Colour image retrieval based on primitives of colour moments

J.-L. Shih and L.-H. Chen

Abstract: A colour image retrieval method based on the primitives of colour moments is proposed. First, an image is divided into several blocks. Then, the colour moments of all blocks are extracted and clustered into several classes. The mean moments of each class are considered as a primitive of the image. All primitives are used as features. Since two different images may have different numbers of features, a new similarity measure is then proposed. To demonstrate the effectiveness of the proposed method, two test databases from Corel are used to compare the performances of the proposed method with other existing ones. The experimental results show that the proposed method is usually better than the others. Furthermore, since for a few special types of images, other methods may have better results occasionally, a relevance feedback algorithm is also provided to automatically determine the best method according to the user's response.

1 Introduction

The recent emergence of multimedia and the availability of large images have made content-based information retrieval an important research topic. The most frequently cited visual contents for image retrieval are colour, texture, and shape. Among them, the colour feature is most commonly used. It is robust to complex background and independent of image size and orientation.

The colour histogram [1] is the best known colour feature and is used by the QBIC system [2]. It is invariant to rotation, translation and scaling. To take into account similarities between similar but not identical colours, the QBIC system introduced the quadratic distance to measure the similarity between two histograms.

To overcome the quantisation effects of the colour histogram, Stricker and Orengo [3] used colour moments as feature vectors for image retrieval. Since any colour distribution can be characterised by its moments, and most information is concentrated in the low-order moments, only the first moment (mean), the second moment (variance) and the third moment (skewness) are taken as features. The method is not appropriate for retrieving partially similar images.

For fast image retrieval on large image databases, the WebSEEk system [4] proposed use of the colour set. The colour set can be obtained from a colour histogram by thresholding the colour histogram. Each image is first mapped onto a colour set in HSV colour space and then segmented into several regions. The spatial relationships among these regions are used for querying similar images. Since similar images may have different segmentation results, this will affect the retrieval results.

To reduce the influence of improper segmentation, the Blobworld system [5], which was developed as part of the digital library project of the University of California at Berkeley, transforms raw image data into a small set of image regions, called blobs. Each blob, which roughly corresponds to an object or a part of an object, is coherent in colour and texture. For specific images, such as the sunset, a zebra, or the sky, proper blobs can be extracted. However, it is hard to find representative blobs for most colour images.

Stricker and Dimai [6] proposed the concept of 'pseudo-segmentation'. For each image, they defined five partial overlapping fuzzy regions: centre, top-left, top-right, bottom-left and bottom-right. The method has the disadvantage that if an object exists in the centre of a query image, any image containing a similar object not located in the centre will not be retrieved.

The colour histogram, colour moments and colour set contain only colour information of each pixel in an image, the local relationship among neighbouring pixels is not involved. Huang *et al.* [7] proposed another kind of feature, called 'colour correlograms', which expresses the spatial correlation of pairs of colour changes with distance. However, it is sensitive to object size in an image.

The MPEG-7 visual standard proposed four kinds of colour descriptors [8–10], including scalable colour, dominant colour [11], colour layout and colour structure, for colour image retrieval. The scalable colour descriptor (SCD) is defined in the HSV colour space with fixed colour space quantisation and uses a novel Haar transform encoding. The dominant colour descriptor gives the distribution of the salient colours in a given region or in an image. Unlike the fixed colour quantisation in SCD, colours in a given region are clustered into a small number of dominant colours. These two colour descriptors provide a compact descriptor that is easy to index.

IEE Proceedings online no. 20020614

DOI: 10.1049/ip-vis:20020614

Paper first received 16th July 2001 and in revised form 3rd May 2002

J.-L. Shih was with the Department of Computer and Information Science, National Chiao Tung University and is now with the Department of Computer Science and Information Engineering, Chung Hua University, No. 707, Sec. 2, WuFu Rd., Hsinchu, Taiwan 300, Republic of China

L.-H. Chen is with the Department of Computer and Information Science, National Chiao Tung University, 1001 Ta Hsueh Rd., Hsinchu, Taiwan 30050, Republic of China

However, they contain only the colour information of each pixel in an image. The colour layout descriptor (CLD) can capture the spatial layout of the dominant colours on a grid superimposed on an image. First, the representative colour for each one of an 8×8 grid is extracted, then a DCT is applied. The resulting coefficients are finally encoded. The CLD is efficient only for sketch-based image retrieval. The colour structure descriptor (CSD) expresses local colour structures in an image. It records the occurrence of each particular colour in each local structure. Thus, two images are considered to be similar only when most of the local colour structures in an image are similar to those in the other image. This means that two images with only partial parts similar cannot be considered to be similar.

