Content-Based Retrieval on the Compressed Domain of 2nd Generation Image Coding

G. Qiu

School of Computer Science, The University of Nottingham
Jubilee Campus, Nottingham NG8 1BB, UK
qiu@cs.nott.ac.uk

Abstract. The so-called 2nd generation image coding methods segment images into homogenous regions of various sizes adaptively. Each region is then coded with an appropriate method according to its perceptual characteristics. One of our research aims is to extend 2nd generation image coding methods to meet not only the traditional rate distortion criterion but also the new "easy content access" criterion [6] of image coding. Towards this end, we introduce the region and colour co-occurrence matrix (RACOM), an image content description feature easily computable directly (without decoding) from the compressed domain of 2nd generation image coding methods, and use it for content-based image retrieval from large image databases. We show that RACOM is a very effective image content descriptor and has a comparable performance to state of the art techniques for content-based image retrieval.

1 Introduction

Image coding is the key enabling technology in the current digital media revolution. Traditionally, rate distortion criterion was the goal of image coding. 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. In the author's view, trying to improve bit rates by a fraction of bit per pixel and distortion by a fraction of a dB on existing image coding frameworks is not the best use of resources. With the explosive increase in visual data (image and video), the new challenge faced by the research community is how to make the vast collection of images and video data easily accessible.

Content-based image indexing and retrieval is a promising technology for managing large image/video databases [1]. Research in this direction has been actively pursued in different disciplines for over a decade. Well-known techniques such as colour histogram [2] and textures descriptors [3], newer methods such as color correlogram [4] and blobworld [5], and many others variations [1] have been developed. In terms of image retrieval accuracy and relevance, all these methods have different strengths and weaknesses. In the present paper, we attempt to address a

H.-Y. Shum, M. Liao, and S.-F. Chang (Eds.): PCM 2001, LNCS 2195, pp. 915–920, 2001.

common weakness associated with traditional approaches to content-based image indexing and retrieval.

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, what would be better, is 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" [6].

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 [7]. However, 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. Completely unconnected to image coding and compressed domain content access, the blobworld system [5] has demonstrated the advantages of segmenting images into meaningful regions for content-based retrieval. In this paper, we present the region and color co-occurrence matrix (RACOM) easily computable from the compressed domain of an adaptive segmentation based image coding method [8]. Using the RACOM as image content descriptor we have successfully applied it to content-based image retrieval from large image databases

This paper is organized as follows. In Section 2, we briefly review a variable block size segmentation based image compression method and describe an adaptive image segmentation algorithm. In Section 3, we present the region and colour co-occurrence matrix (RACOM). Section 5 presents experimental results and concluding remarks are given in Section 6.

2 Segmentation-based Image Compression

Broadly speaking, segmentation based image methods are often classified as 2nd generation image coding [7]. The idea behind this scheme is that if we can classified 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 shape 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 [8]. As has been reported in [8] excellent rate distortion performance can be achieved by such a method. For details of coding methods and achievable bit rates, readers are referred to [8]. Of particular interest to our current project is the way in which each variable size block is coded. In particular, the average colour of the block is coded separately which implies it is readily available with minimal computation. Although many other properties of each variable size block are computable also, we shall only exploit the mean colour and study the use of other properties in the future.

Following the essential idea of [8], we have developed our own implementation of the variable block size segmentation method, and we will not repeat the original implementation (the complete coder) in full. Instead, our interest is to demonstrate that image content description features can be easily computed from the compressed bit streams (the block mean colour) of this type of image coding techniques for content-based image retrieval. 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 x N; 3) within each region, the pixels must have similar colours. Therefore an image is segmented into squared regions of sizes 1 x 1, 2 x 2, 3 x 3, 4 x 4, ..., N x N. The procedures is as follows.

Step 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.

Step 2. If p(x, y) has not been assigned a region label, then do the following

2.1. Set S = 1

2.2. Calculate

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

2.3. If e(x, y) < EL; where EL is a pre-set error limit value, and p(x+S, y) and p(x, y+S) have not been assigned a region label, and S < = N, where N is a pre-set maximum region size, then increase S by 1 and go to 2.2, otherwise go to 2.4

2.4 Label all p(x + i, y +i), i = 0, 1, ..., S-1, with the next available region label.

Step 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. Notice that we have not implemented the full coder, because the mean colour of each block will be readily available in a real coder [8] we shall use the mean colour of the segmented blocks directly to derive image content description features in the next section.

