A NOVEL IMAGE RETRIEVAL

Jianfeng REN[1], Yuli SHEN[2] and Lei GUO[3]
[1,3]Department of Automatic Control,
North Western Polytechnic University, P.R. CHINA
[2] Department of Electrical Engineering,
Zhan Jiang Ocean University, P.R.CHINA
rjjff@163.com[1] lguo@nwpu.edu.cn[3]

ABSTRACT
In this paper, we present results of a project that seeks to transform the low level features to high level meaning. Firstly we extract the low level features called as the representative colors from the images, and then we present a new approach called WordNet to establish the link from the low level feature vectors to the semantics. In order to improve the retrieval efficiency, the relevance feedback is also applied into our system. Experimental results show that our method is promising.

KEY WORDS
Representative colors, wordnet, relevance feedback.

1 INTRODUCTION

Indexing diverse collections of multimedia data remains a challenging problem. Most current approaches to image retrieval mainly focused on two aspects: one is the visual features [1,2,3,4], such as the color, texture and shape, the other is the distance metrics [5,6,7].

However, in the visual feature approaches, the interfaces supplied to the user are non-intuitive and unnatural. The user searching for visual data usually has some idea of the image content and the image layout of the desired image, where the image content is described in terms of objects and global features rather than low-level features such as the color and texture. However, the image objects can be correctly and clearly expressed through keywords, which have powerful query abilities.

All the images from the database in our experiment are nature images, such as rose, blue sky and so on. Through human observation, these objects have obvious color characteristics. For example, rose appears red. This paper is conducted to explore such links between the colors and nature objects, which are expressed through keywords.

Even though significant progress has been made toward developing effective content-based descriptors, such as the standard descriptors by MEPG-7 [8], there is a difficulty to narrow the gap between the low-level features in image analysis and image understanding at the semantic level. Because people analyze, understand and classify the image content according to its semantic features.

Due to the importance of semantics, some approaches have been provided to bridge the gap between the low-level features and semantic level features. Shi-F Chang [9] proposed the novel idea of Semantic Visual Templates (SVT) to narrow this gap. Each template represents a personalized view of concepts (e.g. slalom, meetings, sunsets etc). The SVT is represented using a set of successful queries, which are generated by a two-way interaction between the user and the system. Aleksandra Mojsiovic and Bernice Rogowitz [10] proposed a method for semantic categorization and retrieval of photographic images based on low-level image descriptors. In this method, they first used multidimensional scaling and hierarchical cluster analysis (HCA) to model the semantic categories into human observers organizing images.

Through a series of psychophysical experiments and analyses, they refined the definition of these semantic categories, and used these results to discover a set of low-level image features. The main drawback of this method is that we need to do a lot of psychophysical experiments. Wiiam I.Grosky and Rong Zhao [11] presented the techniques, latent semantics indexing (SCI), to negotiate the gap. First, a corpus is formed of documents (in this case, images with a caption) from which features are computed. Then by singular value decomposition (SVD), the dictionary covering the captions is correlated with features derived from the pictures. However these methods seem to be inefficient due to the size of image collections.

In order to overcome the shortfalls mentioned above, in this paper, we propose a new approach based on the color semantics to narrow the gap between the low-level and semantic level, considering the intuitive characteristics of color. Firstly, the low-level features based on the color are presented. Then we use the WordNet in order to narrow the gap between the low-level feature and semantic level features through the training sample images and human interactions. In order to more efficiently retrieve images, relevance feedback has been adopted in our strategy. Experimental results demonstrate that this method could correctly retrieve images not only through the specified images but also keywords in a given domain.

This paper is organized as followed: we introduce the overview of our method in section2. In section 3, the low-level features based on the representative colors are developed. In section 4, we will present the WordNet to bridge the gap between low level features and high-level features. We will provide the different query strategies in section 5. In section 6, the relevance feedback is applied into our strategy. In section 7, a detailed analysis based on the experimental results will be presented. Finally, Conclusions and remarks are given in section 8

2 OVERVIEW

Our method mainly includes the following steps:

Step1: image decomposed into five regions14. Because there are no good methods to segment different objects in the images at present, we mainly apply the simple method provided in [14] to segment the images. After segmented, each region should include one object as possible as it could.

