Available online at www.sciencedirect.com

SCIENCE DIRECT®

Image and Vision Computing 22 (2004) 1211–1220

image and vision COMPUTING

www.elsevier.com/locate/imavis

Content based image retrieval using motif cooccurrence matrix

N. Jhanwar[a,*], S. Chaudhuri[b], G. Seetharaman[c], B. Zavidovique[d]

[a]School of Computing, National University of Singapore, Singapore, Singapore
[b]Department of Electrical Engineering, Indian Institute of Technology, Mumbai 400076, India
[c]Center for Advanced Computer Studies, University of Louisiana, Lafayette, LA 70504-4330, USA
[d]Institut d'Electronique Fondamentale, University of Paris-Sud, 91405 Orsay, France

Received 20 February 2003; received in revised form 16 December 2003; accepted 17 March 2004

Abstract

We present a new technique for content based image retrieval using motif cooccurrence matrix (MCM). The MCM is derived from the motif transformed image. The whole image is divided into 2×2 pixel grids. Each grid is replaced by a scan motif that minimizes the local gradient while traversing the 2×2 grid forming a motif transformed image. The MCM is then defined as a 3D matrix whose (i, j, k) entry denotes the probability of finding a motif i at a distance k from the motif j in the transformed image. Conceptually, the MCM is quite similar to the color cooccurrence matrix (CCM), however, the retrieval using the MCM is better than the CCM since it captures the third order image statistics in the local neighborhood. Experiments confirm that the use of MCM considerably improves the retrieval performance.

Keywords: Content-based retrieval; Peano scan; Optimal scan; Cooccurrence matrix; Image query; Scan motif

1. Introduction

Content based image retrieval (CBIR) has become an intensely researched area [1–3] over the last decade. The steady decline in the cost of mass storage, the increase in on-line accessibility of remotely stored images, and the widespread usage of images in many civilian applications have broadened the scope of CBIR. Some of the earliest avatars of CBIR include query-by pictorial example [4], terrain correlators and automatic target recognizer [5], and the picture archival and communication systems. The first general purpose CBIR system, QBIC [6], was developed nearly 10 years ago. A number of general purpose CBIR systems have been built since then [7–9], which are capable of supporting new algorithms as they are developed. The access to these tools for benchmarking purposes continues to fuel the growth in research on CBIR.

In principle, a CBIR system is required to help retrieve images based on visual properties such as color, texture or pictorial entities such as shape of an object. The primary goal of the CBIR system is to construct meaningful descriptions of physical attributes from images to facilitate efficient and effective retrieval. Both the physical features and the mathematical features have been used in the literature to achieve the same.

Many approaches have been proposed to extract physical features such as color [10], texture [11], sketch [12,13], shape [14], structure or a combination of two or more such features. Swain and Ballard [10] proposed a CBIR color indexing method that is based on color distribution. It is assumed that the spectral content is held constant since the color distribution over an image depends on illumination environment. Tieu and Viola [15] propose to mix one or more features by computing a very large number of highly selective features and comparing these features over a set of known relevant images and using only those features, which capture the similarity amongst these sets. In practice, the feature spare is not always illumination invariant as assumed in the modeling stage, and hence the retrieval is ineffective in some cases.

The color indexing [10] and color histogram [16] based techniques remain popular primarily due to their simplicity. However, these methods produce false positives, since a color histogram lacks spatial information. In particular, the degradation becomes more obvious as the number of images in the database becomes large. The approach is extended to capture the spatial information using color correlogram

* Corresponding author.
E-mail address: nitin@comp.nus.edu.sg (N. Jhanwar).

0262-8856/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.imavis.2004.03.026

as suggested in Refs. [17,18]. The color correlogram is a 3D table where the (i, j, k) entry specifies the probability of finding a pixel of color j, at a distance k from a pixel of color i in the image. Color cooccurrence histogram [19] is just a normalized version of color correlogram.

Wavelet transforms have also been applied [20] to decompose the query and database images, which are then compared using the most significant wavelet coefficients. A hierarchical histogram approach [21,22] is also used in some cases to analyze the images in a multiresolution framework. A more direct hash-table driven approach based on optimum indexing keys such as the Fisher linear discriminant feature has been used in Ref. [23]. Another conventional approach [24] is to compute the discrete Fourier transform of the isochromatically separated components of the query and candidate images.

