q \in \theta} d(p_{i,j}, q) \tag{4}$$

where the intensity difference d is computed by *Gaussian* distance, θ is the neighborhood region and q is the neighboring pixel of the centered pixel p at (i,j). From pixel-to-pixel contrast addition, $C_{i,j} = Ci,j(L) + Ci,j(U) + Ci,j(V)$. Furthermore, normalizing the contrast values of all of the individual pixels to the scale $[0,255]$ generates a saliency map, as shown in Fig. 7c. From (Ma & Zhang, 2003), the attention center (x_0, y_0) can be computed as:

$$\begin{cases} x_0 = \frac{1}{C_M} \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} C_{i,j} \times i \\ y_0 = \frac{1}{C_M} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} C_{i,j} \times j \end{cases} \tag{5}$$

where $C_M = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} C_{i,j}$ is the 0th order moment of the saliency map and the image size is $M \times N$. Fig. 7d provides several illustrative examples of extracted attention centers.

3.3.2. Gaussian weighting model

According to user's visual perception, each region is assigned an importance weight, which is inverse proportional to the distance between the pixels of the region and the attention center. The Gaussian model of the distance between the pixel and the attention center used to evaluate the pixel importance PI_i can be defined as:

$$PI_i = \exp(-dis(i, C_0)/\sigma) \tag{6}$$

where *dis* is the Euclidean distance of the location difference between pixel i and the attention center C_0, and σ is the standard deviation of all distances between each pixel in the entire image and the attention center. By considering the region sizes, the region weighting importance w_i is given by the sum of the importance of the individual pixels inside region R, i.e.

$$w_i = \sum_{j \in R_i} PI_j \tag{7}$$

And the sum of the total region weighting importance of each image should be normalized to 1. That is, the constraint of the region weighting importance is as:

$$\sum_{i=1}^{m} w_i = 1 \tag{8}$$

where m is the total number of regions in the image.

Fig. 8 shows an example of the region weights calculated using the proposed weighting scheme (left and red) and the area percentage (AP) method (Li et al., 2000), individually. In terms of human visual perception, the weight of the tiger in the image should be assigned a greater importance. As shown, the weight of the tiger assigned using the proposed weighting scheme is higher than that assigned by the AP method by 10%.

3.4. Image similarity measure

Since the image segmentation may not be perfect, the integrated region matching (IRM) scheme (Li et al., 2000) allows one region of an image to be matched to several regions of another image. Compared with image retrievals based on individual region-to-region similarity comparisons, IRM is more robust to inaccurate image segmentations, as shown in Fig. 9.

Assume that an image I_P contains m regions and an image I_Q contains n regions. A matching between regions p_i and q_j is assigned a significance credit $s_{i,j}$, where this credit represents the importance of the matching in determining the similarity between

Fig. 7. Contrast-based image attention analysis: (a) original *MxN*-pixel (width * - height) image, (b) $M/2xN/2$-pixel wavelet LL-subband, (c) $M/2xN/2$-pixel saliency map, and (d) extracted attention center of the original image.

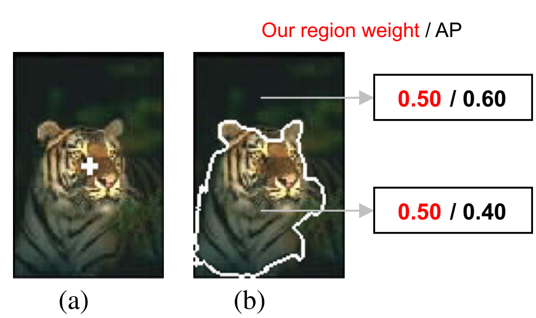


Fig. 8. (a) Extracted attention center, and (b) proposed region weight (left and red) and area percentage (AP) weight of image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

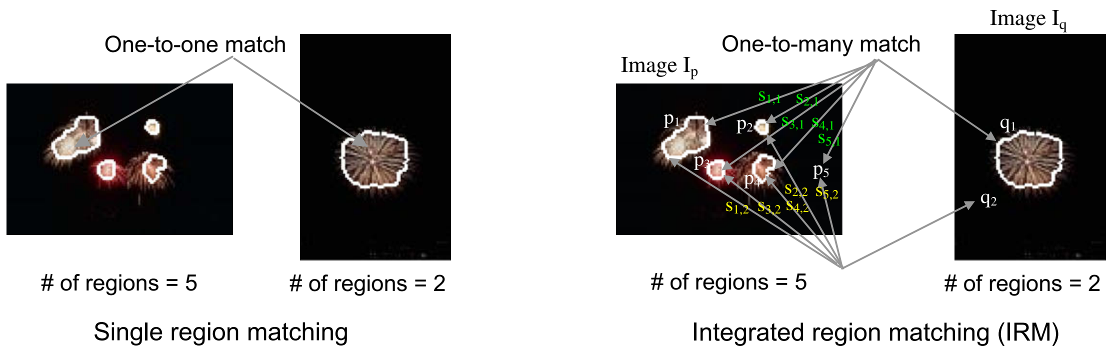


Fig. 9. Robustness of integrated region matching (IRM) to inaccurate image segmentation.

the two images. Furthermore, let $d(p_i, q_j)$, i.e. the region feature distance between p_i and q_j, be the Euclidean distance. The IRM distance between I_P and I_Q is given by the weighted sum of all the similarities between the region pairs, i.e.

$$d(I_P, I_Q) = \sum_{i=1}^{m} \sum_{j=1}^{n} s_{i,j} d(p_i, q_j) \tag{9}$$

The problem of defining the similarity between the two images then becomes to choose the significance credit of all of the region pairs.

