



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

J. Vis. Commun. Image R. 15 (2004) 507–521

JOURNAL OF
VISUAL
Communication &
IMAGE
Representation

www.elsevier.com/locate/jvcir

Embedded colour image coding for content-based retrieval

Guoping Qiu*

School of Computer Science, The University of Nottingham, UK

Received 21 September 2001; accepted 26 November 2003

Available online 10 February 2004

Abstract

We present an image coding method explicitly designed for easy content access, i.e., content-based image retrieval. Based on a number of well-studied conventional image coding methods, namely segmentation-based image coding (SBIC), vector quantization (VQ), and a recently developed coloured pattern appearance model (CPAM), we have developed an image coding method with a compressed stream from which effective image content descriptors can be derived with very little computation. A colour image is first segmented adaptively into homogeneous regions of various sizes. Each region is then decomposed into three channels according to the CPAM and VQ is employed to represent the chromatic and achromatic spatial patterns efficiently. The image content descriptors are the joint probability distributions of the segmented region sizes and their achromatic and chromatic spatial patterns' VQ codebook indices. From image indexing and content-based retrieval perspective, this work can be regarded as a method effectively exploiting/employing image coding technologies to develop novel and effective image descriptors for content-based image retrieval. We have applied the newly developed image content descriptor to content-based image retrieval from a large colour photo image database. Experimental results demonstrate that the new method is comparable to state of the art methods, such as colour correlogram and the latest MPEG7 colour structure descriptor in content-based image retrieval.

© 2003 Elsevier Inc. All rights reserved.

Keywords: Segmentation-based image coding; Vector quantization; Image coding; Colour imaging; Content-based indexing; Image retrieval; Image database

* Fax: +44-115-951-4254.

E-mail address: qiu@cs.nott.ac.uk.

1. Introduction

We believe it is possible and advantageous to integrate image coding and indexing for efficient storage and easy content access. It is generally agreed that image coding is one of the most successful fields compared to related disciplines such as computer vision in the sense that many practical image coding technologies have been developed and in daily use. In fact, image coding is the key enabling technology in the current digital media revolution. Traditionally, rate-distortion criterion was the goal of image coding; i.e., it is desirable for image coding algorithms to achieve high compression and low distortion. Over the past several decades, much effort has been spent on reducing the bit rate and improving the distortion performances. It can be argued that to a certain extent, the rate-distortion criterion has been met by many modern image coding techniques. Trying to improve bit rates by a fraction of bit per pixel and distortion by a fraction of a decibel on existing image coding frameworks will unlikely make a dramatic impact on the field and may not be the best use of resources. Furthermore, with the rapid development in hardware, storage technologies and high-speed networks, a slightly higher bit rate or a little bit more complexity will be acceptable in many practical applications. With the explosive increase in visual data (image and video), a new challenge faced by the research community is how to make the vast collection of images and video data easily accessible. It is therefore our motivation in this work to make compressed contents easily accessible, and as such our emphasis will be on balancing the performances of image retrieval and compression.

Since content-based indexing and image coding are closely related in the sense that both have the objective of developing compact representations of the original image data, our other motivation of the current work is to explore the achievements of image coding community to develop new and effective image indexing features.

Content-based image indexing and retrieval is a promising technology for managing large image/video databases (Rui et al., 1999). Research in this direction has been actively pursued in different disciplines for over a decade. Well-known techniques such as colour histogram (Swain and Ballard, 1991) and textures descriptors (Manjunath and Ma, 1996), newer methods such as colour correlogram (Huang et al., 1997), blobworld (Carson et al., 1999), and many others variations, such as those listed in the latest international standard (MPEG7 FCD, 2001) have been developed.

In terms of image retrieval accuracy and relevance, all these methods have different strengths and weaknesses. Evaluation of a certain technique's performance in content-based retrieval is often difficult because up to this date, there is no agreed measuring criterions and benchmark testing data set to compare different methods. However, a common weakness of all current methods is that all pre-compute indexing features and store them as meta-data alongside the imagery data itself. One of the objectives of our present work is to develop more flexible and more efficient methods for content-based retrieval; a method which can derive the indexing features effortlessly from the compressed image bit streams.