Step2: computing the frequency of eight domain color components in each region.

Step3: establishing the WordNet from the color features to semantics through the training samples and human interactions

Step4: re-define the query feature from the similar images using the relevance feedback technique until the user satisfy the results main goal of *pre-processing* is to prepare raster cartographic images in such a way as to simplify them and increase the reliability of their recognition in the automatic system.

3 COLOR FEATURE EXTRACTION

Color is perhaps the most expressive of all the visual features and has been extensively studied in image retrieval during the last decade. Some low-level features are developed by using the color histograms[2,12,13]. This type of representation tends to produce false positives. Color histogram give statistics of the image pixels but do not provide spatial, relational or content information in terms of objects in the scene. To improve performance, spatial information was incorporated into color-based image retrieval system by allowing multiple color histograms, representing different locations in the image [2,12,13].

In this paper, we provide eight representative colors in a region of interest. The reason that we employ this method is as following:

1. It is provided mainly based on the observation that a small number of colors are usually sufficient to characterize the color information in image region.
2. Representative color approach can overcome the drawback of the traditional histogram. The number of bins in a typical color histogram range from few tens to a few hundreds. The high dimensionality of the feature vectors result in high computational cost in distance calculation for similarity retrieval, and inefficiency in indexing and search.
3. It also considers the spatial information and relationships between the different regions in the images.

The process of our method is as follows: Firstly the image can be decomposed into five fixed regions14, which is shown in figure 1, and then representative color histogram is computed in each region.

Figure 1. Image decomposition

3.1 COLOR SPACE

The different color spaces used in MPEG-7 include the familiar monochrome, RGB, HSV, YCrCb, and the new HMMD. Although many approaches are mainly based on the HSV and YCrCb rather than RGB due to the fact these methods can improve the perceptual uniformity [15], the RGB color format is the most common color format for digital images. The primary reason for this is because it retains compatibility with computer displays.

3.2 REPRESENTATIVE COLORS

According to human vision perception, a small number of image colors are sufficient to express the color information. Therefore, in our method, we regard eight colors as the representative colors, which are mapped into corresponding points in the RGB space, which are shown in table 1.

Red	(255,0,0)
Green	(0,255,0)
Blue	(0,0,255)
Yellow	(255,255,0)
Magenta	(255,0,255)
Cyan	(0,255,255)
Black	(0,0,0)
White	(255,255,255)

Table 1. Relationships between the representative color and its corresponding points

3.3 COLOR CLUSTERING

In order to compute the percentage of representative colors in the region, the colors in the image region should be clustered based on the nearest neighbor algorithm.

Now we assume eight representative colors $\{C_1,C_2,C_3,C_4,C_5,C_6,C_7,C_8\}$ as the clustering center, and $\# C_i$ denotes the total number of pixels which are grouped into the clustering center Ci , and $\# C$ denotes the total number of pixels in the image, so $\frac{\# C_i}{\# C}$ can be regarded as the percentage that representative color account for in the image. For example, in this paper, $\frac{\# C_1}{\# C}$ denotes that the percentage of the red component in the image.

Because the image are decomposed into five fixed regions such as the left-up, right-up, left-down, right-down and center, we should express the spatial relationships between the different regions in the feature vectors. Here, let expresses the corresponding region in the image. For example, in a given region , after color clustering $[\frac{\#C_1}{\#C},\frac{\#C_2}{\#C},\frac{\#C_3}{\#C},\frac{\#C_4}{\#C},\frac{\#C_5}{\#C},\frac{\#C_6}{\#C},\frac{\#C_7}{\#C},\frac{\#C_8}{\#C},s5]$ is obtained

Now we have obtained the representative color features in the corresponding regions. However, these feature vectors could not be directly recognized by humans, just because people understand image at the semantic level. So a new approach called WordNet .is provided to solve this problem.