IRM attempts to fulfill the significance credits between regions by assigning as much significance as possible to the region pair with minimum distance. The matching problem is solved by this greedy algorithm under the following constraints:

$$s_{i,j} \geqslant 0, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n \tag{10}$$

$$\sum_{j=1}^{n} s_{i,j} = w_{p_i}, \quad 1 \leqslant i \leqslant m \tag{11}$$

$$\sum_{i=1}^{m} s_{i,j} = w_{q_i}, \quad 1 \leqslant j \leqslant n \tag{12}$$

In addition, for normalization:

$$\sum_{i=1}^{m} w_{p_i} = \sum_{j=1}^{n} w_{q_j} = 1 \tag{13}$$

Consequently, for assigning significance $s_{i,j}$ consistently, the procedure steps of the IRM algorithm are summarized as follows:

1. Set $M = \{(p_i, q_j) | i = 1, \ldots, m; j = 1, \ldots, n\}$
2. Choose the minimum $d(p_{i'}, q_{j'})$ for $(p_{i'}, q_{j'}) \in M$.
3. $s_{i',j'} = \min(w_{pi'}, w_{qj'})$.
4. If $w_{pi'} < w_{qj'}$, set $s_{i',j} = 0$ for $j \neq j'$; else set $s_{i,j'} = 0$ for $i \neq i'$.
5. $w_{pi'} = w_{pi'} - \min(w_{pi'}, w_{qj'})$ and $w_{qj'} = w_{qj'} - \min(w_{pi'}, w_{qj'})$
6. $M = M - \{(p_{i'}, q_{j'})\}$
7. If $\sum_{i=1}^{m} w_{p_i} > 0$ and $\sum_{j=1}^{n} w_{q_j} > 0$ then go to Step 2; else stop.

The values of w_{pi} and w_{qj} are chosen to represent the region weighting importance (i.e. significance) of regions p_i and q_j in the images I_P and I_q, respectively. Both values are assigned by the method described in section 3.3.2.

4. Heuristic pre-clustering relevance feedback based on region-based GBDA

In interactive region-based or content-based image retrieval processes, the system must re-calculate the similarities and corresponding feature weights between query image and all images in the database based on the user's feedbacks to refine the retrieval results. According to Nakazato and Huang (2002), using multiple positive and negative groups can improve the relevance feedback in CBIR systems. Namely, the system supports the user to manipulate image groups directly. Nevertheless, grouping images is not an intuitive process for traditional users. Therefore, the pre-clustering algorithm in following section is developed to assist users in grouping these samples (images or feedbacks). Subsequently, by modifying the Group Biased Discriminant Analysis (GBDA) method (Nakazato & Huang, 2002), this study develops the region-based relevance feedbacks to re-estimate both the similarities and corresponding feature weights between images. The image similarity ranking process in Nakazato and Huang (2002) is then speeded up to become an online calculation of the discriminating transformation matrix in the feature space. Thus, the retrieval results are improved by the feedbacks of the user specification in the process of the online module (Fig. 3).

4.1. Guiding pre-clustering

The approach of Query-by-Groups in Nakazato and Huang (2002) is proposed to analyze the multiple positive and negative image groups, which are manually grouped by users according to their interests. However, for ordinary users, it is much easier to select the positive or negative samples than to group them. Grouping these positive samples is not intuitive, let alone to group images well. Therefore, the system proposed in this study provides an objectively heuristic pre-clustering algorithm to automatically assist user in easily and subjectively grouping these positive feedbacks.

As shown in Fig. 4, users can easily select the positive samples in the left block of the interface, and the proposed pre-clustering algorithm groups these positive samples in the top-right block. The pre-clustering algorithm, as summarized in Fig. 10, commences by computing and sorting the IRM distances between any two positive feedbacks. The two positive images, I_u and I_v, with the minimal IRM distance are added into the first positive class (or group) $PClass_1$. Subsequently, the image, *ChosenSample*, with the shortest distance from one of remaining positive samples to the first class is chosen. If when adding this positive sample to the current positive class, i.e. *NowClass*, the distortion between the images in this positive class is less than a predefined threshold, then this image can be inserted into the current positive class; else a new positive class is created for this positive sample.

Note that the distortion of one positive class is defined as

$$class\_dis = \sum_{(I_P, I_Q) \in PClass_i} d(I_P, I_Q) / pair\_num_i \tag{14}$$

where (I_P, I_Q) is a pair of any two positive samples in this class and $d(I_P, I_Q)$ is the IRM distance between these two samples, and $pair\_num_i$ is the total pair-wise numbers in this class, where $pair\_num_i = C_2^{n_i}$ and n_i is the number of positive samples. The

```
Input: the PSet contains N positive samples (feedbacks) I_1, I_2,…, I_N
Output: the M positive classes PClass_1, PClass_2,…, PClass_M
Method: HeuristicPreCluster(PSet, N)
M ← 1
Create a new class PClass_1 and PClass_1 ← Φ
Initialize PairIndexMatrix
ClassID_{1~N} ← -1
NowClass ← PClass_1
IRMDisMatrix ← ComputeIRMDist(PSet, N)
Sort(IRMDisMatrix, N)
PairIndexMatrix ← ComputePairRelation(IRMDistMatrix, PSet, N)
( I_u, I_v) ← FindMinDisPair(IRMDistMatrix, PairIndexMatrix)
Insert samples I_u and I_v to PClass_1
ClassID_u ← M    and    ClassID_v ← M
WHILE PSet ≠ Φ
    ChosenSample ← SearchSample(NowClass, PSet, PairIndexMatrix, N)
    TmpClass ← NowClass + ChosenSample
    IF (ClassDist (TmpClass ) < thre )    // The distortion of the TmpClass (Eq. (14) )
        Insert ChosenSample to Class_M
    ELSE
        M ← M+1
        Create a new class Class_M
        Insert ChosenSample to Class_M
    ENDIF
    ClassID_{ChosenSample} ← M
    NowClass ← Class_M
    PSet ← PSet – { ChosenSample }
END WHILE
```

Fig. 10. The heuristic pre-clustering algorithm.

threshold value which decides whether the chosen positive sample can be inserted into the current positive class is given as

$$thre = c \times AveFeedbacksDis \tag{15}$$

where *AveFeedbacksDis* is the average of the IRM distances of all positive sample pairs. The c is a constant value between 0 and 1, and controls the pre-clustering results. If c is too small, the clustering results contains too many groups, and vice versa. The value of c is set to be 0.8 in our implantation. Then these processing steps are repeated iteratively until each of the positive feedbacks, I_k, has been assigned to a corresponding class $ClassID_k$.

Fig. 11 shows an example of the online guiding pre-clustering algorithm and this example assumes that four positive feedback samples exist, i.e. $P_1 \sim P_4$. The IRM distances between any two samples are calculated and sorted. Based on this algorithm, the positive images are then classified into corresponding positive classes.