Naïve approaches to content-based image indexing store images using standard compression techniques and store low-level image features as explicit side information as indices. However, these methods are unsatisfactory because extra side information leads to data expansion; the use of pre-computed indexing features restricts the flexibility of retrieval methods; and computing new indexing features requires either partial or full decompression, which demands extra computation. Since image coding is an essential component of an imaging system, it is therefore desirable to make the coded bits accessible “midstream” without the need to decompress the image. Making image data accessible in the compressed domain, or midstream, has been advocated as an extra criterion for image coding, the so-called “4th criterion” (Picard, 1994). Midstream content access, e.g., query and retrieval, for large image database is advantageous because, (1) the compressed domain has fewer bits to look through, (2) it saves time decompressing unwanted data, (3) it saves space storing decompressed data, and (4) the compressed domain has the same information as pictures at the end.

Segmentation-based image coding, also known as the second generation image coding, was once regarded as a very promising image compression method and was actively pursued by researchers in the 1980s (Ebrahimi and Kunt, 1998; Kunt et al., 1985). However, we believe its full potential was never realized due to the algorithmic and computer hardware limitations of the time. We believe this line of research will be useful in managing large image databases. In particular it can be used to compress images to moderate bite rates and to enable fast processing in the compressed domain for easy content access (Qiu, 2001b). Completely unconnected to image coding and compressed domain content access, the blobworld system (Carson et al., 1999) has demonstrated the advantages of segmenting images into meaningful regions for content-based retrieval.

Recently, the author developed the coloured pattern appearance model (CPAM) for image coding and indexing (Qiu, 2001a; Qiu, 2002). CPAM models the appearance of a small patch of colour image based on the colour and pattern separable model (Poirson and Wandell, 1993) and opponent colour theory (Kaiser and Boynton, 1996) of human colour vision. The use of separate channels to represent the average intensity, the chromatic and achromatic spatial patterns enables efficient representation/coding of small image patterns (patches) using vector quantization (VQ). It has been demonstrated that simple statistics of the VQ codeword usage, e.g., the histogram of the VQ codewords used to encode the patches of the image, which can be computed very easily enabled effective content-based image retrieval from large image databases (Qiu, 2002).

In this paper, we present a method which combines the constrained adaptive segmentation-based image coding method (Qiu, 2001b; Vaisey and Gersho, 1992) and CPAM for image coding and indexing. The image is first segmented into squared homogeneous regions (based on a controllable criterion). Each segmented region is then coded into three channels of the CPAM and efficient representation is achieved using vector quantization. Similar to the region and colour co-occurrence matrix of our earlier work (Qiu, 2001b), an effective image descriptor can be derived from the representation stream with little computation—as if the image

content descriptor is implicitly “embedded” in the compressed stream. We will show that this new image descriptor can be used for effective content-based retrieval and has comparable performances to state of the art methods such as the MPEG7 colour structure descriptor (MPEG7 FCD, 2001) and the colour correlogram method (Huang et al., 1997).

This paper is organized as follows. In Section 2, we briefly review a variable block size segmentation-based image compression method and describe a simple adaptive image segmentation algorithm. In Section 3, we briefly describe the coloured pattern appearance model and the use of vector quantization for the efficient representation of chromatic and achromatic spatial image patterns. Section 4 presents a technique for combining adaptive segmentation and CPAM for image coding and derive easily computable image descriptors from the coded image stream. Section 5 presents experimental results on content-based image retrieval from large image databases. Section 6 concludes the presentation.

2. Segmentation-based image coding

Broadly speaking, segmentation-based image coding (SBIC) methods are often classified as 2nd generation image coding (Ebrahimi and Kunt, 1998; Kunt et al., 1985). The idea behind this scheme is that if we can classify image regions into different classes, then we can allocate different number of bits to different regions according to the properties of the region thus achieving optimality in rate-distortion performance. However, in unconstrained image segmentation, both the numbers of the regions and their shapes are determined solely by the image being examined. This fact implies that a very large number of bits may be needed to represent the shape and location information. Therefore, certain constraints have to be imposed in segmenting the image into regions for efficient coding. Many methods based on this principle have been developed over the years. One such technique, which represents one of the most promising segmentation-based approaches to image coding was the variable block size segmentation technique developed by Vaisey and Gersho (1992). In this method, an image is segmented into regions of different sizes. Each region is classified into one of several perceptually distinct categories, and each category is coded using a distinct coding procedure, such as VQ, DCT, and a combination of the two. As has been reported in Vaisey and Gersho (1992) excellent rate-distortion performance can be achieved by such a method. For details of coding methods and achievable bit rates, readers are referred to Vaisey and Gersho (1992).