Besides the colour features, texture features, such as water-filling [12] and wavelet moments [13], are well known features for image retrieval. The water-filling algorithm operates on the edge map of the image as a 'flooding of connected canal systems'. The filling time, the number of edge pixels, the number of simple edge loops and the number of edge branches are extracted as features. However, if the edge map does not correspond to the perceptual meaning or edges are not well-defined, water-filling performs worse than wavelet moments. For wavelet moments, each input image is first fed into a wavelet filter bank and decomposed into 10 subbands. The mean and standard deviation of the wavelet coefficients in each subband are extracted to form 20 feature components. Although texture feature concerns the detailed context, it is suitable only for images full of regular textures. Thus, for a natural image, if users take only the texture features to effect retrieval, a good result will not be expected. Thus in the MARS system [14], a feedback algorithm [15] is used to automatically adjust the weights among colour, texture and shape features interactively with users. However, in image retrieval, sometimes using only the most appropriate feature vector can achieve better retrieval results than combining several feature vectors.

In general, there are two kinds of search strategies. One is to find images totally similar to the query one. The other is to find images partly similar to the query one. For example, if a query image is a rose and some roses are in a garden image, then the garden image will be considered as different from the query image under the first search strategy, but as similar under the second strategy. Methods using the second search strategy can meet the aims of the first search strategy but not vice versa. All of the methods using the above-mentioned features take the first search strategy.

In this paper, a region-similarity colour image retrieval method using the second search strategy is proposed to solve the disadvantages of the first search strategy. First, an image is segmented into several blocks. Then, the colour moments of all blocks are extracted and clustered into several classes. The mean vector of each class is considered as a primitive of the image. All primitives are used as feature vectors. Then, a similarity measure is proposed to perform colour image retrieval. To show the effectiveness of the proposed method, comparisons between the proposed method and other existing methods are provided. The experimental results show that the proposed method is superior to other existing methods for most kinds of image. For a few special kinds of images, some other methods will have better results. To obtain the best result, we also implement other methods to form a colour image retrieval system. In this system, a relevance feedback algorithm is used to automatically determine the best method according to the user's response.

2 Extraction of primitives of colour moments

In this Section, we first define the primitives of colour moments. For colour images, the YIQ colour model is used in this paper. Because a probability distribution is uniquely characterised by its moments [3], the colour distributions of the Y, I and Q components of an image can be represented by its colour moments. The first colour moment of the ith colour component ($i = 1, 2, 3$) is defined by

$$M_i^1 = \frac{1}{N}\sum_{j=1}^{N} p_{i,j}$$

where $p_{i,j}$ is the colour value of the ith colour component of the jth image pixel and N is the total number of pixels in the image. The hth moment, $h = 2, 3, \ldots$, of the ith colour component is then defined as

$$M_i^h = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{i,j} - M_i^1)^h\right)^{1/h}$$

Take the first H moments of each colour component in an image s to form a feature vector, $\boldsymbol{CT}$, which is defined as

$$\begin{aligned}\boldsymbol{CT} &= [ct_1, ct_2, \ldots, ct_Z] \\ &= [\alpha_1 M_1^1, \alpha_1 M_1^2, \ldots, \alpha_1 M_1^H, \alpha_2 M_2^1, \alpha_2 M_2^2, \ldots, \\ &\quad \alpha_2 M_2^H, \alpha_3 M_3^1, \alpha_3 M_3^2, \ldots, \alpha_3 M_3^H]\end{aligned}$$

where $Z = H \cdot 3$ and $\alpha_1, \alpha_2, \alpha_3$ are the weights for the Y, I, Q components. Based on the above definition, an image is first divided into X non-overlapping blocks. For each block a, its hth colour moment of the ith colour component is defined by $M_{a,i}^h$. Then, the feature vector, $\boldsymbol{CB}_a$, of block a is represented as

$$\begin{aligned}\boldsymbol{CB}_a &= [cb_{a,1}, cb_{a,2}, \ldots, cb_{a,Z}] \\ &= [\alpha_1 M_{a,1}^1, \alpha_1 M_{a,1}^2, \ldots, \alpha_1 M_{a,1}^H, \alpha_2 M_{a,2}^1, \alpha_2 M_{a,2}^2, \ldots, \\ &\quad \alpha_2 M_{a,2}^H, \alpha_3 M_{a,3}^1, \alpha_3 M_{a,3}^2, \ldots, \alpha_3 M_{a,3}^H].\end{aligned}$$