4.2. Region-based Group Biased Discriminant Analysis (GBDA)

Briefly, GBDA (Nakazato & Huang, 2002) attempts to group each positive class (or group), while scattering negative samples away from each positive class, as shown in Fig. 12a. GBDA achieves this via maximizing the following criterion function:

$$J(W) = \arg\max_{W} \left| \frac{W^T S_{PN} W}{W^T S_W W} \right| \tag{16}$$

where S_w is the sum of the within-class scatter matrix of the positive classes and S_{PN} is the sum of the between-class scatter matrix of the positive-to-negative classes.

4.2.1. Pseudo group mean representation

Assume all regions of the samples in one positive class can represent that particular class. Then, all of the regions in this positive class can simply be combined into one region set, which is regarded as a segmented pseudo image with many regions and

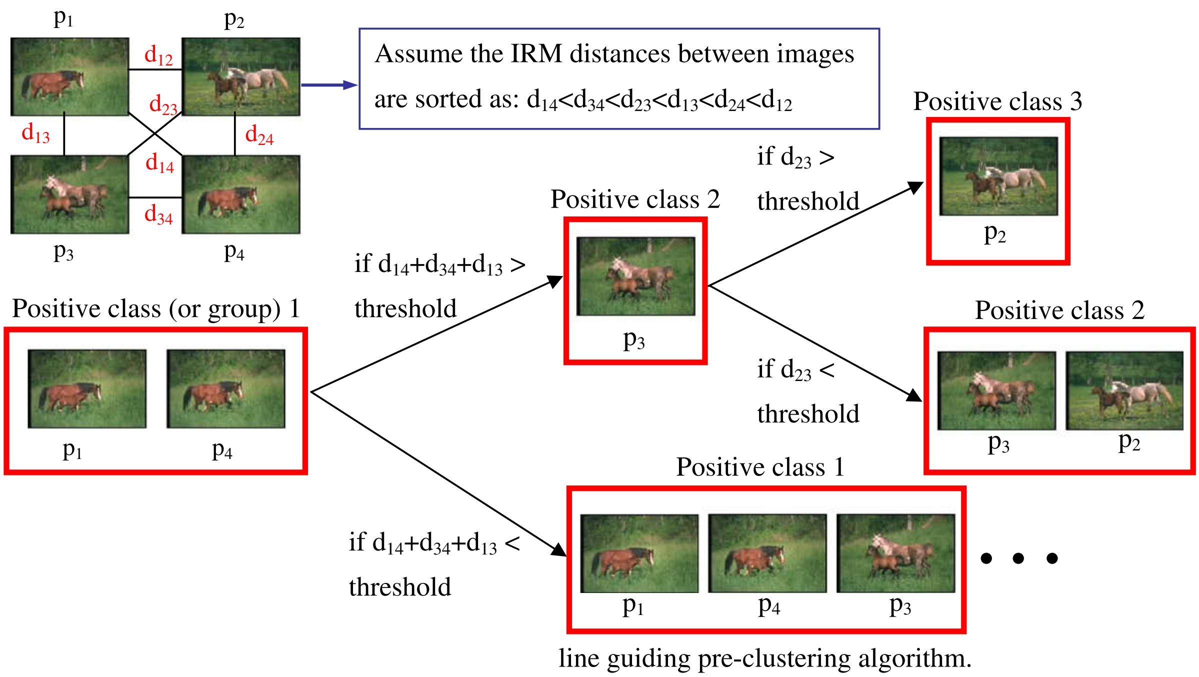


Fig. 11. Illustrative example of online guiding pre-clustering algorithm.