Following the essential idea of Vaisey and Gersho (1992), we have developed a simple implementation of the variable block size segmentation method (Qiu, 2001b). The constraints we used in this work were as follows: (1) the shapes of the regions were restricted to be square; (2) the maximum size of the regions must not exceed $N \times N$; (3) within each region, the pixels must have similar colours. Therefore an image is segmented into squared regions of sizes 1×1 , 2×2 , 3×3 , 4×4 , ..., $N \times N$. The procedures are as follows.

2.1. A constrained adaptive segmentation algorithm (CASA)

1. Scan the image from left to right top to bottom direction. If the current pixel $p(x, y)$ has been assigned a region label, then move to the next pixel. Notice we are working on colour images and $p(x, y)$ is a vector.
2. If $p(x, y)$ has not been assigned a region label, then do the following
 - 2.1 Set $B = B_{\min}$, where B_{\min} is the smallest region size.
 - 2.2 Calculate

$$e(x, y) = \frac{1}{B^2} \sum_{i=0}^{B-1} \sum_{j=0}^{B-1} \|p(x+i, y+j) - m(x, y)\|, \text{ where}$$

$$m(x, y) = \frac{1}{B^2} \sum_{i=0}^{B-1} \sum_{j=0}^{B-1} p(x+i, y+j). \quad (1)$$

- 2.3 If $e(x, y) < EL$; where EL is a pre-set error limit value, and $p(x+B, y)$ and $p(x, y+B)$ have not been assigned a region label, and $B \leq B_{\max}$, where B_{\max} is a pre-set maximum region size, then increase $B+ = B_{\text{step}}$, where B_{step} is block incremental step, and go to 2.2, otherwise go to 2.4.
 - 2.4 Label all $p(x+i, y+i)$, $i = 0, 1, \dots, B-1$, with the next available region label.
3. If not reach the end of the image go to 1. Otherwise stop.

As an example, Fig. 1 shows an image segmented by the algorithm. Depending on the error limit (EL) set by the user, the image can be segmented into having different

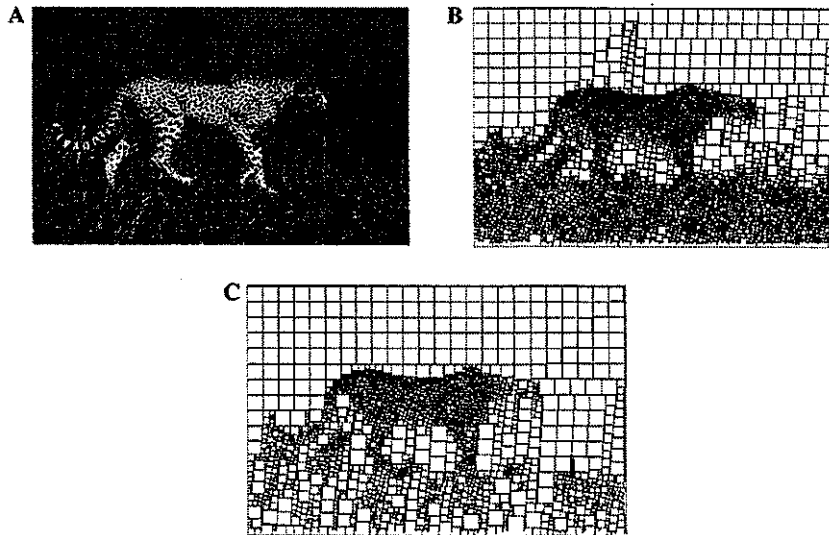


Fig. 1. (A) Original image, (B and C) segmented blocks for $EL = 10$ and 20 . For all results $B_{\min} = 1$, $B_{\max} = 16$, $B_{\text{step}} = 1$.

numbers of various sizes of blocks, thus can be coded to achieve different bit rates and quality levels. In this work, we will use the CPAM method described in the next section to code the segmented blocks. It is worth mentioning that our objectives in this line of work are twofolds, i.e., compression and easy content access, rather than the traditional compression only image coding and the merits of this work should be evaluated as such.