From the above definition we can obtain X feature vectors. However, there are many similar $\boldsymbol{CB}_a$ s among these feature vectors. To speed up the image retrieval, we will find some representative feature vectors to stand for these feature vectors. To reach this aim, a progressive constructive clustering algorithm [16] is used to classify all $\boldsymbol{CB}_a$ s into several clusters and the central vector of each cluster is regarded as a representative vector and called the primitive of the image. The central vector, $\boldsymbol{PC}_k$, of the kth cluster is defined by

$$\begin{aligned}\boldsymbol{PC}_k &= [pc_{k,1}, pc_{k,2}, \ldots, pc_{k,Z}] = \frac{\sum_{j=1}^{n_k} \boldsymbol{CB}_j^k}{n_k} \\ &= \left[\frac{\sum_{j=1}^{n_k} cb_{j,1}^k}{n_k}, \frac{\sum_{j=1}^{n_k} cb_{j,2}^k}{n_k}, \ldots, \frac{\sum_{j=1}^{n_k} cb_{j,Z}^k}{n_k}\right] \quad (1)\end{aligned}$$

where $\boldsymbol{CB}_j^k$, $j = 1, 2, \ldots, n_k$, belongs to the kth cluster and n_k is the size of the kth cluster. Note that in the algorithm, a predetermined distance threshold T_d is used as the maximum radius of each cluster. Since it is fixed for all images, the number of clusters varies for different images. To treat this situation, a method to evaluate the similarity between two images with different numbers of feature vectors is proposed.

3 Colour image retrieval

In this Section, a similarity measure between two images with various numbers of primitives is provided first. Then, since each kind of colour feature is especially appropriate for certain types of image, to obtain an optimal retrieval result, a relevance feedback algorithm is proposed to automatically determine the most appropriate feature among the popular colour features, including the proposed one, according to the user's response.

3.1 Similarity measure

Before introducing the similarity measure, we first provide several definitions. The kth primitive of a query image q is represented as: $\boldsymbol{PC}_k^q = [pc_{k,1}^q, pc_{k,2}^q, \ldots, pc_{k,Z}^q]$, where $k = 1, 2, \ldots, m$, and m is the number of primitives in the query image. The λth primitive of a matching image s is denoted as $\boldsymbol{PC}_\lambda^s = [pc_{\lambda,1}^s, pc_{\lambda,2}^s, \ldots, pc_{\lambda,Z}^s]$. The distance between $\boldsymbol{PC}_k^q$ and $\boldsymbol{PC}_\lambda^s$ is defined as follows:

$$D\_PC_{k,\lambda}^{q,s} = \sqrt{\sum_{i=1}^{Z}(pc_{k,i}^q - pc_{\lambda,i}^s)^2}$$

The minimum distance between PC_k^q and all primitives of s is defined by

$$D\_PC_k^{q,s} = \min_\lambda(D\_PC_{k,\lambda}^{q,s})$$

The distance between the query image q and the matching image s is defined by

$$D\_PC^{q,s} = \sum_{k=1}^{m} n_k^q \times D\_PC_k^{q,s}$$

where n_k^q is the size of the kth cluster. The similarity measure between q and s is defined as

$$Sim^{q,s} = \frac{1}{D\_PC^{q,s}}$$

Note that the larger $Sim^{q,s}$ a matching image has, the more similar it is to the query image. Based on the measure, we can find images similar to the query one by taking those with high values.

3.2 Relevance feedback algorithm

In general, the proposed method will, for most images, have the best retrieval result. For a few special types of image, other methods may give better results. This point will be shown in the next Section. To treat this situation, a system containing the proposed method and other methods as features is proposed and a relevance feedback algorithm is provided to determine the best method according to the user's response. First, for a query image, using the proposed method, g images with the highest similarity measures are retrieved. Then their grades, $G\_PC_{q,s}$, are defined as $g, g-1, g-2, \ldots$, and 1, respectively. The image with the highest similarity measure will have the highest grade. In addition, the grade of each non-retrieved image is defined as zero. We also apply the remaining six methods, colour moments of the whole image, colour correlograms, dominant colour, colour layout, colour structure and scalable colour, to evaluate the corresponding grades denoted as $G\_CT_{q,s}$, $G\_Cor_{q,s}$, $G\_Dom_{q,s}$, $G\_Lay_{q,s}$, $G\_Str_{q,s}$ and $G\_Sca_{q,s}$, respectively. For each matching image s, the summation of grades is defined as

$$\begin{aligned} Grade_{q,s} = {} & \omega_1 * G\_PC_{q,s} + \omega_2 * G\_CT_{q,s} \\ & + \omega_3 * G\_Cor_{q,s} + \omega_4 * G\_Dom_{q,s} \\ & + \omega_5 * G\_Lay_{q,s} + \omega_6 * G\_Str_{q,s} \\ & + \omega_7 * G\_Sca_{q,s} \end{aligned}$$