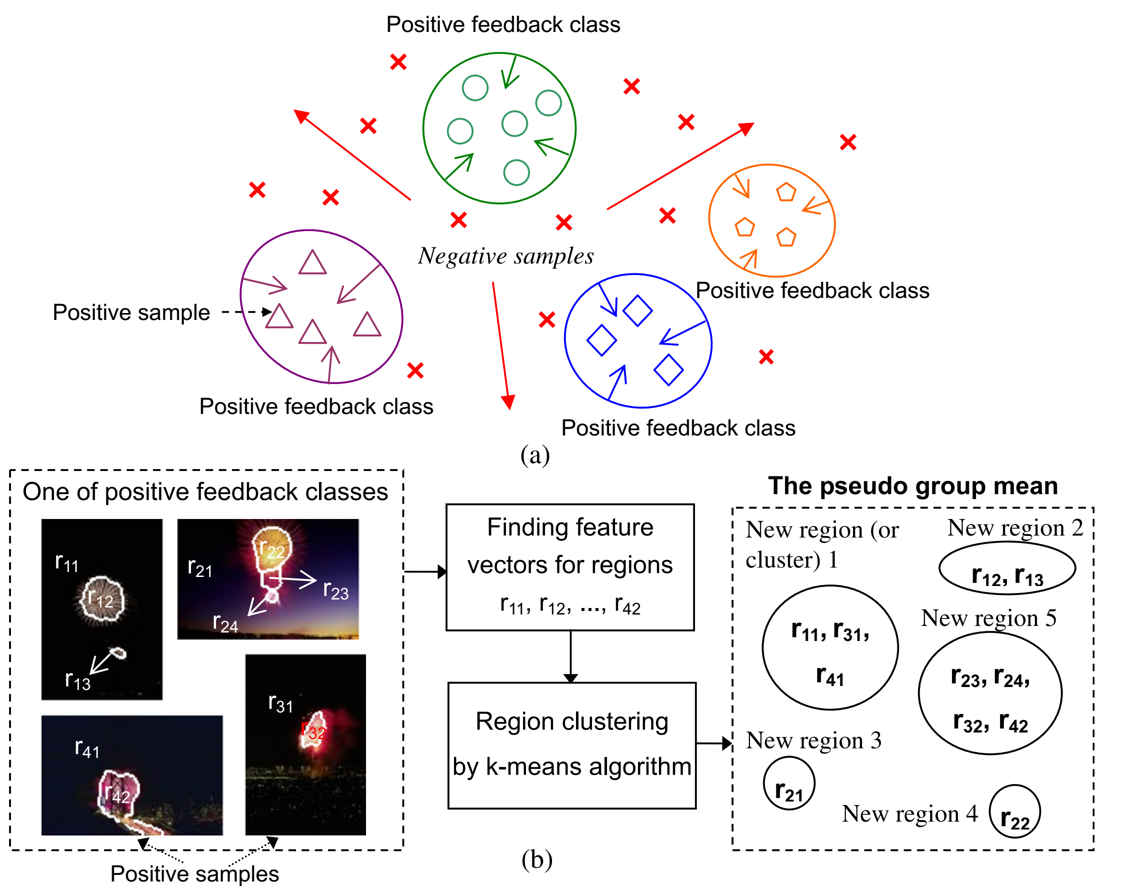


Fig. 12. (a) Concept of GBDA method, which minimizes the sum of the within-class scatter matrix of the positive classes, while maximizes the sum of the between-class scatter matrix of the positive-to-negative classes (Eq. (16)). (b) Flowchart of steps involved in obtaining new regions of pseudo group mean.

referred to as the pseudo (group) mean of the positive class. In other words, in order to obtain the scatter degree of this positive class (Eq. (16)), the pseudo group mean, which is composed of regions, is used to represent the mean of the positive class. Nevertheless, in order to fit the constraint of the region weighting importance, the total importance of the pseudo mean should be normalized

Table 1
Image categories in query set 2 (QS2).

1. Sunset	2. Flower	3. Car	4. Ape	5. Mountain
6. Penguin	7. Tiger	8. Bird	9. Horse	10. Building

to 1. Suppose there are n positive samples in one of the positive classes. Hence, the sum of the total region weighting importance of the pseudo mean is n. To satisfy the constraint, the region weighting importance of the pseudo mean w_m is simply set as:

$$w^k_{m_i} = w^k_i / n \tag{17}$$

where w^k_i is the ith region weighting importance of the kth sample in the positive class.

4.2.2. Region clustering

As the number of feedback iterations increases, the number of regions among all of the positive samples increases rapidly. Furthermore, the execution time required to compare the similarity between images is proportional to the number of regions in those images. Consequently, to avoid slowing the retrieval process, the regions with similar low-level feature vectors are grouped together via clustering. In this study, the k-means algorithm is adopted to group the regions of the samples in the same positive class into a few clusters. Each cluster manifests itself as a new region, which consists of several original regions, within the pseudo mean. The number of clusters k is chosen adaptively by gradually increasing its value. Note that k is initialized to 2 and then increased by 1 at each step. The process stops if the average distortion between all of the positive regions and their corresponding nearest cluster centers falls below a certain threshold, which can be adjusted according to the particular experiment. In the present case, the threshold value is set to 0.01. After clustering, the average feature vector of all original regions in the same cluster is viewed as the feature vector of the new region. The new region weighting importance is given by the sum of all the individual region weighting importance

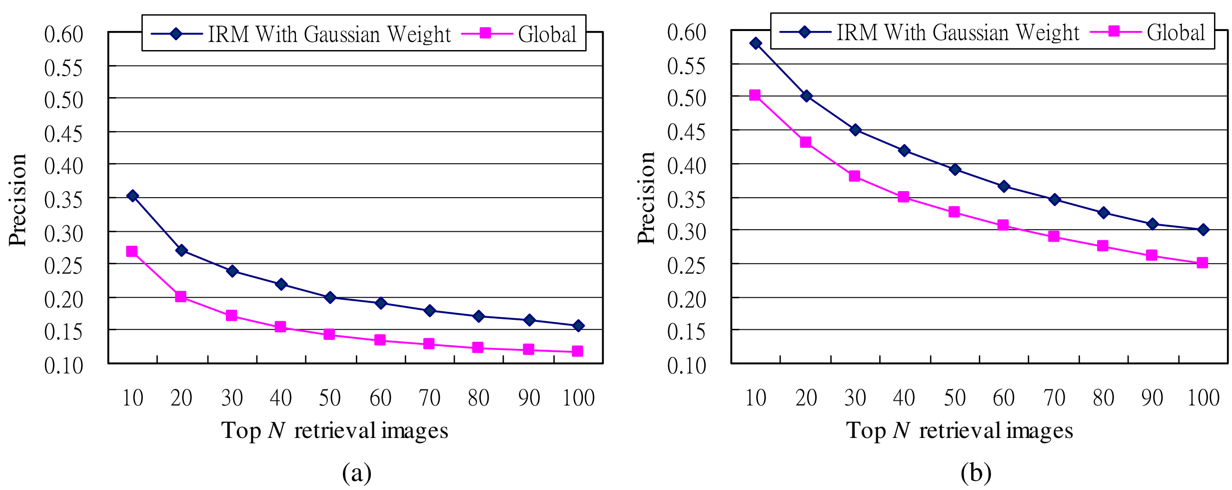


Fig. 13. Average precision comparison between region-based representation and global representation for (a) QS1, and (b) QS2.

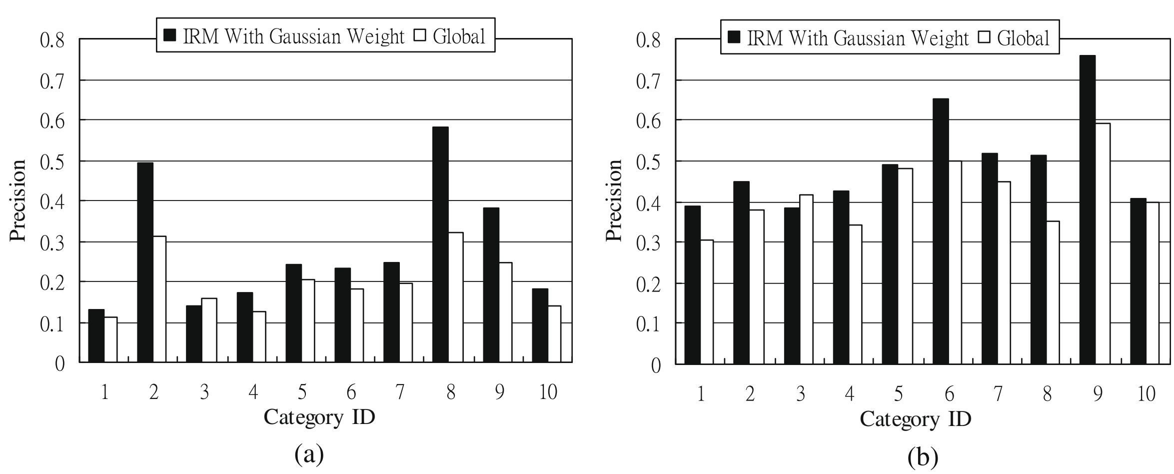


Fig. 14. Precision comparison between region-based representation and global representation in each category for (a) QS1, and (b) QS2.