3. Coloured pattern appearance model and vector quantization

The coloured pattern appearance model is a signal processing model developed loosely based on the colour and pattern separable model (Poirson and Wandell, 1993) and opponent colour theory (Kaiser and Boynton, 1996). A colored image pattern is defined as the spatial and spectral characteristics of a (small) block of pixels in a colour image. To fully exploit the colour vision properties of human visual system, an opponent colour space is used to represent colour images. There are many variations of colour spaces such as the $Y C_b C_r$ space (CCIR, 1990) (or similar Pratt, 1978) can be used. The schematic of the combination of CPAM and VQ for image representation is shown in Fig. 2.

A colored image pattern is modeled by three components: the stimulus strength (SS), the achromatic spatial pattern (ASP), and chromatic spatial pattern (CSP). For a small image area, the SS is approximated by the local mean of the Y component. The pixels in Y normalized by SS form the achromatic spatial pattern. Because C_b and C_r have lower bandwidth, they are subsampled by a factor of 2 in both dimensions. The subsampled pixels of C_b and C_r are normalized by SS, to form the colour pattern (CSP) component of the appearance model. Normalizing the pattern and colour channels by the strength has the effect of removing (to a certain extent) the effects of lighting conditions, making the visual appearance model somewhat “colour constant” (Funt and Finlayson, 1995). In constructing the ASP and CSP pattern representations, it has been taken into account the argument that the appear-

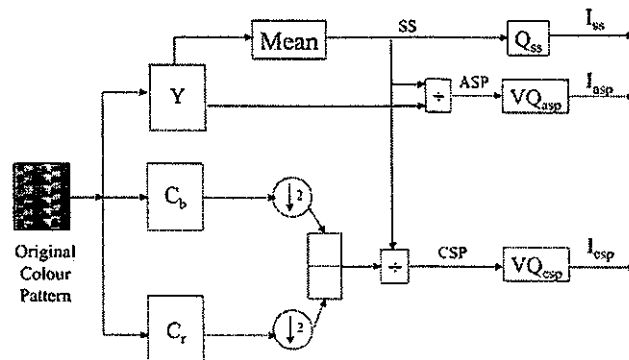


Fig. 2. The CPAM.

ance of a colored image pattern is pattern and colour separable (Poirson and Wandell, 1993). This means that we model the spatial patterns in different chromatic signals independently, hence the chromatic and achromatic patterns are modeled separately. By using an opponent colour representation, it has also been taken into account the fact that opponent chromatic signals have lower bandwidth, thus can be subsampled to form a lower dimensional vector (pattern). By separating achromatic and chromatic signals, it is possible to work on two low-dimensional vectors rather than one very high-dimensional vector. From a computational point of view, low dimensionality is advantageous because higher dimensional signals will suffer from the “curse of dimensionality” phenomenon.

Vector quantization (VQ) (Gersho and Gray, 1992) is a mature method for lossy signal compression/coding in which statistical techniques are used to optimize distortion/rate trade-offs. A vector quantizer is described by an encoder Q , which maps the k -dimensional input vector X to an index $i \in I$ specifying which one of a small collection of reproduction vectors (codewords) in a codebook $C = \{C_i; i \in I\}$ is used for reconstruction, and there is also a decoder, Q^{-1} , which maps the indices into the reproduction vectors, i.e., $X' = Q^{-1}(Q(X))$. The key to the successful use of VQ technology is to obtain a well-designed codebook. There are many methods developed for designing the VQ codebook. The K-mean type algorithms, such as the LGB algorithm (Gersho and Gray, 1992), and neural network-based algorithms, such as the Kohonen feature map (Kohonen, 1989) and the frequency sensitive competitive learning (FSCL) algorithm (Ahalt et al., 1990) are popular tools. To represent the three channels of CPAM, a scalar quantizer can be used to code SS and a separate vector quantizer codes the chromatic and achromatic channels (Qiu, 2002).

4. Integrating SBIC, CPAM, and VQ for colour image coding and indexing

After segmentation using the method described in Section 2, an image will have been divided into squared regions of various sizes. Each such region or block can then be coded using the CPAM. However, if there are many different block sizes, each size needs a separate codebook which will lead to inefficiency because many different codebooks have to be designed and stored. Since the segmentation is based on the homogeneity of the region, a segmented larger block and a segmented smaller block will roughly have the same level of homogeneity. Accordingly all blocks, large or small can be coded with the same level of distortion and similar number of bits. This is in accordance with the principle that a flat area can be coded at lower bit rate than a more detailed area. There are various ways to achieve this such as those developed by Vaisey and Gersho (1992), we describe our implementation in this section.