where $\omega_1, \omega_2, \ldots, \omega_7$ are the weights. Initially by $\omega_1, \omega_2, \ldots, \omega_7$ are set to 1. Based on these grades, a group of images similar to the query one can be retrieved. Note that the seven weights will affect the retrieval results, and it is impossible to determine a set of fixed weights that is appropriate for any kind of image. Thus, we propose a simple relevance feedback algorithm through the user's response to automatically determine the weights. A user can choose r similar images, $I_1, I_2, \ldots, I_r$, from the query results. Based on the grades of these images, the new ω_1 is calculated as

$$\omega_1 = \sum_{j=1}^{r} G\_PC_{q,I_j}$$

$\omega_2, \omega_3, \ldots, \omega_7$ can also be modified. Since using only the most appropriate feature vector will usually get better retrieval results than using combined features, the best feature is selected instead of using combined features. This can be done by modifying ω_i as

$$\omega_i = \begin{cases} 1 & \text{if } \omega_i = \max\{\omega_1, \omega_2, \ldots, \omega_7\}; \\ 0 & \text{otherwise} \end{cases}$$

Using these new weights, a user can obtain a better retrieval result.

4 Experimental results

To evaluate the performance of the proposed method, experiments have been conducted based on the Corel photo library, which is often used by image retrieval research groups [7, 12]. There are two test databases, D1 and D2, selected from Corel in our experiments. D1 is a small and well-classified test database. Based on D1, we implemented other methods using colour moments [3], colour set [4], colour correlograms [7] or the four MPEG-7 colour descriptors [8–10], scalable colour, dominant colour [11], colour layout and colour structure, as features to compare their performances with the proposed scheme.

Zhou *et al.* provided their retrieval results using water-filling [12] or wavelet moments [13] as features on a large test database also obtained from Corel. Thus, we selected images from Corel to establish a large database D2 with the same size as that used in Zhou's experiment. Using D2, we applied the proposed method to compare the performances with water-filling and wavelet moments, without implementing their methods. Finally, we also compared the performance of the proposed scheme with those using colour moment, colour set, colour correlograms, and the four MPEG-7 colour descriptors, on D2.

4.1 Experimental results on small database D1

The small database D1 has 1300 images. These images were selected from 13 Corel photo classes, including the flower, stained glass, women, sunset, sports car, sailboat, ancient architecture, dinosaur, duck, waterfall, painting, underwater world, and gong fu: 100 images were selected randomly from each class. Fig. 1 shows several example

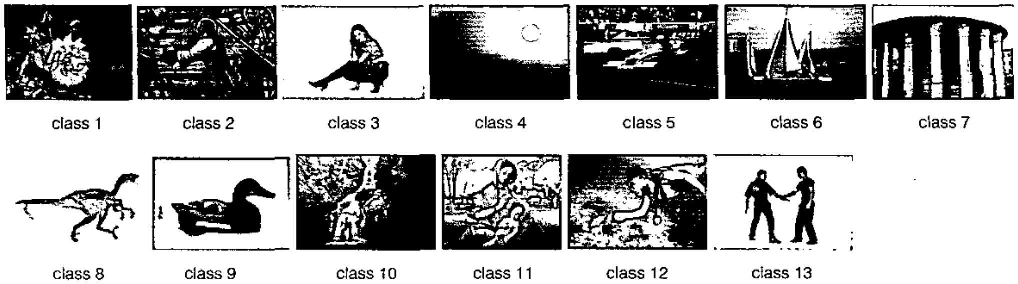

Fig. 1 *Example images of each class from database D1*

images for each class from D1. To demonstrate the effectiveness of the proposed methods on different kinds of colour images, D1 is constructed to contain several kinds of colour images. Some images have smooth colour layout, such as sunset. Some images are full of textures, such as stained glass. Some images have a special colour histogram, such as ancient architecture, waterfall, painting and underwater world. Others have obvious objects in images, such as the flower, woman, sports car, sailboat, dinosaur, duck, and gong fu. For the duck and dinosaur, an object is located in the centre of the image. However, for women and gong fu, the women and men have variable postures, positions, sizes and numbers.

The performance is measured by the 'recall' and 'precision' [17]. Note that the recall, Re, and precision, Pr, are defined by the following equations:

$$Re = \frac{N}{T} \text{ and } Pr = \frac{N}{K}$$

where N is the number of relevant images retrieved, T is the total number of relevant images and K is the total number of retrieved images.

In the experiments, each image is divided into 100 blocks. α_1, α_2 and α_3 are set to 0.4, 1, and 1, respectively. H is set as 4. T_d is set as 15. These values are used because they give the best results.

