適用於圖片搜尋系統之擴充式語意機制之研究

An Enhancement Semantic-Based Mechanism for Image Retrieval

Shi-Ming Huang
Department of Information Management
National Chung Cheng University
smhuang@mis.ccu.edu.tw

Chih-Fong Tsai
Department of AIT
National Chung Cheng University
actcft@ccu.edu.tw

Chia-Ming Chuang
Department of Information Management
National Chung Cheng University
tobe@mis.ccu.edu.tw

摘要

CBIR 最主要的限制是圖片的低階特徵(low-level feature)和人們高階概念(high-level concept)間的語意壕溝(semantic gap)的問題。語意壕溝發生的原因在於使用者偏向用高階的概念或語意進行查詢，而不是使用圖片的低階特徵來進行查詢。

在本研究中，我們提出一個擴充式語意機制(enhancement semantic-based mechanism)來解決此問題。此機制主要結合相關回饋(relevance feedback)和資訊過濾(information filtering)的概念。系統中的圖片是以分類好的圖庫網站為主要來源。另外，我們將色彩語意(如:可愛的、輕快的)視為一種高階的查詢概念，並運用在系統中。系統首先會根據使用者的年齡、職業、性別、興趣來進行分群並且會對使用者的查詢行為進行分析及學習。經由學習階段之後，新的查詢結果會根據使用者所屬的群集和之前使用者回饋來呈現。實驗結果顯示 ESBM 比傳統的關鍵字查詢能提供使用者更有效的查詢結果。

關鍵詞：圖片搜尋、語意壕溝、相關回饋、資訊過濾、色彩語意。

Abstract

One major limitation of content-based image retrieval is the semantic gap between low-level features of images and high-level concepts of human. This is because users usually prefer querying images by high-level concepts or semantics instead of low-level images features.

In this paper, we present an enhancement semantic-based mechanism (ESBM) to solve the problem. The mechanism is mainly based on combining the concept of relevance feedback and information filtering. Our system considers color semantics such as pretty, cheerful, etc. as one high-level conceptual query. In addition, images which have been classified into some conceptual categories are available for keyword-based queries. The system first of all learns users' queries and cluster users into four groups by their ages, occupations, interests and gender. After training, new query results will be based on the feedbacks and user’s groups. Experimental results show that ESBM is able to enhance retrieval effectiveness compared with traditional keyword-based image retrieval systems.

Keyword : image retrieval, semantic gap, relevance feedback, information filtering, color semantic.

一、Introduction

Image retrieval is the trend in the future. Many companies (such Google, Microsoft, Yahoo) have begun to offer users the services of image retrieval today. If they have an effective mechanism of image retrieval to help users, more and more users will use this service.

The performance of content-based image retrieval (CBIR) being unsatisfactory for many practical applications is mainly due to the gap between the high-level semantic concepts and the low-level visual features of images [4]. There are many researches addressing this problem. We can briefly classify two aspects. 1)The first one is focusing on index of the images. In [10], La Cascia and Sclaroff use an integrated approach, which is based on statistical couplings between the content of the document (latent semantic content) and the contents of images (visual statistics). Colombo et al. present a compositional approach increases the level of representation that can be automatically extracted and used in a visual information retrieval system [3]. Tsai et al. present a two-stage mapping model (TSMM) which is intended to minimize the semantic gap by reducing the recognition errors of classifying images [6].

2)Another aspect is to focus on user’s searching behavior. Related technologies include relevance feedback (RF), information filtering (IF) and so on [5][7][8][14]. kanade and Uchihashi propose a new approach to image retrieval that uses users’ feedbacks in the form of interpretation rather than through image features, thus directly exploiting the human perceptive power [12]. Zhou et al. present an approach for image retrieval using on-line analysis of feedback sequence log [13]. In [11], the author presents overlapped subspace clustering and multi-subspace label propagation algorithms. They apply these techniques into the traditional RF further improvement for image retrieval.

In this paper, we focus on the second aspect under the www environment. We assume that mining users’ search behaviors by using RF and IF could enhance image retrieval effectiveness of using keyword-based queries.

RF has been borrowed from textual information retrieval. Content-based image retrieval system can be enhanced by incorporating users’ feedbacks. Today, a lot of researches have proven RF is very useful for image retrieval.