Table 2
Average precision comparison between Gaussian weighting scheme and area percentage (AP) method by using QS2.

Top N Retrieval Images	Top 10	Top 20	Top 30	Top 40	Top 50	Top 60	Top 70	Top 80	Top 90	Top 100
Gaussian Weight	0.583	0.498	0.447	0.411	0.384	0.361	0.341	0.325	0.310	0.298
AP	0.579	0.491	0.441	0.404	0.377	0.354	0.334	0.318	0.303	0.292

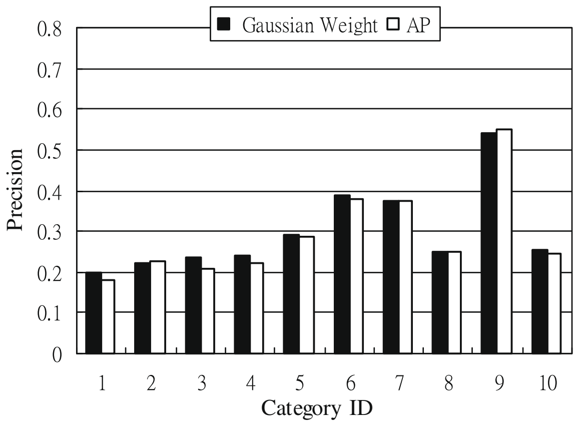


Fig. 15. Precision comparison between two weighting schemes, Gaussian Weight and AP, for QS2 in the initial retrieval result.

in the same cluster. Fig. 12b presents a flowchart of the steps involved in obtaining the new regions (or clusters) of the pseudo group mean in one positive feedback class.

4.2.3. Region-based GBDA formulation

Finally, the terms in (16) are defined as follows:

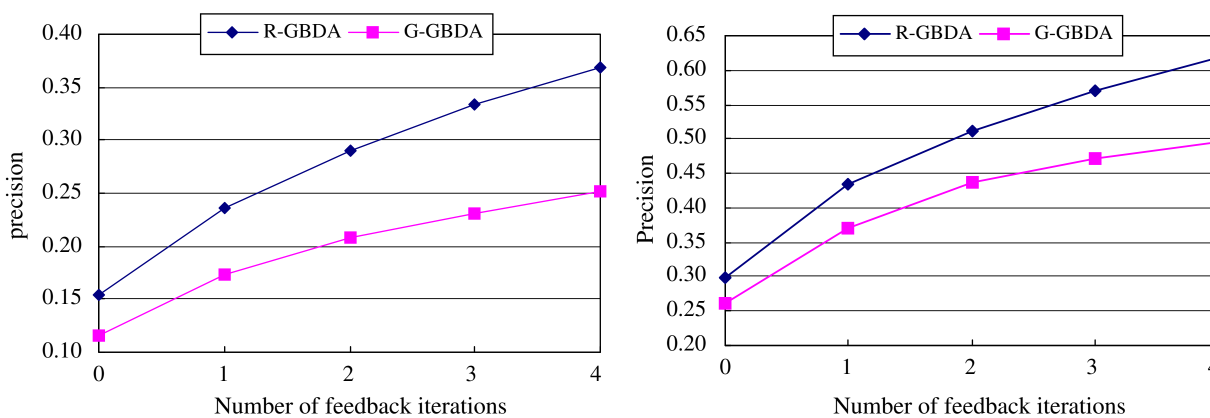
$$S_w = \sum_{k=1}^{c} S_{Pk} \tag{18}$$

$$S_{Pk} = \sum_{x^u \in C_k} s_{m,n} \left(x_m^u - q_n^k\right)\left(x_m^u - q_n^k\right)^T \tag{19}$$

$$S_{PN} = \sum_{k=1}^{c} S_{Nk} \tag{20}$$

$$S_{Nk} = \sum_{y^v \in D} s_{m,n} \left(y_m^v - q_n^k\right)\left(y_m^v - q_n^k\right)^T \tag{21}$$

where x_m^u is the mth region of the uth sample in the kth positive class C_k, y_m^v is the mth region of the vth sample in the negative class, q_n^k is the nth region of the pseudo mean in the kth positive class



(a) (b)

Fig. 16. Accuracy comparison of two relevance feedback algorithms for (a) QS1, and (b) QS2. R-GBDA and G-GBDA denote GBDA algorithm using region-based retrieval and global feature-based retrieval, respectively.

(a) (b)

Fig. 17. (a) First retrieval results in our retrieval system. The top 20 retrieved images are ranking from left to right and from top to bottom, and the image (flower) on the left-top corner is the query image. (***R***/***N***: 7/10 and 15/20). ***R*** /***N***: Total number of relevant images, R, from the top N retrieved images. (b) First feedback (flower) results in our retrieval system. (***R***/***N***: 10/10 and 20/20)

$C_k, s_{m,n}$ is the significance between the mth region of a sample and the nth region of the pseudo mean, c is the number of positive classes and D is the set of negative samples, where a negative sample is considered as one negative class.

As in the Fisher's Discriminant Analysis (FDA), the optimal W that maximizes $J(W)$ is solved as the generalized eigenvector, W_i, associated with the largest eigenvalue, λ_i, i.e.

$$\lambda_i S_w W_i = S_{PN} W_i \tag{22}$$

If S_w^{-1} exists, the solution for W can be found by solving $\lambda_i W_i = \left(S_w^{-1} S_{PN}\right) W_i$. Therefore, the discriminating transformation matrix A becomes:

$$A = \Phi \Lambda^{1/2} \tag{23}$$

where Φ is the matrix whose columns are the eigenvectors of $\left(S_w^{-1} S_{PN}\right)$ and Λ is the diagonal matrix of the corresponding eigenvalues. Once the transformation matrix is available, the distance of the similarity measurement between two images (or samples), e.g. image x with q regions and image y with r regions, can be defined as:

$$distance\ (x, y) = \sum_{m=1}^{q} \sum_{n=1}^{r} s_{m,n} (x_m - y_n)^T A (x_m - y_n) \tag{24}$$

Using this expression, the distance between images in the database and the pseudo mean of each positive class can be compared and sorted.