Image segmentation based approaches [25] seek to partition the images into a number of regions, and make use of the spatial moments of each segment for indexing and comparison amongst the images. Fuzzy logic based techniques have been applied for retrieval in Refs. [26,27], where both the query and the candidate images are analyzed in an appropriately devised fuzzy feature space. The approach in Ref. [27] used a fuzzy feature of image contrast. Graph theoretic methods have also been used to facilitate retrieval [28].

Relevance feedback [18,29,30] has been applied as a tool for refining the search and thus potentially improving the overall retrieval efficiency. It is observed that, in general, if an image is retrieved due to a query image, then the query image must become a strong candidate when the retrieved image is given as a query. This should be a factor in designing the relevance based feedback mechanism.

It is also possible to exploit mathematical features, which perform dimension reduction in a mathematical sense. Since the size of an image is very large, it is practical to represent an image in a lower dimensional feature space rather than in the image space itself. Principal component analysis [31] is one such way to reduce the dimension and to identify similar properties in images. A probabilistic structure based expectation maximization approach has been applied in Ref. [32] to select the most relevant features automatically. This would narrow the gap between the applicability of high level and low level visual features for indexing. Higher level features, relevant in a particular context can be captured [33] through an interactive process designed to highlight the most important features from a class of well known features. The query images is first modified to generate a number of its equivalents (through a set of operators) which are then used interactively to determine the most preferred features in the query image specific to that retrieval.

Recently, some filter bank based classifiers have been used for image classification. The filter response of the filter bank was learnt from training images and represented by clusters [34,35] or histograms [36,37]. The author presented a novel approach [38] for capturing images having similar textures using the joint distribution of intensity values over a small image patch (as small as 3×3 pixel square). The authors in Ref. [39] successfully used cooccurrences of textons for CBIR.

The authors earlier presented a new approach [40] to CBIR based on space filling curves [41,42]. The effectiveness of space filling curves in capturing topologically explicable redundancies [43,44] has been studied in the literature for data compression [45], texture analysis [46,47] and computer graphics [48]. In particular, a specific Peano scan (Z-scan) [49] is used to first map a 2D image data onto a 1D realm. Then, a binary tree called Z-tree [49] is constructed using the Z-scan to facilitate a multiresolution image analysis. A newly defined optimal Z-scan [50] seeks to minimize the net sum of absolute derivative computed over consecutive subscans of length four, at all resolutions. The Fourier Spectrum of the optimal Z-scanned image has been demonstrated [40] to be effective in retrieving the images.

In this paper, we exploit the ability to capture the low level semantics of space filling curves through a cooccurrence measure on a set of six Peano scan motifs at a specific distance. A particular form of Peano scan over a 2×2 grid is called a scan motif and it has been shown later that there can be at most six different motifs (see Fig. 1). The six motifs are defined over a 2×2 grid, each depicting a distinct sequence of pixels starting from the top left corner. These are summarized as follows: Z, ⊐, И, ⊔, ⧖, ⋈. A breadth first traversal of the Z-tree constructed as above, would produce a string, which is wholly explicable using the motif Z. However, when local adaptations are applied over the Z-tree to optimize certain values, for example, the net contrast along the final traversal of pixels, then they would result in a compound string made of all six motifs. The choice of a particular motif over any 2×2 grid depends on the local texture occupying the grid. Then, a statistical feature such as the cooccurrence matrix provides a rich description of the low level semantics. Hence, the motif cooccurrence matrix (MCM) can be assumed to have captured the local statistics in a compact way so that the images can be compared on the basis of their respective MCMs. This is the central idea behind our CBIR approach described in this paper.