4.1. VQ coding of variable size patterns

Our strategy is to design one set of codebook at an intermediate block size and which will be used by all the block sizes. The rationale is that if large enough number

of samples were used to construct the codebook, it should capture statistically most important patterns of various frequency contents. If a block is smaller than the codebook's block size, it means that it has higher frequency components, one can sub-sample the codebook block to accommodate the difference, or equivalently, one can up-sample the image block to the same size as the codebook block. On the other hand, if a segmented block is larger than the block of the codebook, it means that the block is flat and has lower frequency content and therefore can be subsampled to the size of the codebook block for coding. Subsampling can be performed by various orders of interpolation. However, we found the simple bilinear method worked very well (Pratt, 1978).

Assuming that the size of the CPAM pattern (the block size of the codebook) is B_c , and B_s is the block size of a segmented block, the coding procedure is as follows: if $B_s < B_c$ then, up-sample B_s to B_c using bilinear interpolation. If $B_s > B_c$ then sub-sample B_s to B_c using bilinear interpolation. The up/down sampling is done in three channels of the input image block (normally in RGB space) independently. The up/down sampled version is then presented to the CPAM of Fig. 2 to be encoded into three indices. If one wants to reconstruct the image, interpolations have to be performed, down sampling the codeword if $B_s < B_c$, or up sampling the codeword if $B_s > B_c$.

4.2. Construction of image descriptors from SBIC/CPAM/VQ stream

For an image coded with the SBIC/CPAM/VQ approach described in Section 4.1, its bit stream contains three indices of various block sizes. From this set of indices, we can construct a similar image content descriptor to the region and colour co-occurrence matrix (Qiu, 2001b). Let $P_{asp}(i, j)$ be the probability of a block of size i (determined by the segmentation algorithm of Section 2) and whose ASP vector is encoded by the vector quantizer to the j th codeword of VQ_{asp} . Similarly, let $P_{csp}(k, l)$ be the probability of a block of size k and whose CSP vector is encoded into the l th codeword of VQ_{csp} . Because it is a trivial matter to collect the statistics $P_{asp}(i, j)$ and $P_{csp}(k, l)$, there is no need to pre-compute and store these features and they can be obtained easily as and when they are needed. This is in contrast to the conventional compressed bit stream which will require significant effort to extract image content description features (Chang, 1995; Seales et al., 1998). By using $P_{asp}(i, j)$ and $P_{csp}(k, l)$ we can compare the similarity of images by comparing these parameters of the image. We will show in the next section that these image descriptors have very good discriminative power and have comparable performance to state of the art image descriptors in content-based image retrieval from large image databases.

5. Experimental results

In terms of image compression, our method is essentially a VQ-based technique. The rate-distortion performance of this type of techniques has been well documented elsewhere (e.g., IEEE, 1996). As we have pointed out at the beginning of the paper,

we are not interested in a fraction of bit per pixel difference since we believe that that is insignificant compared to the huge task of managing large collection of image databases. Instead our main interest in the present work is to demonstrate that by appropriately using modern signal and image processing technology, we can build image compression methods which have a bit stream from where effective image descriptors can be derived very easily for content-based image retrieval. A different angle to view our work is that we exploit image coding technologies for deriving effective image content descriptors for content-based image retrieval. As an example, Fig. 3 shows samples of the coded image at moderate bit rates of about 1 bpp. We observed very satisfactory image quality at this bit rate for a variety of images. No entropy coding or other routines such as DPCM have been implemented to optimize the bit rates. The bit rate can be easily reduced by using a number of conventional coding strategies.

What we mainly want to demonstrate is how well the descriptors derived from the bit stream will work in retrieving images based on their content. We have tested the proposed method using an image database of 20,000 colour photos from the Corel Photo collection, a source used by many other groups to test content-based retrieval. The codebook's block size was chosen to be 4×4 for moderate computational complexity. The code book was trained using more than 15 million samples from a set of over 200 large size (512×512) images from the MIT VisTex texture database (Qiu, 2001a). Note that these training images were not part of the testing images.