However, there are three drawbacks as follow:

- Iteration : the procedure of RF usually needs a certain number of iterations to meet user’s satisfaction or achieve satisfactory retrieval effectiveness.
- No-willing : very few users are willing to go through huge iterations of feedback. Hence, the number of (ir)relevant images feedback by users could be very few.
- Imbalance of feedback images : in the RF procedure, the number of negative images (non-relevant images) may be larger than the number of positive images (relevant images) that the system needs many positive feedbacks to achieve users’ idea.

On the basis of above-mentioned, we regard RF as the behavior that users click images. In our paper, we will record the ID of those images and click count of those images as ‘implicit’ feedbacks, which avoids the ‘user in the loop’ procedure, as one major limitation of RF. Then the system will show images to new users according to these records.

The enormous growth of visual images makes users more choices, but also brings the problem of information overload. To solve the problem, several approaches have been developed to assist users in searching required images. A major of those is information filtering, which can be used to identify relevant information for each user. The information filtering technology can be classified into two categories: Content-based filtering and Collaborative filtering [1].

Content-based filtering is based on content analysis of the considered objects, e.g. keyword, keyword frequency, and the relationship between keyword and user's preferences. However, there is one drawback as follows:

- It is hard to analyze the content of multimedia data, e.g. voice, video and image. It is not easy to define the dimension of content.

Collaborative filtering is a technique to predict preferences of one person from preferences of other people [9]. In collaborative filtering, items selected for particular users refer to other users who have similar interests. Amazon.com is a representative example.

The system based on collaborative filtering can solve the problems of content-based filtering. Because it doesn't need to analyze the content of object, thus it may be suitable for image retrieval systems.

The rest of the paper is organized as follows. Section 2 describes the system architecture of the enhancement semantic-based mechanism (ESBM). In section 3, we will perform a real case for system evaluation. Experimental results are given in section 4. Finally, Section 5 concludes this paper.

二、System Architecture

The system architecture is illustrated in Figure 1. The system includes two phases of development. The first phase, image feature extraction, is responsible for extracting the low-level feature of the images including classification information, content feature (the type of file) and color feature. The second phase, user semantic extraction, is responsible for finding the relationship between semantics via users' queries and their images by ESBM. In addition, the system will store related content and color features and semantics of images in the index database.

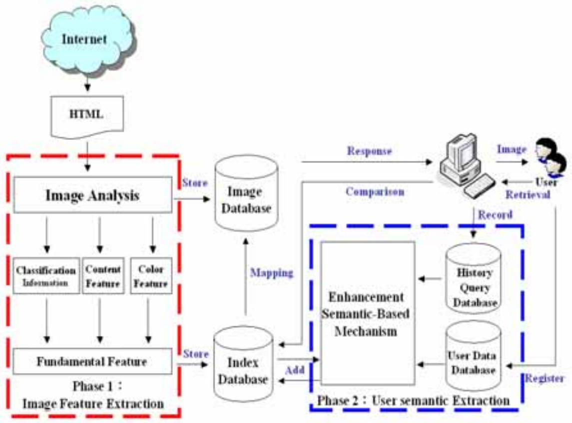


Figure 1. The system architecture

(一) Phase 1 : Image Feature Extraction

Prior to implement the image feature extraction phase, it is necessary to define images resources. We define classified images of a chosen website as our source (e.g. http://www.samsungmobile.com).

This phase consists of three steps (see Figure 2). The term of "designer-based " means the view of the website constructor.

Step 1 : Designer-based image classification information extraction
⇩
Step 2 : Designer-based image content feature extraction
⇩
Step 3: Designer-based image color feature extraction

Figure 2. Three steps of the phase 1

1.Step 1 : designer-based image classification information extraction.

This step is to extract the classification information of the images, according to the classified categories on the website.

2.Step 2 : designer-based image content feature extraction.

This step is to extract content feature of the images (including the type of file, width, height).

3.Step 3 : designer-based image color feature extraction.

This step is to extract the color feature of the images. In this step, we use Hue & Tone 120 color system (Figure 3). Hue includes red, yellow red, yellow, green yellow, green, blue green, blue, purple blue, purple and red purple. Tone includes vivid, strong, bright, pale, very pale, light grayish, light, grayish, dull, deep and dark.