In this paper, images are retrieved using a similarity measure on the MCM. The original image is divided into 2×2 grids. These grids are then replaced by a particular Peano scan motif which would traverse the grid in the optimal sense. Here, the optimality of the Peano scan is with respect to the incremental difference in intensity along

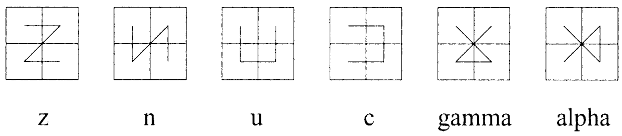

Fig. 1. Primitive scans (called motifs) used to traverse a 2×2 grid.

the scan line minimizing the variation in the intensities in a local neighborhood. In general, 24 different Peano scans (motifs) could traverse a 2×2 grid. But we consider only the Peano scans (motifs) which start from the top left corner of the grid because they represent a complete family of space filling curve, reducing the number of motifs to only 6. The transformed image is used to calculate the probability of finding a motif i at a distance k from a motif j. The distance between the MCMs of two different images is used as the similarity measure while retrieving the images from the database. Hence, the proposed method is quite efficient in terms of both computation and storage requirements. We also show that proposed method is invariant to any monotonic mapping of the image gray levels such as contrast stretching or histogram equalization.

This paper is organized into five sections. Section 2 discusses how to form the MCM for a given image. Section 3 presents the details of processing and retrieval strategies. We present the results and discuss them in Section 4. The paper concludes in Section 5.

2. Optimum Peano scanning

Peano scans are a set of recursively defined space filling curves, spanning over a bounded and simply connected subset of points in two or higher dimensions. Each space filling curve, by definition, is a topological equivalent of a straight line passing through any point exactly once, and spans over the entire set of points in the bounded subspace. The Peano sequences are computed recursively using spatially compact motifs defined over 2×2 grid. The most basic scan is obtained by recursively applying the motif: Ƶ; its topological and computational properties make it a compelling replacement for the standard video rasters. In particular when scalability, hierarchical as well as pipelined computations are among the performance objectives [49–51], the Peano scans are more useful. Also, the Peano scans facilitate a frame work of handling higher-dimensional data for which conventional methods are not easily designable [52].

2.1. Optimal scan pattern

Our effort in search of an optimal Peano scan could be introduced with the following example. Consider a simple run length coding or differential pulse code modulation (DPCM) system used for transmitting a rasterized (video raster) image. The compression system will be maximally effective if the image is made of large number of horizontal runs (constant valued streaks). However, if the underlying image was rotated by 90° about the axis normal to the image plane, before it is rasterized, then the overall compression would drop significantly, and the entropy of the output would not be as low as that of the input data. Suppose if the rasterizer was able to detect the instance and scan the data top-to-bottom and left-to-right, in place of its normal left-to-right and top-to-bottom scanning mode, and instruct the receiving end of the change in scan pattern, then the overall performance improves dramatically. The approach will face its limitations when the image data manifests as a collection of regions, each made of vertical or horizontal runs of unknown values, size and location. Then, one would seek a scheme that will adapt the scan direction locally whenever it detects the need for adaptation.

The locally adapted scan should be such that it minimizes the variation of intensities along the line of scan. Therefore, we use one of the six primitive scans, shown in Fig. 1 to scan the 2×2 grids of the image. The overall effect of using the scan optimal to a certain grid to scan the image is to minimize the abrupt variation in intensities along the scan resulting in compaction of the spectral energy into the low frequency end of the spectrum. Effectively, the image has been transformed into a newer form whose spectral content is concentrated on a narrower zone than the original. Broadly speaking, we locally permute the data to reduce the spectral variation without violating the properties of a space filling curve.

2.2. Motif cooccurrence matrix

Section 2.1 emphasized how we concentrate the spectrum of the image by minimizing the variation of pixels in a local neighborhood. Earlier Jhanwar et al. [40] tried to find the optimal scan with respect to the query image and scan all images in the database using the scan optimal to the query image and compared the transformed images in the frequency domain. Such an algorithm was useful but computationally expensive. Instead of using the spectrum of such an image we can derive a suitable feature by going one step behind in the process of finding the concentrated spectrum. In order to find the concentrated sequence we need to know how every 2×2 grid in the image was scanned. It implies that the information about the way we scan the image defining the local texture is important to derive the highly concentrated spectrum and could be useful to find a suitable feature vector for retrieval purposes.