Fig. 3. Examples encoded images. Left: original image, 24 bits/pixel. Right: encoded image based on the method described in this paper, about 1 bit/pixel.

The similarity of two images as described by the content descriptor developed in this paper can be calculated by the following L_1 norm measure. Let $P_{Aasp}(i, j)$ and $P_{Acsp}(k, l)$, $P_{Basp}(i, j)$ and $P_{Bcsp}(k, l)$ be the image content descriptors of images A and B , respectively. The similarity between A and B can be measured by the following distance:

$$D(A, B) = \lambda_A \sum_i \sum_j \frac{|P_{Aasp}(i, j) - P_{Basp}(i, j)|}{1 + P_{Aasp}(i, j) + P_{Basp}(i, j)} + \lambda_C \sum_k \sum_l \frac{|P_{Acsp}(k, l) - P_{Bcsp}(k, l)|}{1 + P_{Acsp}(k, l) + P_{Bcsp}(k, l)}, \quad (2)$$

where λ_A and λ_C are relative weights given to the chromatic and achromatic pattern features. This measurement was used originally in Huang et al. (1997). A similar measure known as the x^2 -divergence used in statistics to assess the dissimilarity between two probability density functions was used in Schiele and Crowley (2000) for histogram-based object recognition.

In our experiment, we collected 96 pairs of similar images from various sources as the ground truth, a subset of these images are shown in Fig. 4. Each pair of image



Fig. 4. Examples of query image pairs. For each image in set A, there is a corresponding (similar but different) target image in set B, or vice versa.

Table 1
Relation between error limit value and retrieval performance

EL	Ranks				
	1	≤10	≤30	Max	Average
5	136	169	183	435	9
10	121	166	184	207	7
15	126	165	185	104	5
20	124	170	188	81	4
50	80	128	161	231	15

This table should be interpreted as: for $EL = 5$, 136 queries (out of 192) had their target return in the first rank, 169 queries had their target return within the first 10 positions and so on, the highest rank was 435, and the average rank was 9. Similar interpretation applies to other values of EL.

contains similar but different scenes. These query images were embedded in the database. Each of these 192 images was then used as query, the objective was to retrieve the corresponding similar target images. For each query, the returned images were sorted according to the similarity measure (2), the most similar return was placed in the first rank, and the least similar return last. In all results presented, $\lambda_A = \lambda_C = 1$.

In the first experiment, we tested the effect of the value of EL, the error limit in the segmentation stage on the retrieval performance. In this experiment, we used a subset of 2000 images in the database. The results of performing the 192 queries for various values of EL are shown in Table 1. The trend seemed to be that the higher the error limit, the lower the average ranks of the returned target images. But, if the error limit is too high, the opposite is true (not unexpectedly). Obviously, the higher the error limit, the higher the compression, but image quality will be lower. It is desirable to strike a balance between compression and retrieval performance. It is very difficult to systematically optimize the trades off. However, as a guide, setting the reconstruction peak signal-to-noise ratio when using the block mean colour only to about 30 dB, i.e., $EL = 7$, tends to give very satisfactory image quality and reasonable retrieval performance.

Table 2 shows the result of setting $EL = 7$, $B_{\min} = 4$, $B_{\max} = 10$, and $B_{\text{step}} = 2$, and performed the 192 queries on the larger, 20,000 image database. As a comparison, we have also implemented two state of the art methods in the literature, the color

Table 2
The retrieval performances of different methods for 192 queries (96 pairs) on a database of 20,000 images

Methods	Ranks				
	1	≤10	≤30	Max	Average
CC	138	163	169	7197	226
MPEG7 CS	146	164	173	8160	166
NEW	113	146	162	2532	41

The interpretation of this table is similar to that of Table 1, i.e., for the CC method 138 queries returned their target at the first rank, 163 queries returned their targets within the first 10 positions, etc.