To show the performance of the proposed method on database D1, the retrieval results are compared with those using colour moments, colour set, colour correlograms, and the four MPEG-7 colour descriptors. For the colour moments, the first three colour moments of the whole image are extracted in RGB colour space. For the colour set, each image is first mapped onto a colour histogram with 166 colours (18 hues, three saturations, three values and four grey levels) in HSV colour space. Then the colour set is obtained from the colour histogram by thresholding the colour histogram. In the colour correlogram, all colours are first quantised into 64 colours in RGB colour space. Then the autocorrelograms are computed using four distances (1, 3, 5 and 7). Similarity is measured by the L1 norm (sum of absolute differences), except the colour set. For the colour set, the similarity is measured by the quadratic distance.

The four MPEG-7 colour descriptors are implemented according to the MPEG-7 standard [8–10]. It is worth mentioning that, in the MPEG-7 experiment, the scalable colour with the highest retrieval efficiency is the 256-D colour histogram in HSV colour space (16 hues, four saturations and four values). Hence, we compare with the 256-D colour histogram.

As shown in Fig. 2, the proposed method is better than other methods. A detailed comparison of precision for each class is shown in Fig. 3, and again the proposed method is in the top three ranks among all methods for each class.

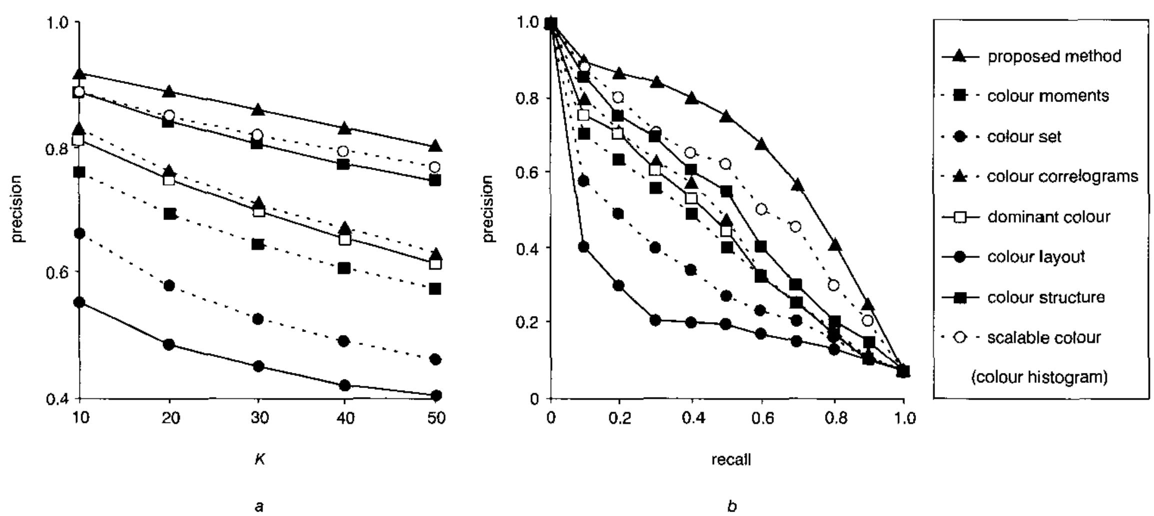

Fig. 2 *Performance comparison between the proposed method and other methods on D1*
a Precision curves
b Precision against recall curves ($T = 100$)

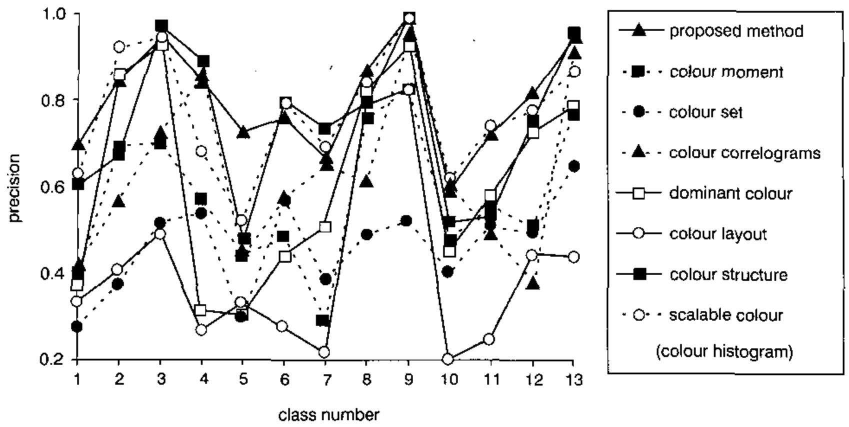

Fig. 3 *Precision comparison for each class from D1 among the proposed method and other methods ($K = 50$)*