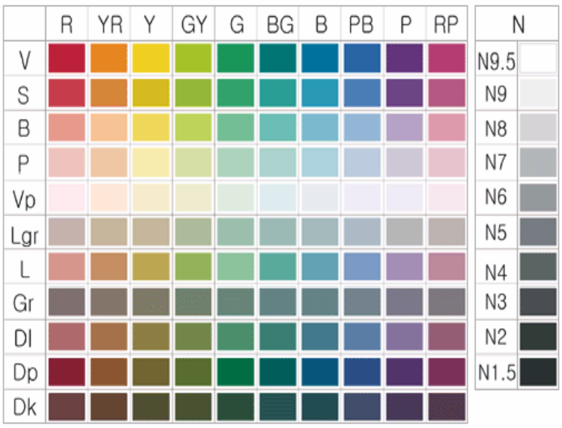


Figure 3. Hue & tone 120 color system

Evidence from the psychological literature seems to support that people find images easier to user particular base colors only [2]. In our paper, we will resolve images and distinguish the percentage of color according to Hue & Tone 120 color system.

Because the color system only includes 120 kinds of color, we have to choose the nearest color from them. In this paper, we adopt Euclidean distance (formula 1) to fine the nearest color of each pixel of one image.

$$dij = \sqrt{(Ri - Rj)^2 + (Gi - Gj)^2 + (Bi - Bj)^2} \quad --(1)$$

Ri, Gi, Bi The RGB value of the number *i* of pixel of one image.

Rj, Gj, Bj The RGB value of the number *j* of color of Hue & Tone 120 Color System.

dij The Euclidean distance between number *i* of pixel and the number *j* of color.

First, we calculate all distances between 120 colors and RGB-value of one pixel. Then, we choose the color which has the shortest distance from Hue & Tone 120 system and record the percentage of the color. The procedure will repeat until all pixels of the image have been processed. Finally, we will record the related color information of the images in the index database.

After the phase 1, we can obtain the fundamental feature of the images (including classification information, content feature and color feature).

(二) Phase 2 : User semantic extraction

This phase is responsible for discovering the relationship between images and their semantics from users' aspect via ESBM.

First, ESBM will produce enhancement semantic according to the percentage of the color information of the images (e.g. pretty, cheerful, and casual). Second, ESBM will use some data mining technology (i.e. clustering) to analyze users' behavior. Finally, ESBM will revise and store the relationship between images and their semantics based on users' feedbacks for improving subsequent retrieval results.

In this phase, user semantic extraction, ESBM will include three steps (Figure 4). The term of user-based means the view of the users.

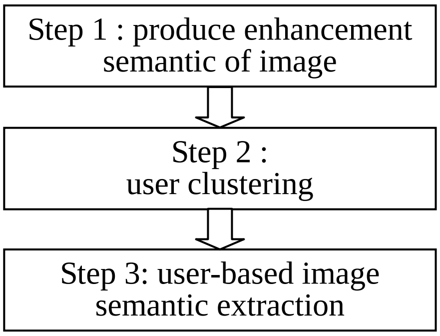
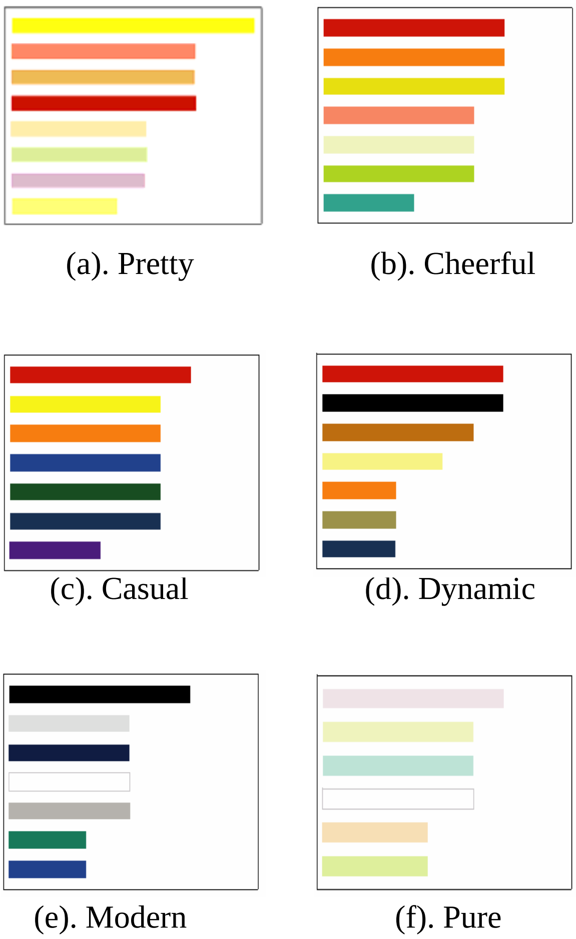


Figure 4. Three steps of ESBM

1.Step 1 : produce enhancement semantic of image.