Table 3
The ranks of the returned targets at a rank higher than 100

Methods	Ranks
	≥ 100
CC	100, 101, 107, 120, 210, 429, 777, 810, 1570, 1857, 2112, 2136, 3572, 3925, 5929, 5989, 6678, 6907, 7197
MPEG7 CS	129, 148, 172, 262, 296, 426, 788, 828, 1070, 1198, 1430, 1596, 4062, 4443, 6527, 8160
NEW	111, 122, 175, 181, 198, 224, 245, 314, 317, 347, 762, 1010, 2532

It is seen that the new methods had fewer higher ranks and lower higher ranks.

correlogram (Huang et al., 1997) and the MPEG7 colour structure descriptor (MPEG7 FCD, 2001). It is seen in this particular setting (notice that no effort has been attempted to optimize the performance of our method), the overall performance is difficult to distinguish. On the one hand, CC and MPEG7 CS had more queries found the target image at lower ranks (better performance). On the other hand, both methods had queries which returned the targets at a much higher ranks (worse performance). Also, the average ranking of the new method is much lower (better performance). Also see Table 3 for the returned targets whose rank was higher than 100. Overall, it is fair to say our new method is at least comparable to these state of the art methods.

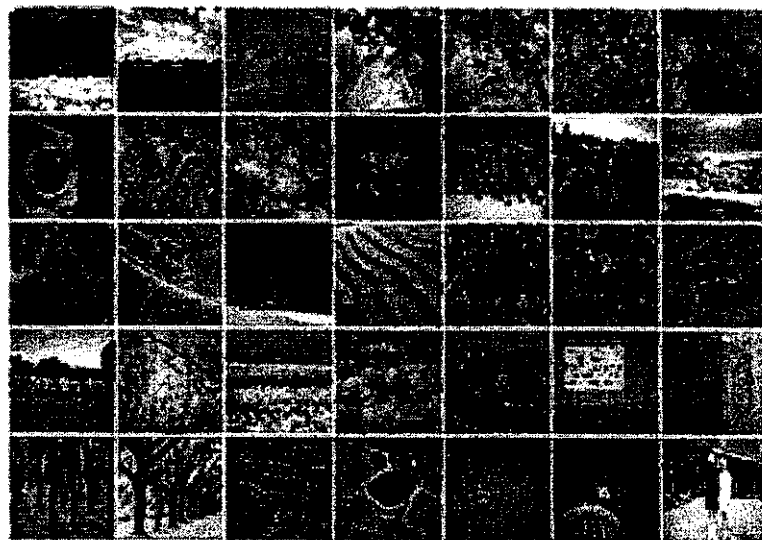
As an example, Fig. 5 shows the first few returned images of a query image for the MPEG7 CS method and our new method, both perform similarly.

6. Concluding remarks

In this paper, we have shown that by integrating a number of well-known image coding methods, it is possible to develop image coding techniques with a compressed image stream from which effective image content descriptors can be easily derived. We have demonstrated that the newly developed image features have comparable performances to state of the art image content descriptors in content-based image retrieval. It is fair to say making compressed image stream easily accessible has not attracted attention of the research community it deserves. Part of the reason may be that historically, the overwhelming demand was on low-bit rate compression. With imaging device becoming ubiquitous, we believe it is high time to place the “4th criterion” (Picard, 1994) of image coding in a more prominent position. From content-based indexing and retrieval perspective, our current work can be regarded as exploiting the mature field of image coding to develop new image descriptors for indexing and retrieval. Image coding and indexing have many things in common in the sense that both are trying to extract from the raw representation the most important image features. For image coding, the extracted features should allow efficient storage and faithful reconstruction of the original image. For image indexing, at low level, the extracted features should be discriminative, and at a higher level these



MPEG7CS



NEW Method

Fig. 5. Examples of retrieved images. The image on the upper left corner is the query, the rest are the returned images arranged in terms of similarity in a canonical order.

features should be semantic. We believe developing methods which meet the requirements of coding and indexing should be a direction worthy pursuing.

Although the 4th criterion in image coding is a philosophically sound idea, as pointed out by one of the reviewers, it may be difficult to implement in practice.

Our current work does not attempt to implement this image coding philosophy, but instead this philosophy has motivated us to investigate the achievements of image coding community to seek new inspirations for the development of new and more effective content-based image indexing features.

Image indexing and retrieval is in general a complex task. In this paper, our contribution has been the development of novel image indexing features inspired by several image coding techniques. Other important issues in image indexing and retrieval, such as similarity measurements, object level image query, scale and rotation invariance, fast indexing and retrieval, etc. were not addressed in the current paper.

Acknowledgments

The author thanks the reviewers for their constructive comments. Part of this work was performed at and sponsored by the Center for Multimedia Signal Processing, Department of Electronic and Information Engineering, Hong Kong Polytechnic University.