3 Region and Colour Co-occurrence Matrix

For the compressed bit streams of the variable block size segmentation based colour image coding, the average colours of the blocks are readily available [8], so are the sizes of the blocks. It is based on the block mean colours and the block sizes we construct the region and colour co-occurrence matrix (RACOM). The RACOM is a two-dimensional array. Each cell, RACOM (m, n), records the probability that a segmented region (SR) of size n x n having an average colour C_m. We know that the number of block sizes are fixed by the coding algorithm, in order to keep the size of

RACOM small, we have to restrict the number of possible colours. Similar to colour correlogram [4], we use a colour table $CT_g = \{C_g(0), C_g(1), \ldots, C_g(M)\}$ (used by all images) to quantize the mean colours in each region to a small fixed set of colours. Notice that the colour table has not been used to code the image but used only for the construction of RACOM. Formally, the RACOM is constructed as

$$RACOM\,(m,n) = \Pr\{Size(SR) = n \mid Q(C_{SR}) = m\}, \forall m,n \tag{1}$$

where C_{SR} is the mean colour of the region and Q is the colour quantizer with a codebook CT_g. The size of RACOM is O(M x N). Since the number of regions in a typical image will be small and it may be possible to construct RACOM on the fly. In addition, we found that RACOM is very sparse and can therefore be stored very efficiently.

4 RACOM for Content-based Image Retrieval

Using RACOM described in the last section as image index, we can perform content-based image retrieval from image databases. To construct the image data database, we can store the RACOM or compute it on-line. Content-based retrieval can be performed by comparing the query image's RACOM with those of the images in the database use an appropriate distance measure. We have found that the following L_1 norm distance measure worked well. Let $RACOM_q$ and $RACOM_d$ be the RACOMs of the query and database images respectively. The similarity of the two images can be measured as the L_1 distance as follows:

$$D(q,d) = \sum_{m=1}^{M}\sum_{n=1}^{N} \frac{\left|RACOM_q(m,n) - RACOM_d(m,n)\right|}{1 + RACOM_q(m,n) + RACOM_d(m,n)} \tag{2}$$

5 Experimental Results

We have tested the RACOM's performance in content-based image retrieval using a database of 20,000 photography color images from the Corel Photo collection. We collected 96 pairs of similar images from various sources as query ground truth. The query images were embedded in the database. The goal was to use one image as query to retrieve the corresponding target image from the database. A subset of the query images is shown in Fig. 2.

In the results presented below, EL = 15, N = 8 and M = 64 (the number of colours). As a comparison we have also implemented the color correlogram of [4] (4 distances and 64 colors). In total, 192 queries were performed, and for majority of the queries, the target image was returned in the first few positions for both colour correlogram (CC) and RACOM. This is in agreement with similar studies by other groups. The cumulative retrieval rate, i.e., the number of retrieved target images above a certain rank, of the CC and RACOM method is shown in table 1. As can be seen, both methods have very similar performance. The average rank of all target images for the

CC method was 265 and for the RACOM method was 129. The average rank was high because a few queries returned their targets in positions greater than 1000. If we consider a query which returned its target at a position higher than 30 as failure, then CC had a success rate of 169/192 (88%) and RACOM had a success rate of 168/192 (87.5%). Examples of retrieved images are shown in Fig. 3.

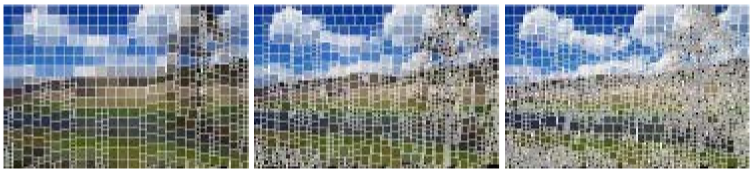


Fig. 1. Images segmented using different values of EL.