5. Experimental results

To evaluate the retrieval performance of the proposed system, this study considered a COREL image database containing 17,695 images with 120 × 80-pixel or 80 × 120-pixle resolution. In the evaluation experiments, two query sets were selected. One set, denoted QS1, included 7000 images taken from 50 categories of the database. The second set, QS2, contained 1000 images selected from 10 different categories. Each category of QS2 contained 100 images, each of which was used as queries. The selected categories are listed in Table 1.

5.1. Evaluation of region-based image retrieval

In this section, the region-based image retrieval technique is compared with the typical global representation method (Niblack et al., 1993; Pentland et al., 1994; Stricker & Orengo, 1995) in an initial retrieval result. For the region-based representation, the segmentation algorithm described in Section 3.1 is applied to all

Fig. 18. (a) First retrieval results for the car in our retrieval system. (***R***/***N***: 5/10 and 7/20). (b) First feedback (car) results in our retrieval system. (***R***/***N***: 8/10 and 17/20). (c) Second feedback (car) results in our retrieval system. (***R***/***N***: 10/10 and 20/20).

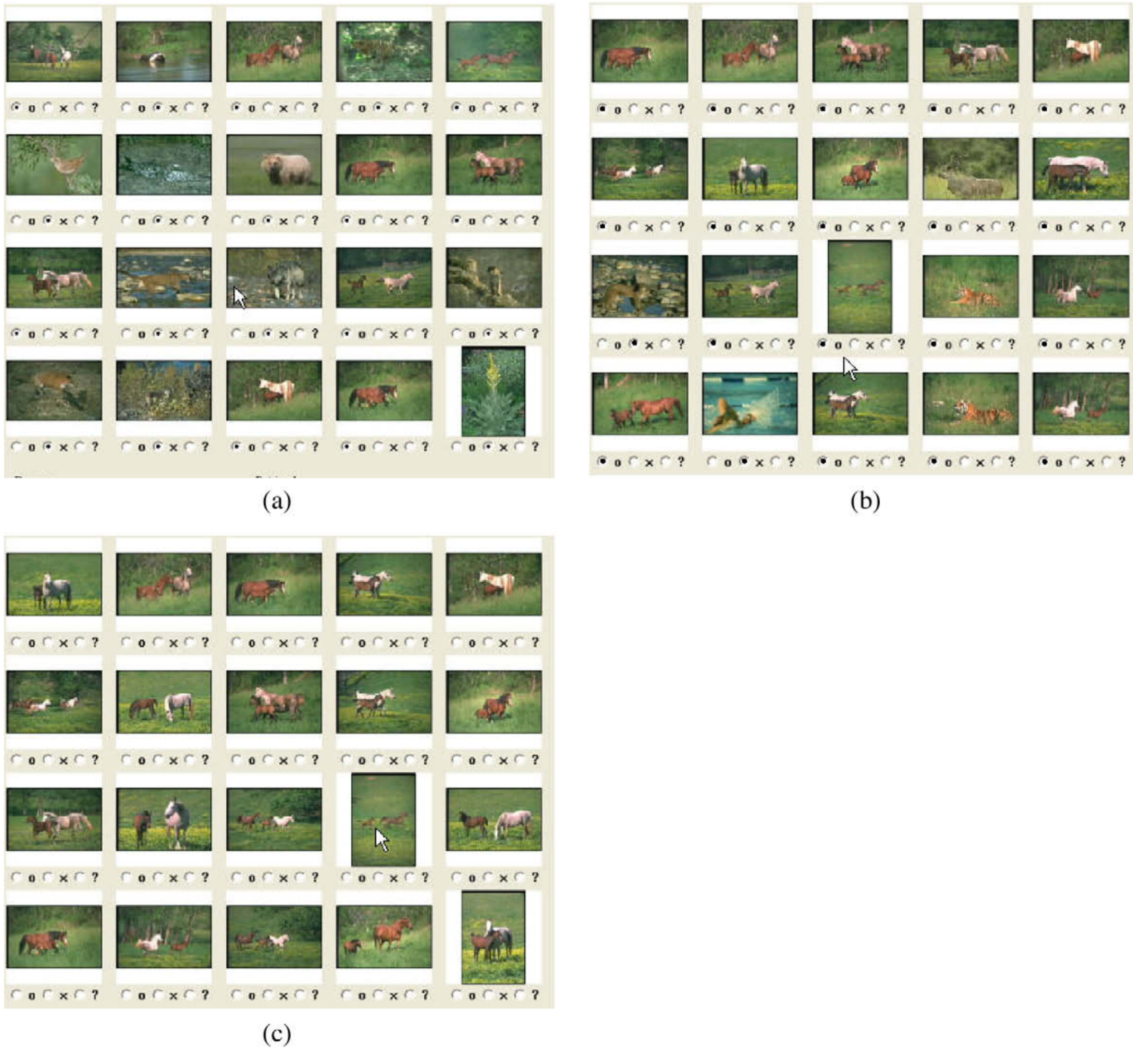


Fig. 19. (a) First retrieval results for the horse in our retrieval system. (***R***/***N***: 5/10 and 9/20). (b) First feedback (horse) results in our retrieval system. (***R***/***N***: 9/10 and 15/20). (c) Second feedback (horse) results in our retrieval system. (***R***/***N***: 10/10 and 20/20).

images in the database. The number of the segmented regions depends on the complexity of the image. To ensure an objective comparison, the same features were used for region-based and global representations. Specifically, the Euclidean distance and the color and texture features described in Section 3.2 were adopted.