References

- Ahalt, S.C. et al. Competitive learning algorithms for vector quantization. *Neural Networks* 3, 277–290.
- Carson, C. et al. Blobworld: a system for region-based image indexing and retrieval. *Proc. Int. Conf. Vis. Inf. Sys.*
- CCIR, 1990. Encoding parameters of digital television for studios, CCIR Recommendation 601-2, Int. audio Consult. Committee, Geneva.
- Chang, S.F. Compressed-domain techniques for image/video indexing and manipulation. *Proc. ICIP*, 314–317.
- Ebrahimi, T., Kunt, M. Visual data compression for multimedia applications. *Proc. IEEE* 86 (6), 1109–1125.
- Funt, B.V., Finlayson, G.D. Color constant color indexing. *IEEE Trans. Pattern Anal. Mach. Intell.* 17, 522–529.
- Gersho, A., Gray, R.M. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, Boston.
- Huang, J. et al. Image indexing using color correlogram. *Proc. CVPR*, 762–768.
- IEEE Transactions on Image Processing*, 1996. Special Issue on Vector Quantization, No. 2, vol. 5.
- Kaiser, P.K., Boynton, R.M. *Human Color Vision*. Optical Society of America, Washington, DC.
- Kohonen, T. *Self-Organization and Associative Memory*. Springer-Verlag, Berlin.
- Kunt, M. et al. Second generation image coding. *Proc. IEEE* 73, 549–574.
- Manjunath, B.S., Ma, W.Y. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intell.* 18, 837–842.
- MPEG7 FCD, 2001. ISO/IEC JTC1/SC29/WG11, March, Singapore.
- Picard, R.W., 1994. Content Access for Image/Video Coding: “The Fourth Criterion,” MIT Media Lab TR No. 295.
- Poirson, A., Wandell, B. Appearance of colored patterns: pattern-color separability. *J. Opt. Soc. Am. A* 10, 2458–2470.
- Pratt, W. *Digital Image Processing*. Wiley, New York.
- Qiu, G., 2001a. Image coding using a coloured pattern appearance model. *Visual Communication and Image Processing 2001*, January, San Jose, CA, USA.

- Qiu, G., 2001b. Constraint adaptive segmentation for color image coding and content-based retrieval. IEEE Workshop on Multimedia Signal Processing, October, Cannes, France.
- Qiu, G. Indexing chromatic and achromatic patterns for content-based colour image retrieval. *Pattern Recogn.* 35, 1675–1686.
- Rui, Y. et al. Image retrieval: current techniques, promising directions, and open issues. *J. Vis. Commun. Image Represent.* 10, 39–62.
- Schiele, B., Crowley, J.L. Recognition without correspondence using multiresolution receptive field histogram. *Int. J. Comput. Vis.* 36 (1), 31–50.
- Seales, W.B. et al. Object recognition in compressed imagery. *Image Vis. Comput.* 16, 337–353.
- Swain, M., Ballard, D. Color indexing. *Int. J. Comput. Vis.* 7, 11–32.
- Vaisey, J., Gersho, A. Image compression with variable block size segmentation. *IEEE Trans. Signal Process.* 40, 2040–2060



G. Qiu is a Lecturer of Computer Science at the School of Computer Science and IT, University of Nottingham, UK. Before joining Nottingham in October 2000, he was a Lecturer at the School of Computing, University of Leeds, UK; and the School of Computing and Mathematics, University of Derby, UK. He received his PhD in neural networks and image processing in 1993 from the University of Central Lancashire, Preston, UK, and his B.Sc. in Electronic Measurement and Instrumentation in 1984 from the University of Electronic Science and Technology of China. From 1987 to 1990 academic years, he was also a postgraduate student at the Department of Electronic Engineering, Beijing Institute of Technology, studied and researched digital signal processing. In the summers of 1999 and 2000, he worked at British Telecommunications Laboratories as a Short-term Research Fellow. His research covers many aspects of visual computing including colour imaging and image databases. He has published some 70 papers in these areas. His research is funded by European Union, Industry and Charitable Funds. He has been a consultant to UK and multinational companies. Since 1993, he has developed and taught many undergraduate and graduate courses, including courses in colour imaging, image processing and computer vision.