The step is to produce enhancement semantic of images according to the percentage of the color information of the images. First, we will get related color information of images from the index database. Then, we will compare it with fourteen color combinations which are defined by I.R.I (Figure 5).



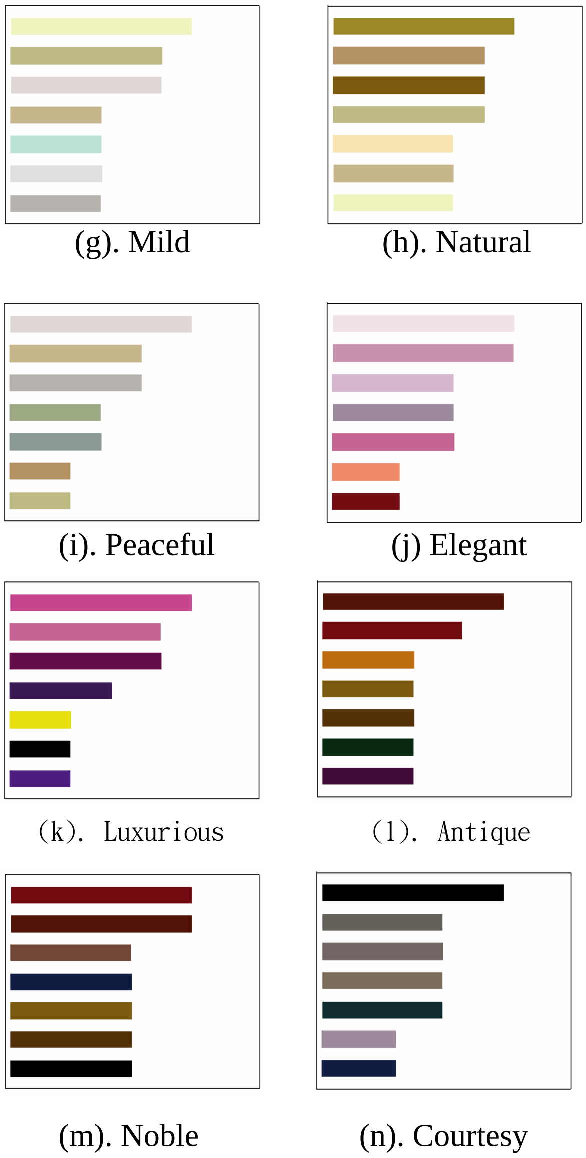


Figure 5. color semantics

Finally, we can produce the enhancement semantic of the images for semantic retrieval. (e.g. pretty, cheerful, casual, dynamic, modern, pure , mild, natural, peaceful, elegant, luxurious, antique, noble and courtesy.)

2.Step 2 : user clustering.

The step is mainly to adopt the concept of collaborative filtering described in section 1. We will divide users into different groups and analyze users' behavior of different groups including semantic-based query results and clicks of downloading some particular images from the users. In our paper, we will divide users via two-stage clustering approach. We regard personal data as the standard of clustering. The personal data includes gender, age, occupation, and interest.

3.Step 3 : user-based image semantic extraction.

The step is to discover the relationship between images and their semantics based on different clusters. In the step, we will revise the relationship between them according to users' behaviors (i.e. feedback).

In our paper, we propose the formula 2 for ranking the images.

$$RS(i) = \alpha * CSi + (1-\alpha) * CLi \text{------(2)}$$

RS(i)	The ranking score of image i
α	Weight value, $0 < \alpha < 1$
CSi	The ratio of color similarity of image i
CLi	The ratio of click count of image i

RS(i) represents the ranking score of image i. The query result will be arranged by it. 'α' represents the weight value normalized between 0 and 1.