Colour moments and colour set have the worst performances for most classes. Colour correlograms have bad performance for underwater world images (class 12). This is because the diver, coral and fish in these images have variable sizes and colours, and these variations are sensitive to the colour correlograms. For the MPEG7 colour layout, the performance is not good, except for the dinosaur (class 8) and duck (class 9) images. Each image from these two classes contains an object at the image centre. This shows that for two similar images, if both contain a similar object at different locations, these two images will be considered to be different based on colour layout. Colour structure performs worse than the proposed method for images with complex colour layout and texture, such as flower (class 1), stained glass (class 2), sports car (class 5), dinosaur (class 8), waterfall (class 10), painting (class 11) and underwater world (class 12). Finally, the proposed method performs much better than scalable colour on the sports car (class 5), since the cars are located in different backgrounds in D1 and the scalable colour cannot retrieve images with similar objects appearing in different backgrounds.

4.2 Experimental results on large database D2

Water-filling [12] and wavelet moments [13] have been used to carryout colour image retrieval on a large database consisting of 17 695 images from Corel with 400 aeroplanes and 100 American eagles. To make an objective comparison, we used a test database, D2, from Corel that also has 17 695 images with 400 aeroplanes and 100 eagles. For D2, we took the first 17 195 images from Corel according to Corel's own serial numbers plus 400 aeroplanes and 100 eagles. That is, the D2 images are varied and not classified.

Table 1 shows a comparison, in terms of average number of relevant images that are retrieved, by taking 100 aeroplanes and 100 eagles as query images on D2. Besides water-filling and wavelet moments, we also use colour moments, colour set, colour correlograms, and the four MPEG-7 colour descriptors as features to compare precisions. Table 1 shows that the proposed method is superior to those using water-filling and wavelet moments. However, for aeroplanes, this method is worse than methods using colour moments and colour structure. For eagles, the proposed method is worse than methods using colour moments, colour correlograms, colour structure and scalable colour. This is reasonable, since each image of the aeroplanes and eagles (Fig. 4) has a similar sky background occupying a large area, and the proposed method takes a partial matching strategy; sometimes the two kinds of images will be considered similar.

Since it is impossible to find a colour feature that is suitable for every kind of image, a relevance feedback algorithm is provided to determine the most appropriate feature among these colour features, according to the user's response. In these experiments, we chose 10 relevant images from the top 50 retrieved images and query again. As shown in Table 1, after using the relevance feedback algorithm, each Pr approaches the best Pr among the results obtained when using each colour feature individually.

Table 1: Average number of relevant images retrieved by taking 100 aeroplanes and 100 eagles as query images from D2

	Aeroplanes Pr ($K=20$)	Pr ($K=80$)	Eagles Pr ($K=20$)	Pr ($K=80$)
Proposed method	0.487	0.376	0.273	0.128
Water-filling	0.315	0.225	0.167	0.085
Wavelet moments	0.288	0.213	0.141	0.082
Colour moment	0.537	0.464	0.328	0.203
Colour set	0.323	0.235	0.236	0.171
Colour correlograms	0.416	0.249	0.441	0.245
Dominant colour	0.388	0.276	0.186	0.074
Colour layout	0.433	0.393	0.135	0.068
Colour structure	0.606	0.460	0.529	0.369
Scalable colour (colour histogram)	0.492	0.353	0.32	0.184
Combination with relevance feedback	0.563	0.401	0.457	0.266

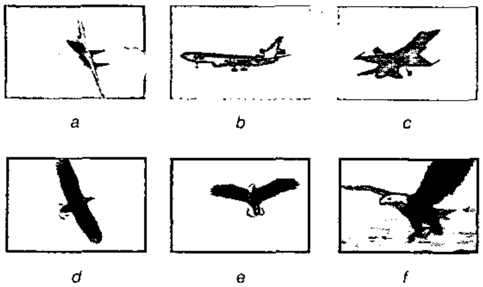

Fig. 4 *Example images of aeroplane and eagle classes from D2*
a–c Aeroplane images
d–f Eagle images

To compare the retrieval results for various images, we randomly selected 40 classes and five images for each class from D2 as query images. These classes included the tiger, boat, underwater world, sunset, eagle, lion, penguin, elephant, horse, flower, mountain, swim pool, fox, waterfall, goat, cat, dog, architecture, pyramid, duck, monkey etc. Since it is very time-consuming to count the total number of relevant images of each class in the large database in advance, we just compare the precision (Fig. 5). Fig. 5 shows that the proposed method is superior to other methods on images from the large database. Moreover, the relevance feedback algorithm is effective, and can help reach a better result. Finally, the retrieval results (see Fig. 6) for an elephant image are given to show