A retrieved image was considered as relevant if it belonged to the same category as the query image. For each individual category, the retrieval accuracy was computed as the average precision rate of 100 queries from the top N retrieved images. Note that, the precision rate for each query image is defined as R/N, where R is the total number of relevant images and N is the total number of top retrieved images. The results of the evaluations for the QS1 and QS2 query sets are shown in Fig. 13a and b, respectively. It is clear that the average retrieval accuracy for the total of 1000 queries processed by the region-based retrieval scheme is superior to that of the global representation approach. Fig. 14a and b show the average precision rate of each category when retrieving the top 20 images of QS1 and QS2, respectively. In QS1, for the category of *bird*, the performance of the region-based retrieval method is nearly double that of the global feature-based retrieval approach. Meanwhile, in QS2, for the categories of *penguin*, *bird* and *horse*, which all contain simpler backgrounds, the precision rate of the region-based retrieval scheme is higher than that of global representation by approximately 15%. However, there are some categories for which region-based retrieval is unsuitable, e.g. the category of *car* in both QS1 and QS2. The reason for this may be that this particular category contains diverse and complicated scenes (backgrounds) and various (foreground) object features taken under different view angles. One solution is the use of the object-based image retrieval or the SIFT (scale invariant feature transformation) descriptor to describe the images (Lowe, 1999; Mikolajczyk, Leibe, & Schiele, 2006; Wang et al., 2005).

5.2. Evaluation of gaussian weighting model

As shown in Fig. 8, the importance of the tiger region using the proposed weighting scheme is higher than that assigned in the AP method (Li et al., 2000). To demonstrate the influence of the Gaussian weighting model, Table 2 summarizes the initial retrieval results obtained by using the Gaussian weighting model (Gaussian Weight) and the Area Percentage (AP) method, individually, for QS2. The average precision rate of the total 1000 queries in the 10 categories shows that the proposed weighting scheme is slightly better than that of the AP method. Fig. 15 shows the average precision rate of the two weighting schemes for different semantic categories. Since in current study the Gaussian weighting model is dominated by the region size, its performance improvement is limited.

Fig. 20. (a) First retrieval results for the plane in our retrieval system. (***R***/***N***: 5/10 and 9/20). (b) First feedback (plane) results in our retrieval system. (***R***/***N***: 9/10 and 18/20).

Fig. 21. (a) First retrieval results for the building in our retrieval system. (***R***/***N***: 2/10 and 4/20). (b) First feedback (building) results in our retrieval system. (***R***/***N***: 6/10 and 10/20). (c) Second feedback (building) results in our retrieval system. (***R***/***N***: 10/10 and 13/20). (d) Third feedback (building) results in our retrieval system. (***R***/***N***: 10/10 and 15/20).

(a) (b)

Fig. 22. (a) First retrieval results for the stained glass in our retrieval system. (*R*/*N*: 5/10 and 9/20). (b) First feedback (stained glass) results in our retrieval system. (***R***/***N***: 10/10 and 19/20).

5.3. Evaluation of region-based GBDA

To evaluate the integrated region-based image retrieval and GBDA scheme, the performance of GBDA using global representation was compared. The same color and texture features described in Section 3.2 were adopted. The Gaussian weighting scheme was chosen to assign the weights of the regions. To simulate the user's feedback, all of the retrieved results in the same category as that of the initial query image were regarded as positive samples, while those images in categories different from that of the query were regarded as negative samples. The corresponding results for query sets QS1 and QS2 are shown in Fig. 16a and b, respectively. As shown, for the average precision rate of the total 7000 (1000) queries, the performance of GBDA using region-based retrieval is better than that of GBDA using global representation by 11.71% (12.42%) for QS1 (QS2) after four feedback iterations.

5.4. Examples of using this proposal retrieval system

Fig. 17–22 show some retrieval results. The number of iterations for the relevance feedback is decided by the increasing rate of the performance. And *R* denotes the number of relevant images from the top *N* retrieved images.

6. Conclusions and future works

The major contribution of this study is its integration of RBIR with the relevance feedback algorithm using multiple positive and negative groups. Compared to a single region matching scheme, the overall similarity measure eases the burden on the user and reduces the uncertainty of the automatic region segmentation. A region weighting scheme based on human visual perception has been developed utilizing the properties of the color contrast saliency map. Additionally, color contrast extraction has been conducted in the LL-subband. This not only preserves the basic content of the image, but also lowers the computational cost significantly.

The proposed system guides the user in clustering the positive feedbacks by providing objectively heuristic pre-clustering results. The user can then easily, subjectively and manually revise the clusters by referring to the guiding cluster results. In order to obtain the scatter degree of the positive groups, all of the regions of the positive samples in the group are combined into a region set representing the pseudo group mean of that group. The *k*-means algorithm is adopted to accelerate the feedback process. Finally, the similarity between the query and the other images in the database is obtained by region-based Group Biased Discriminant Analysis.

In a future study, the authors intend to refine the retrieval performance of the developed system by assigning the importance of regions on the basis of user feedback information. Other features of the images, such as the region shape information (Wong, Shin, & Su, 2007) or the spatial relationship between regions, and the use of keywords will also be considered.

Acknowledgements

The authors would like to express their gratitude to professors Q. Tian, M. Nakazato and T.S. Huang for their provision of GBDA source codes.

References

Carson, C., Belongie, S., Greenspan, H., & Malik, J. (2002). Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transaction on Pattern Analysis and Machine Intelligence, 24*(8), 1026–1038.

Ceccarelli, M., Musacchia, F., & Petrosino, A. (2006). Content-based image retrieval by a fuzzy scale-space approach. *International Journal of Pattern Recognition and Artificial Intelligence, 20*(6), 849–868.

Chan, Y. K., & Chen, C. Y. (2004). Image retrieval system based on color-complexity and color-spatial features. *International Journal of System and Software, 71*, 65–70.

Chen, Y., & Wang, J. Z. (2002). A region-based fuzzy feature matching approach to content-based image retrieval. *IEEE Transaction on Pattern Analysis and Machine Intelligence, 24*(9), 1252–1267.

Jing, F., Li, M., Zhang, H. J., & Zhang, B. (2002). Unsupervised image segmentation using local homogeneity analysis. In *Proceedings of the IEEE international symposium on circuits and systems* (Vol. 4, pp. 145–148).