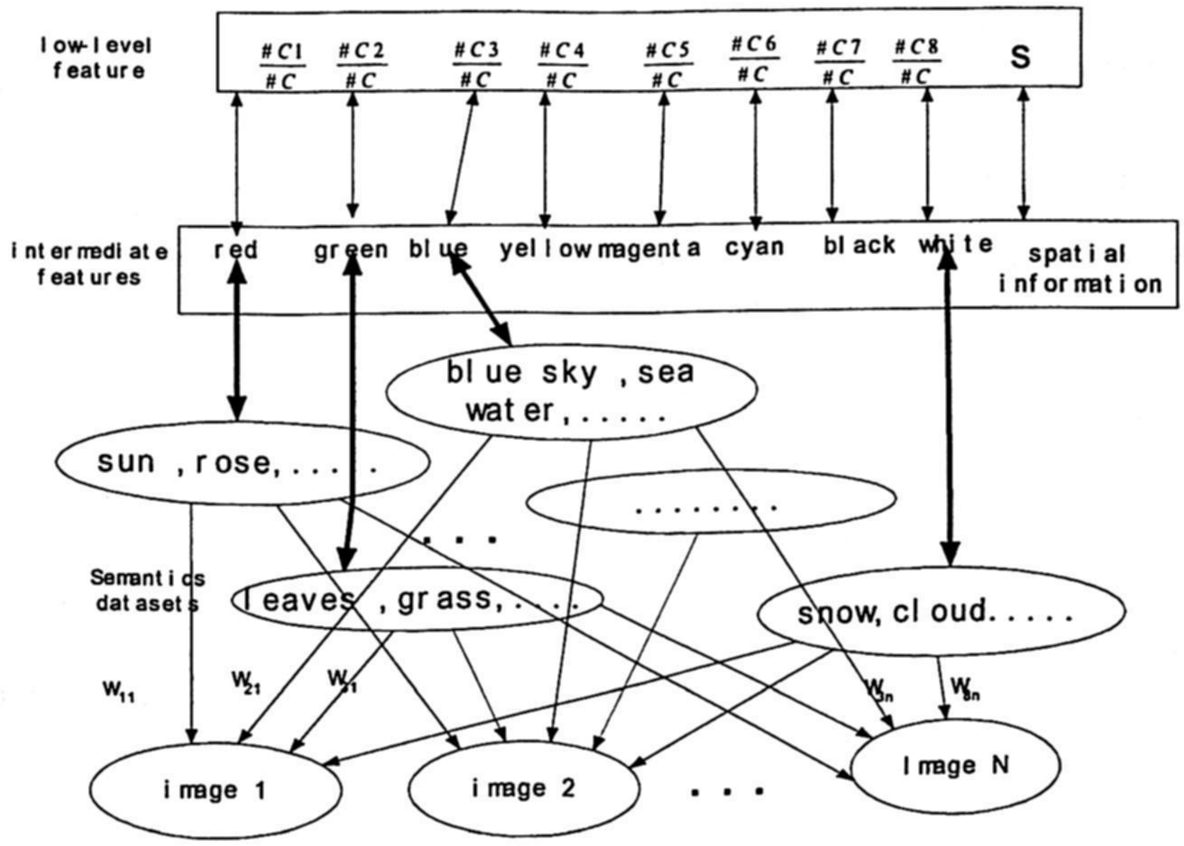

Figure 2. Link between representative colors and semantic keywords

4 ESTABLISH THE LINKS

In our paper, all the images are from the scene images, which have obvious color characters. We can easily find the link between the representative color and nature object. For example blue can express the meaning of blue sky or sea water. A detailed link between the color and the corresponding keywords is shown in figure 2

Through the above links, we can easily narrow the gaps between the low-level and semantic level features. And more, we also see that the intermediate features play an essential role in bridging the gap between them. Such links are stored in the database as the look-up table. In addition, the semantic datasets can be expanded through human interactions and relevance feedback, in other words, the user can add and delete the keywords in the semantic datasets.