CS is abbreviated from color similarity. CSi represents the ratio that users agree the enhancement semantic of image i and download image i. CSi is calculated by formula 3.

$$\frac{\text{The amount of users agree the enhacement sematic of image } i}{\text{The amount of all sers}} \text{--(3)}$$

CL is abbreviated from click. CLi represents that the ratio of click count of image i. CLi is calculated by formula 4.

$$CLi = \frac{\text{The amount of click count of image } i}{\text{The amount of all clicks}} \text{-----(4)}$$

For example, there are four users and the semantic of image 1 is 'pretty'. Now, we want to calculate the ranking score of image 1. We can divide it into two parts for further calculation. First, we calculate the color similarity (CS) of image 1. Given the amount of users is four. The amount of users that agree the enhancement semantic of image 1 and download it is two. CS1 is calculated as follows.

$$CS1 = \frac{2}{4} = 0.5$$

Second we calculate the CL1. Given the amount of all clicks by all users is fifty. The amount of click count of image 1 is two. CL1 is calculated as follows.

$$CL1 = \frac{2}{50} = 0.04$$

If we assume the weigh value of 'α' is 0.5. So, RS(1) is calculated as follows.

$$\begin{aligned} RS(1) &= \alpha * CS1 + (1-\alpha) * CL1 \\ &= 0.5 * 0.5 + (1-0.5) * 0.04 \\ &= 0.27 \end{aligned}$$

After we collect all related information of users' behavior (i.e. feedbacks), we will re-calculate the ranking score (RS) of the images when there are new users' feedbacks. The new ranking score will be stored in the database. Next time when (new) users query images, the query result will be based on the newest RS, which is calculated before this query.

三、System Evaluation

(一) Scenario

Here, we take Samsung as our case because image search and download is very important for phone companies. Figure 6 shows the website. The website consists of 15 categories, which contains 592 images.

Figure 6. Samsung website

Samsung is the largest mobile-phone company in Korea. According to Garnter (2004), Samsung Electronics held the No. 2 position with a market share of 17.1% in the world market following Nokia's 30.8%.

(二) Experimental process

There are two stages in our experiment. In the stage 1, thirty users manipulated the system. We calculated the RS of the images based on users' behavior (relevance feedback). In the stage 2, another set of thirty users manipulated the system again. The images were arranged by the RS calculated from stage 1. Then we record the users' behavior of stage 2. The performance will be illustrated in section 4.

1.Stage 1 : learning stage.

In this stage, the first set of thirty users must register their personal information (Figure 7). Then we divide thirty users into four clusters based on their personal information (gender, age, occupation and interest) via two-stage clustering approach (Table 1). The field of cluster of the userdata table means the user belong to which cluster.

Figure 7. Member data

Table 1. Partial record of userdata table

U_id	Password	gender	age	occupation	interest	cluster
axlu	******	Female	24	Student	4	1
eatbd	****	Female	28	Student	2	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
skj	******	Male	25	Student	4	4

After clustering, users first need to choose one enhancement semantic (e.g. pretty) from the color semantic list and press the "search "button on the left side (Figure 8). The matched images will be showed on the right side. Then users need to click any images they think match the enhancement semantic they choose (e.g. pretty).

There are fourteen semantics on our system. Each of the thirty users must repeat the procedure fourteen times. The related click information will be recorded on the querydata table (Table 2).

Figure 8. Experimental screen

Table 2. Partial record of query data table

U_id	P_id	semantic	time
axlu	8	Pretty	200563014112
eatbd	33	Pretty	200563014112
⋮	⋮	⋮	⋮
johnny	8	Courtesy	200563014115
ggi	67	Courtesy	200563014115

In this stage, the thirty users have been divided into four clusters. We will calculate the RS of each image according to related querydata of different clusters. That is, there are four RS values per image based on the four clusters. In the next stage, the images will be re-arranged by the RS as new users are involved during image retrieval.

2.Stage 2 : testing stage.

In the stage 2, there are another thirty users to test the system. They also need to register their personal information. In the stage 1, the first thirty users have been divided into four clusters. In this stage, we use the centroids of the clusters of stage 1 as the starting point for clustering of stage 2. Therefore, the thirty users of stage 2 will be divided into one of the four clusters produced by stage 1.

When the user chooses the enhancement semantic, the retrieved images of a particular query will be arranged by RS calculated from stage 1. Here, we set the threshold of RS is 0.25. That is, the RS of those images which are retrieved are all exceed 0.25.

The procedure of the test is the same as stage 1. First, users choose the enhancement semantic. Then users click the image they think it conform to the semantic.