Jing, F., Li, M., Zhang, H. J., & Zhang, B. (2004). Relevance feedback in region-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(5), 672–681.

Li, J., Wang, J. Z., & Wiederhold, G. (2000). IRM: Integrated region matching for image retrieval. In *Proceedings of the 8th ACM international conference on multimedia*, October 2000 (pp. 147–156).

Lin, H. J., Kao, Y. T., Yang, F. W., & Wang, S. P. (2006). Content-based image retrieval trained by Adaboost for mobile application. *International Journal of Pattern Recognition and Artificial Intelligence, 20*(4), 525–542.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the international conference on computer vision* (Vol. 2, pp. 1150–1157).

Ma, L., Zhou, Q., Chelberg, D., & Celenk, M. (2004). Shape-based image retrieval with relevance feedback. In *Proceedings of the IEEE international conference on multimedia and expo*, June 2004 (Vol. 2, pp. 779–782).

Ma, W. Y., & Manjunath, B. S. (1997). NETRA: A toolbox for navigating large image databases. In *Proceedings of the IEEE international conference on image processing*, Santa Barbara, CA, October 1997 (Vol. I, pp. 568–571).

Ma, Y. F., & Zhang, H. J. (2003). Contrast-based image attention analysis by using fuzzy growing. In *Proceedings of the 11th ACM international conference on multimedia*, November 2003 (pp. 734–381).

Mezaris, V., Kompatsiaris, I., & Strintzis, M. G. (2004). Region-based image retrieval using an object ontology and relevance feedback. *Eurasip Journal on Applied Signal Processing, 2004*(6), 886–901.

Mikolajczyk, K., Leibe, B., & Schiele, B. (2006). Multiple object class detection with a generative model. In *Proceedings of the IEEE international conference on computer vision and pattern recognition* (Vol. 1, pp. 26–36).

Nakazato, M., & Huang, T. S. (2002). Extending image retrieval with group-oriented interface. In *Proceedings of the IEEE international conference on multimedia and expo*, August 2002 (Vol. 2, pp. 201–204).

Natsev, A., Rastogi, R., & Shim, K. (1999). WALRUS: A similarity retrieval algorithm for image databases. In *Proceedings of the ACM SIGMOD international conference management of data* (pp. 395–406).

Niblack, W., Barber, R., Equitz, W., Flickner, M. D., Glasman, E. H., & Petkovic, D., et al. (1993). The QBIC project: Querying images by content using color, texture, and shape. In *Proceedings of the SPIE*, San Jose, CA, Feb. 1993 (Vol. 1908, pp. 173–187).

Pentland, A., Picard, R., & Sclaroff, S. (1994). Photobook: Content-based manipulation of image databases. In *Proceedings of the SPIE Storage and Retrieval for Image and Video Databases II*, San Jose, CA, Feb. 1994 (pp. 34–47).

Rui, Y., Huang, T. S., Ortega, M., & Mehrotra, S. (1998). Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology, 8*(5), 644–655.

Rui, Y., & Huang, T. S., (2000). Optimizing learning in image retrieval. In *proceedings of the IEEE conference on computer vision and pattern recognition*, South Carolina, June 2000 (Vol. 1, pp. 236–243).

Smith, J. R., & Chang, S. F., (1994). Transform features for texture classification and discrimination in large image databases. In *Proceedings of the IEEE international conference on image processing*, November 1994 (pp. 407–411).

Stricker, M., & Orengo, M. (1995). Similarity of color images. *Proceedings of the SPIE Storage and Retrieval for Image and Video Databases*, 381–392.

Su, Z., Zhang, H. Z., Li, S., & Ma, S. (2003). Relevance feedback in content-based image retrieval: Bayesian framework, feature subspaces, and progressive learning. *IEEE Transactions on Image Processing, 12*(8), 924–937.

Su, W. T., & Lien, J. J. (2006). Heuristic pre-clustering relevance feedback in region-based image retrieval. *Asian Conference on Computer Vision, 2*, 294–304.

Sun, Y., & Ozawa, S. (2003). Semantic-meaningful content-based image retrieval in wavelet domain. In *Proceedings of the 5th ACM SIGMM international workshop on multimedia information retrieval*, November 2003 (pp. 122–129).

Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics–sensitive integrated matching for picture libraries. *IEEE Transaction on Pattern Analysis and Machine Intelligence, 23*, 947–963.

Wang, J., Zha, H., & Cipolla, R. (2005). Combining interest points and edges for content-based image retrieval. In *Proceedings of the IEEE international conference on image processing*, Nov ember 2005 (Vol. 3, pp. 1256–1259).

Wong, W. T., Shin, F. Y., & Su, T. F. (2007). Shape-based image retrieval using two-level similarity measures. *International Journal of Pattern Recognition and Artificial Intelligence, 21*(6).

Wood, M. E., Campbell, N. W., & Thomas, B. T. (1998). Iterative refinement by relevance feedback in content-based digital image retrieval. In *Proceedings of the 5th ACM international conference on multimedia*, Bristol, UK, September 1998, (pp. 13–20).

Wu, Y., Tian, Q., & Huang, T. S., (2000). Discriminant EM algorithm with application to image retrieval. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, South Carolina, June 2000 (Vol. 1, pp. 222–227).

Yoo, H. W., Jung, S. H., Jang, D. S., & Na, Y. K. (2002). Extraction of major object features using vq clustering for content-based image retrieval. *Pattern Recognition, 35*, 1115–1126.

Yoo, H. W., Jung, S. H., Seo, K. K., & Lee, M. E. (2004). Retrieving images by comparing homogeneous color and texture objects in the image. *International Journal of Pattern Recognition and Artificial Intelligence, 18*(6), 1093–1110.

Yoshizawa, T., & Schweitzer, H. (2004). Long-term learning of semantic grouping from relevance feedback. In *Proceedings of the 6thACM SIGMM international workshop on multimedia information retrieval*, New York, NY, USA, October 2004, (pp. 165–172).

Zhou, X. S., & Huang, T. S., (2001). Small sample learning during multimedia retrieval using BiasMap. In *Proceedings of the IEEE international conference on computer vision and pattern recognition*, Hawaii, December 2001 (pp. 11–17).