For a give region R5 located in the center of the image, the specified process from the low-level to semantic level is expressed as following:

If #C1 in the feature vector is maximum
Then the object can be {rose, sun....}
If #C3 in the feature vector is maximum
Then the object can be {see water, blue sky....}

Through the above algorithm, each region can be indexed one representative color, which has a set of keywords associated. Accordingly, an image that is decomposed into five fixed regions will have not more than five sets of associated keywords, because the different regions may have the same representative color.

The weight associated on each link of the keywords with the image represents the degree of relevance in which the keyword describes the linked image's semantic content.

From above discussed, we see that each image object is associated with three different level features respectively at the low-level feature, intermediate level feature and semantic level feature.

5. QUERY STRATEGY

From Figure2, an image can be expressed at three levels: low-level features, intermediate level and semantic level. Accordingly, there are three different modes of user interactions involved in typical retrieval systems.

Firstly, we can index image and retrieval image from the semantic level. In this case, the user types in a list of keywords representing the semantic contents of the desired image. For instance, for the image shown in the figure 1, we can list our keywords: we want to find cloud in the center and the blue sky in the surrounding. In this way, we not only describe the semantic content but also spatial information in the images.

Secondly, at the intermediate level, we can use the combinations of the representative colors to describe the image. For the image shown in figure1, we can describe it as following: white in the center and blue in the surrounding.

Thirdly, we also can retrieval the images based on the low-level features which are obtained by using the algorithm provided in section 3.

6. RELEVANCE FEEDBACK

Recently, relevance feedback based CBIR techniques [16,17] merged as a promising research direction. The central idea of the relevance feedback is that it does not require a user to provide accurate initial queries, but rather estimate the user's ideal initial query by using positive and negative examples feedback by user. The fundamental goal of this method is to estimate the ideal query parameters accurately and robustly.

All the approaches [16,17] perform relevance feedback at the low-level feature vector level, but failed to take account into the actual semantics for the image databases. The inherent problem with these approaches is that the low-level features are often not as powerful in representing complete semantic content of images as keywords in representing text documents.

In our paper, we provide relevance feedback not only at the low-level feature level but also at the semantic level. At the low-level feature level, we try to improve the estimate of the" ideal query point" by moving it towards good examples point and away from bad examples points. Here, we update the estimates by using the Rocchio's formula given below (1) for sets relevant documents D'_R and non-relevant documents D'_N given by the user [18].

$$Q' = \alpha Q + \beta\left(\frac{1}{N_{R'}}\sum_{i\in D'_R} D_i\right) - \gamma\left(\frac{1}{N_{N'}}\sum_{i\in D'_N} D_i\right) \quad (1)$$

However, semantic based relevance feedback can be performed relatively easily compared to low-level feature counterpart. The basic idea behind it is a simple scheme to update the weights W_{ij} associated with each link shown in Figure 2. And the weight updating process is described as below:

1. Initialize all weights W_{ij} to 1. That is, every keyword has the same importance.
2. Collect the user query and the positive and negative feedback examples
3. For each keyword in the input query, check to see if any of them is not the keyword database. If so, add them into the database without creating any links
4. For each positive example, check to see if query keyword is not linked to it. If so, create a link with weight 1 from each missing keyword to this image. For all other keywords that already to this image, increment the weight by 1.
5. For each negative example, check to see if any query keyword is linked with it. If so, set the new weight $W'_{ij} = W_{ij}/2$. If the weight W_{ij} on any link is less than 1, delete that link.

7. EXPERIMENTAL RESULT

We have implemented the presented algorithms in the image retrieval system designed by us. In this image retrieval system, it can support three modes of query which have been previously discussed in Section5. All the experimental evaluations were implemented under Microsoft Windows 2000. The machine we used is a Pentium-3650M PC with 126MB DRAM main memory. The software is Matlab 6.0.

We download about 10000 scene images in the BMP format from the website. Every representative color may have more than 1000 scene images. In order to guarantee that nature objects have obvious colors, these images mainly are the combinations of the following nature that nature objects have obvious colors, these images mainly are the combinations of the following nature objects such as flowers, mountain, blue sky and so on. According to its semantic content of the image and the corresponding representative colors, each representative color can be associated with a few sets of keywords. For example, the keywords such as blue sky and sea waters may be classified into the representative color "blue". Through the algorithms provided in section 3 and section 4, each image and associated keywords are put into the database. Figure 3 show the main interface of our image retrieval system.