四、Experimental Results

There, we use accuracy to evaluate our performance based on the thirty users of stage 2. In other words, the first set of the thirty users can be thought of as performing 'initial' relevance feedbacks to 'train' the system. We will calculate the accuracy of each cluster. The retrieval accuracy is defined as formula 5

$$Accuracy = \frac{Total\ images\ clicked}{Total\ images\ retrieved} \text{---------(5)}$$

As shown in figure 9 when the threshold of RS increases, the avg. accuracy of the four clusters will also increase. On the other hand, when the RS of one image is higher than others, this image will be arranged in front. It means the image has a higher probability to be clicked as relevance by users. It also means the image semantics are more relevant to users' opinions.

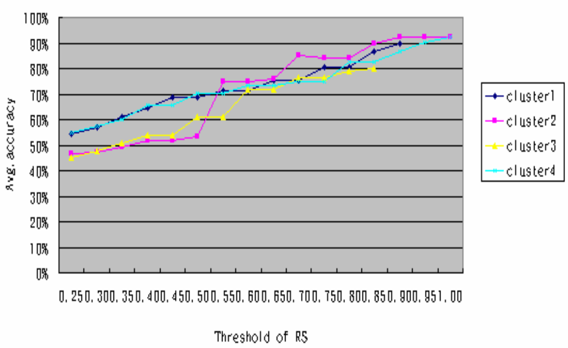


Figure 9. Threshold of rs vs. avg.accuracy

Figure 10 shows the avg. accuracy of the pretty category via ESBM and traditional image retrieval system (by keyword/without ESBM). When we use keyword (pretty) to search, we can get 70 images. There are 237 images belonging to the pretty category via ESBM. We set the threshold of RS of images to be 0.25. So there are 131 images exceed 0.25. On the stage 2 of experiment, we asked users to click the images conform to ‘pretty’ via ESBM and keyword respectively. We use cluster 1 as an example for a comparison.

There are 58 images that the users regard them as relevant to the semantic of ‘pretty’ in 70 images. There are 94 images that users identify in 131 images. Then 49 images are repeated between 58 and 94 images. So we can search 103 images (58+94-49=103) which conform to ‘pretty’.

On the other hand, the avg. accuracy of cluster 1 with ESBM is 91.26% (94/103 * 100%=91.26%) and the avg. accuracy of cluster 1 without ESBM is 56.31% (58/103*100%=56.31%). We can retrieve more relevant images via ESBM than keyword-based image retrieval.

However, a shortcoming of our mechanism is that there are only 49 images out of the 58 images (49/58*100%=84.48%) can be retrieved via ESBM. That is, we can’t use ESBM alone to find all relevant images which are retrieved by keywords.

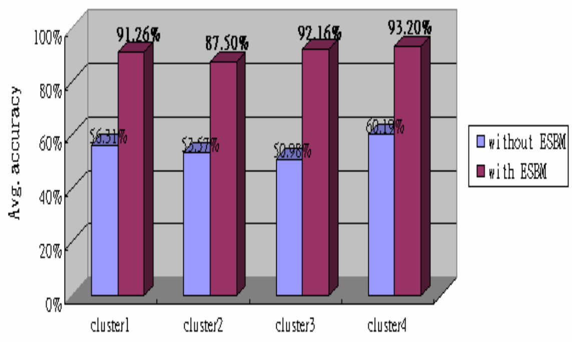


Figure 10. Avg.accuracy of ESBM vs without ESBM

Figure 11 shows that the count of feedback person is higher, the avg. accuracy will also be higher. When the amount of feedback person of the cluster increases in the stage 1, the avg. accuracy of each cluster in the stag 2 will also increases. It proves the relevance feedback is very helpful for accuracy of image retrieval. However, as one major limitation of relevance feedback is the requirement of a certain number of feedbacks to reach satisfactory retrieval effectiveness, reducing the count of the feedback person to achieve similar retrieval accuracy as this experimental result is not the focus of this research.

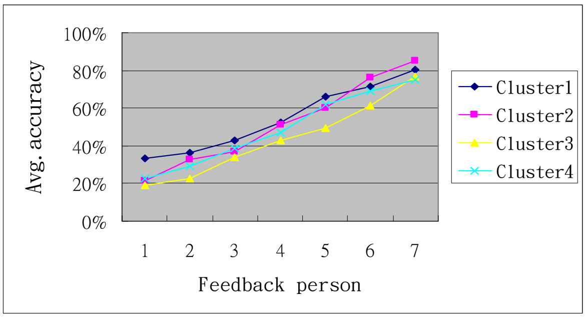


Figure 11. Feedback iterations vs. avg.accuracy

According to the above experimental results, we believe that our mechanism is able to enhance image retrieval effectiveness. Our enhancement semantic-based mechanism (ESBM) produces the enhancement semantic of images by their color. Furthermore, we combine the relevance feedback (users’ feedbacks) that is very useful for image retrieval and collaborative filtering (clustering) which can be used to identify relevant information for each user.