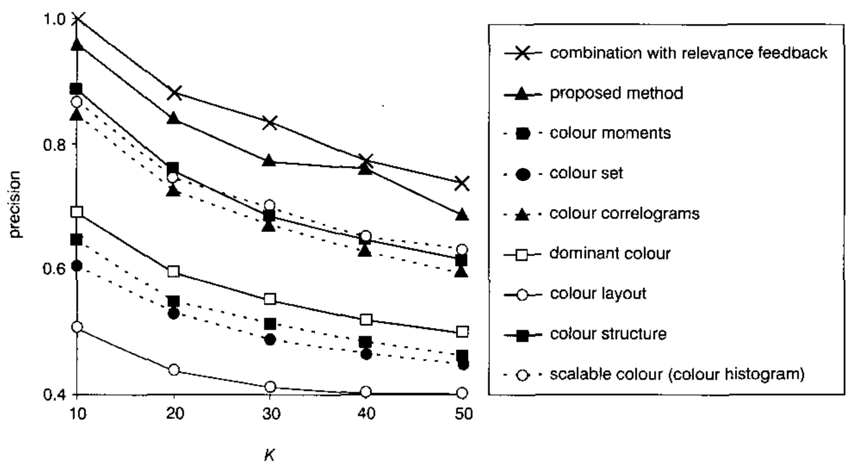

Fig. 5 *Comparison of precision of the proposed method, those using other colour features and a combination with relevance feedback, on D2 images*

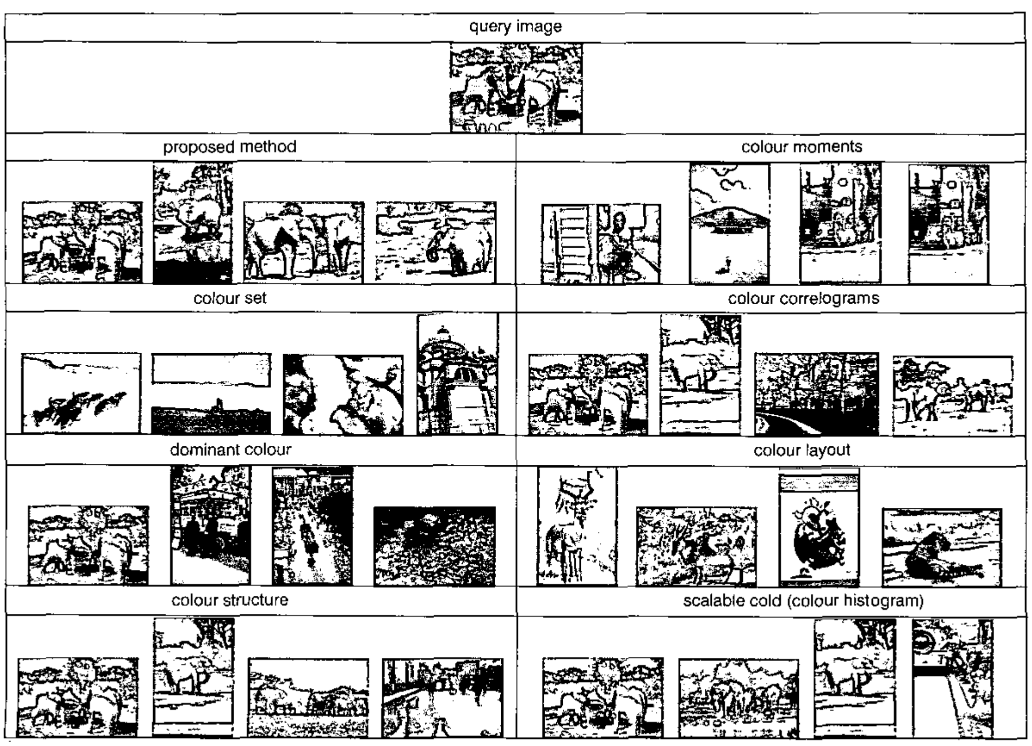

Fig. 6 *Retrieval results for D2 database for an elephant query image*

the characteristics of the different methods. Fig. 6 shows that the proposed method has the greatest ability to find those images containing elephants without restricting the position and size of the elephants. Note that, if the colour moments of Stricker and Dimai [6] are used as features, only those images with elephants at the image centre can be retrieved.