The experimental research is concerned primarily with the retrieval time and the accuracy of the retrieved images.

The first experiment is conducted to evaluate the retrieval time. Table 2 shows the retrieval time at three different levels.

Query level	Retrieval time (s)
Query by Example image	132.13
Query by intermediate color	0.300
Query by keywords	0.270

Table 2 the retrieval time

From the above table, we can see that query by keywords can improve the retrieval speed greatly compared with the query by example image.

The second experiment is implemented in order to evaluate the retrieval accuracy based on the relevance feedback. A retrieved image is considered as a relevant one if it has the similar representative color to the query image. The retrieval accuracy is defined as:

$$R = \frac{\text{relevant images retrieved in top N returns}}{\text{N}} \qquad (2)$$

In our experiment, N is usually set to 20. Thus, the retrieval accuracy is denoted by R20.

From the database, four random images are selected as querying the whole image database at the semantic level and low-level feature respectively. Figure 4 shows the accuracy based on the relevance feedback at the semantic level, and figure 5 shows the accuracy based on the relevance feedback at the low level feature.

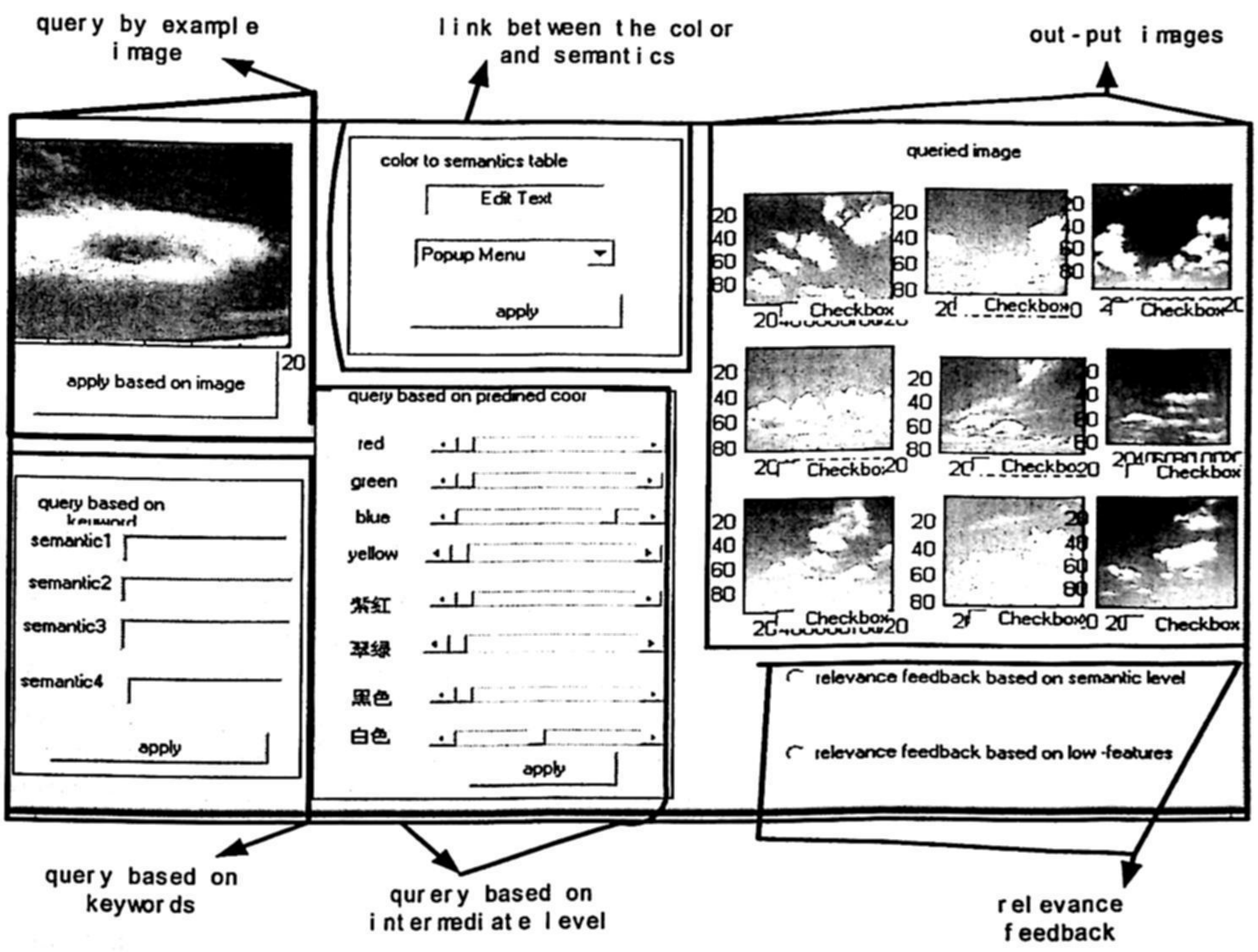

Figure 3: System Interface