That is, when the enhancement semantic of images was produced by color, the images will go through the procedure of relevance feedback and collaborative filtering. On the next time, the system can offer the relevant images that the previous users clicked to new users. On the other hand, our mechanism can solve the problem of semantic gap effectively.

五、Conclusion

According to the above experimental results, we believe that our mechanism is able to enhance image retrieval effectiveness. Our enhancement semantic-based mechanism (ESBM) produces the enhancement semantic of images by their color. Furthermore, we combine the relevance feedback (users' feedbacks) that is very useful for image retrieval and collaborative filtering (clustering) which can be used to identify relevant information for each user.

That is, when the enhancement semantic of images was produced by color, the images will go through the procedure of relevance feedback and collaborative filtering. On the next time, the system can offer the relevant images that the previous users clicked to new users. On the other hand, our mechanism can solve the problem of semantic gap effectively.

六、References

[1] Basu C., Hirch H.,and Cohen W.W. "Recommendation as classification : Using social and content-based information in recommendation", the Fifteenth national conference on artificial intelligencent AAAI/IAAI, pp.714-720, 1998.

[2] B.M. Mehtre, M.S. Kankanhalli, A.D. Narasimhalu, and G.C. Man, "Color Matching for Image Retrieval," Patten Recognition Letters, Vol. 16, pp. 325-331, 1995.

[3] C.Colombo ,A. Del Bimbo, and P.Pala,"Semantics in visual information retrieval",IEEE Multimedia, Vol. 6, pp.38-53, July-Sept.1999.

[4] C.Dorai and S.Venkatesh, "Bridging the semantic gap with computational media aesthetics", IEEE Multimedia, Vol. 10, pp.15-17, 2003.

[5] Chen Jiewei, Xu Weiran, Guo Jun, "An approach to intelligent information filtering in Chinese document images based on garbage model", International Symposium on Intelligent Multimedia, Video and Speech , pp.198 – 201, Oct. 2004.

[6] Chih-Fong Tsai,Ken McGarry, John Tait, " Using Neuro-Fuzzy techniques Based on a Two-Stage Mapping Model for Concept-Based Image Database Indexing ",IEEE International Symposium on Multimedia Software engineering, pp.6-12, Dec. 2003.

[7] Feng Jing; Mingjing Li, Hong-Jiang Zhang, Bo Zhang, "Relevance feedback in region-based image retrieval", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, pp.672 – 681, May 2004.

[8] Junhua Chen, Zhonghua Yang, "A learning multi-agent system for personalized information filtering", the Fourth Pacific Rim Conference on Multimedia, Vol. 3, pp.1864 – 1868, 2003.

[9] McnNee, S. M., Lam,S.K.,Konstan, J.A., Riedl., J., "Interfaces for Eliciting New User Preferences in Recommender Systems", The 9th international Conference On User Modeling, pp.178-188, 2003.

[10] M.La Cascia, S.Sethi, S.Sclaroff, "Combining textual and visual cues for content-based image retrieval on the World Wide Web", IEEE Workshop on Content-Based Access of Image and Video Libraries, pp.24-28, 21 June 1998.

[11] Tao Qin, Tie-Yan Liu, Xu-Dong Zhang, Wei-Ying Ma and Hong-Jiang Zhang, "Subspace Clustering and Label Propagation for Active Feedback in Image Retrieval" , Multimedia Modelling Conference, pp.72 – 179, Jan. 2005.

[12] T. Kanade and S. Uchihashi, "User-Powered Content-Free Approach to Image Retrieval", Proceedings of International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society, pp. 24 - 32 , March, 2004.

[13] Xiangdong Zhou, Qi Zhang, Li Liu, Liang Zhang and Baile Shi, "An image

retrieval method based on analysis of feedback sequence log**”,** Pattern Recognition Letters Vol. 24, pp.2499 – 2508, Oct. 2003.

[14] Yong Rui, T.S. Huang, M.Ortega, S.Mehrotra, “Relevance feedback: a power tool for interactive content-based image retrieval”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, pp.644 – 655, Sept. 1998.