We derive a transformed image to accommodate the information about the motifs used to scan the corresponding grid in the original image. The transformed image, therefore, shows how the (i,j)th grid was traversed in the original image while minimizing the variation in pixel intensity in that grid. Since every entry in the transformed image represents a 2×2 grid in the original image, the size of the transformed image for a $N \times N$ image would be $N/2 \times N/2$. Fig. 2(a) shows an 8×8 image and Fig. 2(b) shows the corresponding 4×4 transformed image. This transformed image is then used to calculate the MCM. The MCM is constructed using the transformed image whose (i,j,k) entry represents the probability of finding a motif i at a distance k from the motif j. The intuition

(a)

202	53	149	54	255	255	255	124
78	55	84	52	57	190	186	250
129	68	35	128	160	38	36	255
183	29	140	68	54	31	144	182
176	52	47	43	47	53	145	156
145	38	61	45	40	62	140	176
150	186	95	188	220	211	87	167
99	196	189	174	155	159	151	106

(b)

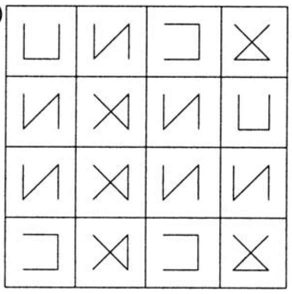

Fig. 2. Illustration of scan motifs. (a) An 8 × 8 image. (b) Motif transformed image of (a).

(a)

53	149	54	255	255	255	124	202
55	84	52	57	190	186	250	78
68	35	128	160	38	36	255	129
29	140	68	54	31	144	182	183
52	47	43	47	53	145	156	176
38	61	45	40	62	140	176	145
186	95	188	220	211	87	167	150
196	189	174	155	159	151	106	99

(b)

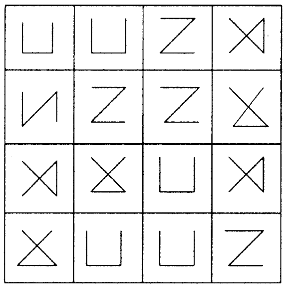

Fig. 3. Illustration of sensitivity of motif transformed image to a translational shift. (a) A horizontally shifted image derived from Fig. 2(a). (b) The motif transformed image corresponding to the shifted image in (a).

behind using a MCM is that if there were a common object in the query and database images the grids corresponding to the object in both the images would be optimally scanned using the same motif structure and the spatial relationship between the motifs would be conserved inside the common object in both cases making the MCM highly effective in image retrieval. Moreover, the size of the feature vector thus derived would be only 6×6 for every color plane irrespective of the image size making the MCM highly suitable for CBIR problem. What we try to achieve by representing the feature through the MCM is to efficiently encode the third order statistics of the intensity field. The motif transformed image encodes the second order (over a 2×2 grid) property and the relative distribution of scan motifs for a distance k (say (0,1) pixels) encodes the third order statistics. Compare this to the color cooccurrence matrix (CCM) which encodes only the first order relationships. Hence, the retrieval based on the MCM is expected to perform better. However, the dimension of the feature space is still very low—only 6×6 for every color plane. It may be noted that we are capturing the MCMs of two images for a fixed distance k in this study.

It may be interesting to compare the motif transformed image to a texton based analysis of the same image. One may consider the different motifs to correspond to different textons and a particular 2×2 grid to be represented by the most relevant texton. The MCM would then represent a texton cooccurrence matrix. Thus, the motif transformed image can be considered as a special case of limited vocabulary texton representation of the image.

2.3. Translation invariance

The feature vector which we had derived in Section 2.2 is very sensitive to translation. If we shift the image by one pixel in any direction, the feature vector corresponding to the new image could be very different from the feature vector of the original image. This is because the four pixels in the corresponding grids of the original image and the image translated by one pixel, which are being compared, are quite different and, therefore, have a different spatial relationship among them and may give an entirely different optimal motif which would traverse the grid. Fig. 3(a) shows the same 8×8 image as in Fig. 2(a) but shifted horizontally by one pixel. Fig. 3(b) shows the corresponding motif transformed image which is quite different from the same derived from the original image (given in Fig. 2(b)). Therefore, a completely different MCM may be recognized for the shifted image during the retrieval, reducing the effectiveness of the MCM.