As we can see from the results, our system achieves the high accuracy (>80%) after a few of relevance feedback for any given query. Unlike other methods where more user relevance feedback [17,18] may lead to lower the retrieval accuracy, experimental results demonstrate our method can be stable.

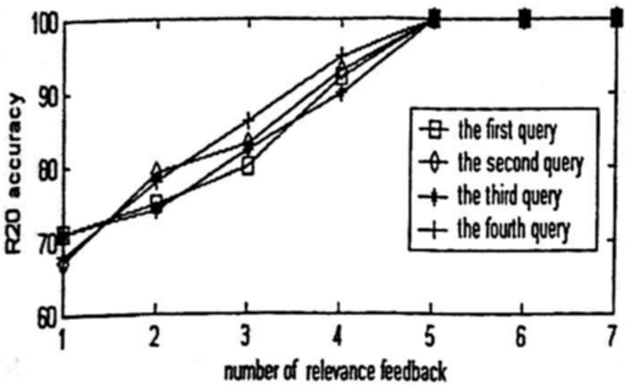

Figure 4. Accuracy based on the relevance feedback at the semantic level

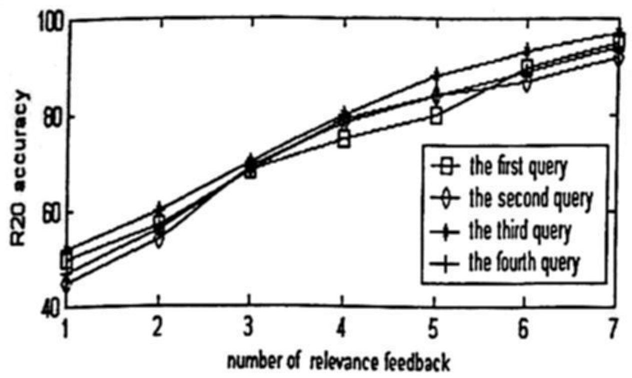

Figure 5. Accuracy based on relevance feedback at the low level

In addition to verifying the effectiveness of our proposed method, we also compared our method against the other techniques. We have chosen to compare our method with the retrieval technique used in [19]. The comparison is made through four random queries based on relevance feedback. Figure 6 shows the comparisons with the method in[19].

It is easily seen from results, our methods based on relevance feedback will improve the retrieval accuracy substantially.

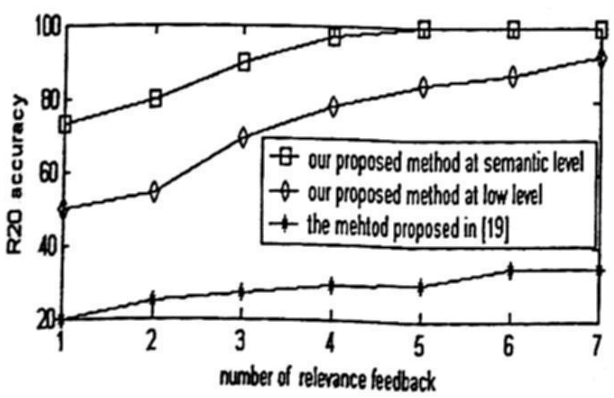

Figure 6. Performance comparison

8. CONCLUSIONS

In this paper, we provide the representative colors for indexing the images. Our method not only captures the color content of the images but also characterizes the spatial information of color in the image. In addition, we present a new approach called the WordNet to narrow the gap between the low-level features from the representative colors and semantic content of the images.