	Cumulative Retrieval Rate		
Method	<= 10	< = 30	<= 100
CC	163	169	174
RACOM	157	168	174

Table 1. Cumulative retrieval rate performances

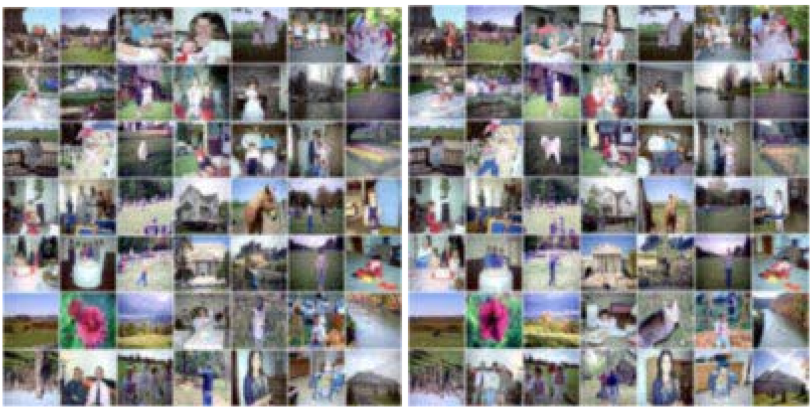
6 Concluding Remarks

In this paper, we have introduced an effective image content descriptor easily computable from the compressed domain of 2nd generation image coding methods. We demonstrated its usefulness in content-based retrieval of colour images in large image database. It is fair to say that 2nd generation coding methods were not favored based on the traditional rate distortion only criterion of image coding. However, with the new demand for easy content access, this type of image coding techniques may have overall advantage over current MPEG and JPEG frameworks in developing image coding and representation methods which will not only satisfy the traditional rate distortion criterion but also the new easy content access criterion of image coding.

References

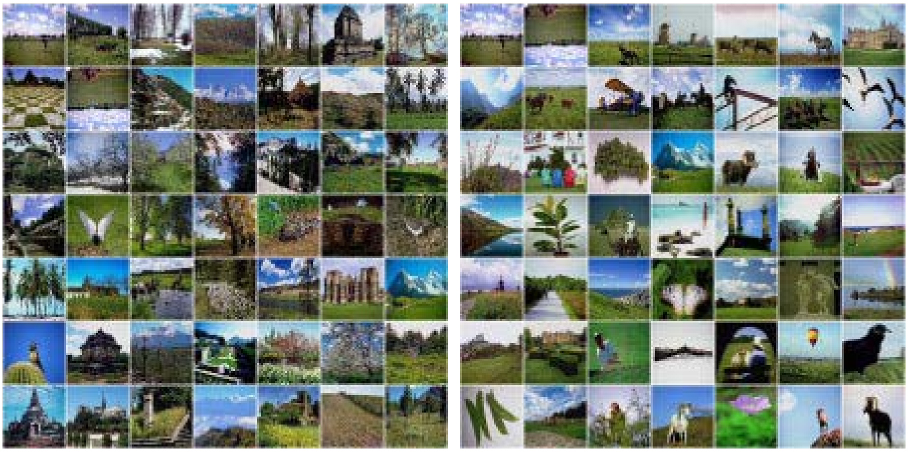
1. 1. Y. Rui et al, “Image retrieval: current techniques, promising directions, and open issues”, Journal of Visual Communication and Image Representation, Vol. 10, pp. 39-62, 1999
2. M. Swain and D. Ballard, "Color indexing", International Journal of Computer Vision, Vol. 7, pp. 11-32, Year 1991
3. B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data”, IEEE Trans Pattern Analysis and Machine Intelligence, vol. 18, pp. 837 – 842, 1996

4. J. Huang, et. al., "Image indexing using color correlogram", Proc. CVPR, pp. 762-768, 1997
5. C. Carson et al, “Blobworld: A system for region-based image indexing and retrieval”, Proc. Int. Conf. Vis. Inf. Sys. 1999
6. R.W. Picard, "Content Access for Image/Video Coding: "The Fourth Criterion"", MIT Media Lab TR No. 295. 1994
7. M. Kunt et al, “Second generation image coding”, Proc. IEEE, vol. 73, pp. 549 –574, 1985
8. J. Vaisey and A Gersho, “Image compression with variable block size segmentation, IEEE Trans Signal Processing, vol. 40, pp. 2040 – 2060, 1992



Set A Set B

Fig. 2. .Query image pairs. For each image in **Set A,** there is a target image in **Set B**



(a) (b)

Fig. 3. , Examples of retrieved images. The top-left corner image is the query and subsequent ones from left to right top to bottom are returned images ordered according to their similarity to the query. (a) results of colour correlogram. (b) results of RACOM.