5 Conclusions

A new colour image retrieval method based on primitives of colour moments is proposed. First, an image is divided into several blocks. Then, the colour moments of all blocks are extracted and clustered into several classes based on a fast non-iterative clustering algorithm. The mean vector of each class is considered as a primitive of the image. All primitives are used as feature vectors. Then, a specially designed similarity measure is used to perform colour image retrieval. The proposed method, unlike other methods, contains the detail colour information of each important part in an image. Comparison with other methods reveals that for most types of image, the proposed method outperforms other methods using colour set, colour moments, colour correlograms, water-filling, wavelet moments, and the four MPEG-7 colour descriptors, as features. Since each feature will be the most suitable for a particular kind of image, to utilise this phenomenon, a colour image retrieval system is also designed. It includes the proposed method and others using the above-mentioned features. In this system, to meet the preferences of users, a relevance feedback algorithm is proposed to automatically determine the most appropriate feature, according to the user's response. The proposed system can be used as part of a digital library for content-based image retrieval (CBIR).

6 Acknowledgment

This research was supported in part by the National Science Council, ROC, under Contract NSC 89-2213-E-009-148.

7 References

1 SWAIN, M., and BALLARD, D.: 'Color indexing', *Int. J. Comput. Vis.*, 1991, **7**, (1), pp. 11–32
2 FLICKNER, M., SAWHNEY, H., NIBLACK, W., ASHLEY, J., HUANG, Q., DOM, B., GORKANI, M., HAFNER, J., LEE, D., PETKOVIC, D., STEELE, D., and YANKER, P.: 'Query by image and video content: The QBIC system', *IEEE Computer*, 1995, **28**, (9), pp. 23–32
3 STRICKER, M., and ORENGO, M.: 'Similarity of color images', *Proc. SPIE - Int. Soc. Opt. Eng.*, 1995, **2420**, pp. 381–392
4 SMITH, J.R., and CHANG, S.F.: 'Visually searching the web for content', *IEEE Trans. Multimed.*, 1997, **4**, (3), pp. 12–20
5 CARSON, C., BELONGIE, S., GREENSPAN, H., and MALIK, J.: 'Region-based image querying'. Proceedings of IEEE workshop on Content-based access of image and video libraries, in conjunction with CVPR'97, 1997, pp. 762–768
6 STRICKER, M., and DIMAI, A.: 'Color indexing with weak spatial constraints', *Proc. SPIE - Int. Soc. Opt. Eng.*, 1996, **2670**, pp. 29–40
7 HUANG, J., KUMAR, S.K., MITRA, M., ZHU, W., and ZABIH, R.: 'Image indexing using color correlograms'. Proceedings of CVPR Int. Conf., 1997, pp. 762–768
8 MANJUNATH, B.S., OHM, J.-R., VASUDEVAN, V.V., and YAMADA, A.: 'Color and texture descriptors', *IEEE Trans. Circuits Systs. Video Technol.*, 2001, **11**, (6), pp. 703–715
9 MPEG-7 Information Technology-Multimedia Content Descriptor Interface-Part3: Visual. Final Committee Draft, ISO/IEC/JTC1/SC29/WG11/N4062, Mar.2001
10 MPEG-7 Visual Part of Experimentation Model Version 9.0, ISO/IEC/JTC1/SC29/WG11/N3914, Jan.2001, Yamada, A., Pickering, M., Jeannin, S., Cieplinski, L., Ohm, J.-R. and Kim, M., Eds.,
11 DENG, Y., MANJUNATH, B.S., KENNEY, C., MOORE, M.S., and SHIN, H.: 'An efficient colour representation for image retrieval', *IEEE Trans. Image Process.*, 2001, **10**, (1), pp. 140–147
12 ZHOU, X.S., and HUANG, T.S.: 'Image retrieval: feature primitives, feature representation, and relevance feedback'. Proceedings of IEEE Workshop on Content-based access of image and video libraries, in conjunction with CVPR'00, 2000, pp. 10–14
13 LAINE, A., and FAN, J.: 'Texture classification by wavelet packet signatures', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1993, **15**, (11), pp. 1186–1191
14 MEHROTRA, S., RUI, Y., MICHAEL, O.B., and HUANG, T.S.: 'Supporting content-based queries over images in MARS'. Proceedings of IEEE Int. Conf. Multimedia Computing and Systems, 1997, pp. 632–633
15 RUI, Y., HUANG, T.S., ORTEGA, M., and MEHROTRA, S.: 'Relevance feedback: A power tool in interactive content-based image retrieval', *IEEE Trans. Circuits Syst. Video Technol.*, 1998, **8**, (5), pp. 644–655
16 AKROUT, N., PROST, R., and GOUTTE, R.: 'Image compression by vector quantization: a review focused on codebook generation', *Image Vis. Comput.*, 1994, **12**, (10), pp. 627–637
17 DENG, Y., and MANJUNATH, B.S.: 'An efficient low-dimensional color indexing scheme for region-based image retrieval'. Proceedings of IEEE Int. Conf. on Acoustics, speech, and signal processing, 1996, vol. 6, pp. 3017–3020