To diminish the effect of translation we derive four feature vectors (MCM) for the query image, instead of a single MCM. They correspond to the MCMs corresponding to the original image, and the images shifted by one pixel horizontally, vertically and diagonally. The spatial relationship among the motifs of the database images and that of all translates of the query image would be conserved in one of the four derived images independent of the actual amount of translation. In general, if an image is translated x pixels horizontally and y pixels vertically in comparison to the original image, then the translated image would produce an MCM identical to the MCM of the same image translated by (x mod 2, y mod 2) pixels, indicating that one of the four derived images would actually be able to minimize the effect of translation. This is due to the fact that the scan motifs are defined over a 2×2 grid. Needless to say we are neglecting the effect due to boundary pixels. Therefore, when we compare two images to find out the distance between them, all four translates of the query image should be compared with the data base images. The minimum distance which we get by comparing the translates of the query image is considered to be the actual distance between the query image and the database image, and this is used for retrieval purposes. The above procedure makes the feature space invariant to translation, a property very much required in CBIR.

The above strategy though useful would only detect motifs, which would lie on even pixel boundary (i.e. separated by a block of pixels of size 2×2). An alternate approach to achieve translational invariance is to transform the image into $N \times N$ image by determining the motif corresponding to every pixel location. This is equivalent to a non-linear filtering of the image at all pixel locations with an appropriate filter bank (in our case we search for the best fit motif). This would make the transformed image translation invariant as well as provide a better estimation of the MCM as there are three times more observations. The scheme is very sensitive to image being more robust to scaling of images because it would look for motifs on even boundaries of that translated image which has the least effect of scaling at even boundaries.

It is interesting to note that the MCM of an image is invariant to any monotonic mapping of the intensity values in each color plane. This is due to the fact that the intensities in each 2×2 grid is replaced by a suitable scan motifs. Hence, the actual gray levels in the image are irrelevant. They can be transformed by any monotonic function and motifs remain unchanged. Thus, if there is any change in the image contrast or brightness due to camera AGC, the MCM remains unchanged. Similarly, if image is histogram equalized, the MCM is again the same. Since one can change each color plane independently, without affecting the MCMs the proposed method can handle a wide range of color variation due to change in illumination or preprocessing.

3. Image retrieval

In Section 2, we discussed what is a MCM. Conceptually, the MCM is quite similar to the CCM. The primary difference between the MCM and the CCM is that while the CCM encodes the relationship between pixel intensities at two different locations, MCM encodes the relationship between intensity variation along specified scan directions in the image. Consider the 2×2 grid as shown in Fig. 4(a). This grid could be traversed optimally using both the Z and N-type scans as shown in Fig. 4(b). Therefore, in terms of implementation where the CCM encodes the relationship between pixel intensities at two locations uniquely, the MCM has to keep in account that there could be more than one motifs which could traverse the grid optimally. Even though the case shown in Fig. 4(a) may be termed as a pathological case but such a case would be very frequent in regions having a fairly homogeneous texture. One way to take care of this is to update all relevant entries of the MCM.

The information that a particular motif contains is how the pixel intensities vary in a local neighborhood, i.e. over a 2×2 grid. Consider two grids at a distance k from one another, which are used to update the probability assignment in the MCM. Consider that both the grids could individually be traversed optimally using all the six scans. This implies that all the pixel values in these two grids are the same. Therefore, both of these grids belong to a homogeneous region and there is no texture local to both of these grids which could be captured. Updating probabilities using these two grids would give us the wrong information since our feature would suggest that these two grids are actually traversed optimally by the respective motifs. Moreover, such errors does not have equal bearing on the feature vector. The effective change in probabilities would be more pronounced for the pairs of motifs which occur less frequently in the image than the pairs which occur much more frequently. Hence, such instances of non-unique motif transformation are discarded while calculating the MCM. A similar approach was used in Ref. [53] to discard gray scale information to achieve illumination invariance and robustly handle large amount of noise.

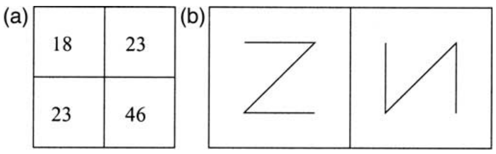

Fig. 4. Illustration of scan ambiguity. (a) A 2×2 grid under consideration. (b) The motifs which would both optimally scan the same grid.