Finally, in order to improve the retrieval performance, we applied the relevance feedback to retrieve images both at the semantic level and at low-level feature. Experimental results show this method is stable. In the future work, we will further narrow the gap from the shape descriptor to semantics after we have obtained the representative colors.

REFERENCES

[1] M. Flicker, H, Sawhney, and W. Niblack, "query bay image and video: The QBIC system," IEEE Computer, 28, 1995.

[2] M, Swain and D. Ballard, " Color indexing" International Journal of Computer Vision,7,pp11-32,1991.

[3] C.C. Gotlieb and H.E. Kreyszig, "Texture descriptors based on co-occurrence martrices," Computer Vision, Graphics, and Image Processing, 51,1990.

[4] J.Z. Wang et al, "Content-based image indexing and searching using Daubechies's wavelets," Int J. Digital libraries, 1, pp 311-328,1998.

[5] S. Santini and R. Jain, "Similarity measures," IEEE Trans. On PAMI, 9,1999.

[6] A, Natses, R. Rasto, "Walrus: a similarity retrieval algorithm for image database," SIGMOD Record, 28, 1999

[7] D, Bimbo and P, Pala, "Visual image retrieval by elastic matching of user sketches," IEEE Trans on PAMI, 19, 1997.

[8] Text of ISO/IEC 15938-3 Multimedia Content Description Interface-Part 3: Visual, Final Committee Draft, ISO/IEC/JTC1/SC29/WG11, Doc, N4062, Mar, 2001.

[9] S.-F. Chang, W. Chen, and H. Sundaram, "Semantic Visual Templates: linking visual features to semantics Semantic Visual Templates: Linking Visual Features to Semantics," Proceedings 1998 International Conference on Image Processing (ICIP'98), Chicago, Illinois, October 4-7, 1998.

[10] Mojsilovic, and B. Rogowitz, "Capturing image semantics with low-level descriptors", Proc. Int. Conference on Image Processing, ICIP 2001, Thessaloniki, Greece, September 2001.

[11] Rong Zhao and William I.Grosky, "Narrowing the Semantic Gap---Improved Text-Based Web Document Retrieval Using Visual Features," IEEE Transaction on Multimedia, Vol,4, No.2, pp189-200, 2002

[12] W. Niblack, R.Barber, W.Equitz, M.Flickner, E. Glasman, D. Petkovic, P. Yanker, C.Faloutsos, and G.Taubin. "The Q BIC project: Querying images by content, using color, texture, and shape.". In SPIE Conference on Storage and Retrieval for Image and Video Databases. Volume 1908, pp 173-187,1993.
[13] T, Kato et al. "A sketch retrieval method for full color image database-query by visual example." In proceedings, 11th IAMPR International Conference on Pattern Recognition, pp530-533, 199
[14] Markus Stricker, Alexander Dimai," Color Indexing with Weak Spatial Constraints" SPIE Conference, Feb, 96,San jose.
[15] G.wyszecki and W. S. Stiles,Color Science: Concepts and Methods, John Wiley & Sons,1982.
16] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in Proc. IEEE Int. Conf. On Image Proc., 1997.
[17] Y. Ishikawa, R. Subramanya, and C.Faloutsos, "Mindread: Query Databases through multiple examples" in Proc. Of the 24th VIDB Conference,(New York),1998.
[18] RUI,Y, huang ,T.S, and Mehrotra, S "content-based image retrieval with relevance feedback in MARS," in proc. IEEE Int Conference. On image proceeding, 1997.
[19] Rui,Y., Huang, T.S. "A Novel Relevance Feedback Techniques in Image Retrieval", ACM Multimedia,1999.