Given two MCMs corresponding to query and database images, one now needs to effectively and efficiently generate the similarity measure. The similarity measure that we used for the purpose of retrieval was simply the product of difference between the corresponding entries in the MCM and an associated weight. Numerically for every color plane c this could be expressed as

$$d_c(I_1, I_2) = \sum_{i=1}^{6} \sum_{j=1}^{6} \alpha_c(i,j) |M_c^{I_1}(i,j) - M_c^{I_2}(i,j)| \tag{1}$$

where I_1, I_2 are the two images, M is the MCM for a fixed distance k (say, (0,1) pixels used in this study), the subscript c is the color plane being considered and $\alpha_c(i,j)$ represents an associated weight. The weight should be such that all the entries corresponding to the MCM should contribute fairly equally while calculating the similarity measure. This suggests that the weight should be dependent on the individual values of the two MCM components which we are comparing. The weight used in our study was

$$\alpha_c(i,j) = \frac{1}{M_c^{I_1}(i,j) + M_c^{I_2}(i,j) + \nu} \tag{2}$$

where ν is a small number used to avoid possible numerical instabilities.

The use of such an weight would normalize the contribution of every element in the distance to a certain extent. The weight would alter the contributions of different elements of the matrix in such a way that

1. A higher priority is given to the elements if both $M_c^{I_2}(i,j)$ and $M_c^{I_1}(i,j)$ are small. A higher priority enables these features to make a substantial contribution to the distance.
2. A lower priority is given to the elements if both $M_c^{I_2}(i,j)$ and $M_c^{I_1}(i,j)$ are large. Since both the elements are large, the difference is likely to be moderately large and such a weight would help in pulling down its contribution in the distance.

Table 1
Contextual description of the database under study

Description	Buildings	Bark	Brick	Fabric	Flowers	Food	Leaves
No. of images	13	9	11	20	8	12	17
Description	Metal	Paintings	Sand	Stone	Terrain	Tile	Miscellaneous
No. of images	6	14	7	6	11	11	25

3. A moderate priority to elements if only one of the MCMs is large. This is where the actual difference comes in the two images. A moderate priority is given to such elements as there is no need to pull down the distance: these are elements which make the two images different. Further there is no need to pull up the distance any more as the distance contribution would already be quite large. In effect, this component contributes the maximum in the overall similarity measure.

Accommodating the recommendation given in Section 2.3 in Eq. (1) to make the feature translation invariant, the modified distance $\bar{d}_c(I_1, I_2)$ between two images I_1 and I_2 for a particular color space could be given as

$$\begin{aligned}\bar{d}_c(I_1, I_2) = \min(&d_c(I_1^{(0,0)}, I_2), d_c(I_1^{(0,1)}, I_2),\\ &d_c(I_1^{(1,0)}, I_2), d_c(I_1^{(1,1)}, I_2))\end{aligned} \quad (3)$$

where $I_1^{(x,y)}$ denotes a translate of the image I_1 by (x, y) pixels. We now need to consider the contribution in the distance measure from all color planes. However, we cannot add the measure $\bar{d}_c(I_1, I_2)$ for the three color planes as we may have situation when we may find distances in the MCMs of two images being minimized for different values of translational shifts. This is not physically justifiable. Hence, we convert the images in to gray-tone (using HSI coordinates) and obtain the translate (x_0, y_0) which minimizes the distance between these two MCMs. We now include the contributions from all color planes, and the overall distance measure finally becomes

$$D(I_1, I_2) = \sum_{c=1}^{3} d_c(I_1^{(x_0,y_0)}, I_2). \quad (4)$$

4. Experimental results

As discussed in Section 1, there are a number of CBIR schemes in the currently available literature and they all report a varied amount of success in terms of retrieval accuracy. In order to test the efficacy of the proposed method we perform a number of experiments but only a few of them are reported in this section for brevity. We compare the performance of the proposed method with that of other techniques that are very similar in approach to demonstrate the superiority of our proposal. It is beyond the scope of current study to compare the performance with all existing CBIR schemes as the feature space is very diverse in many of these algorithms and hence, we restrict the comparison to that of some of the most popular ones. We compare the performance with only those methods that have similar computational and storage requirements. For the presentation of results, we consider only the Vistex database from the MIT Media Lab. The database covered a wide range of categories shown in Table 1. For each query, a sample image is selected and is used for the search of similar images in the entire collection.

We earlier said that the MCM scheme is very much similar to the CCM. Fig. 5 shows the retrieval results corresponding to the query image shown in Fig. 6 using the CCM which captures just the first order statistics in a local neighborhood. The retrieval accuracy is far from being satisfactory as one retrieves only five images correctly out of the top 10 matches.

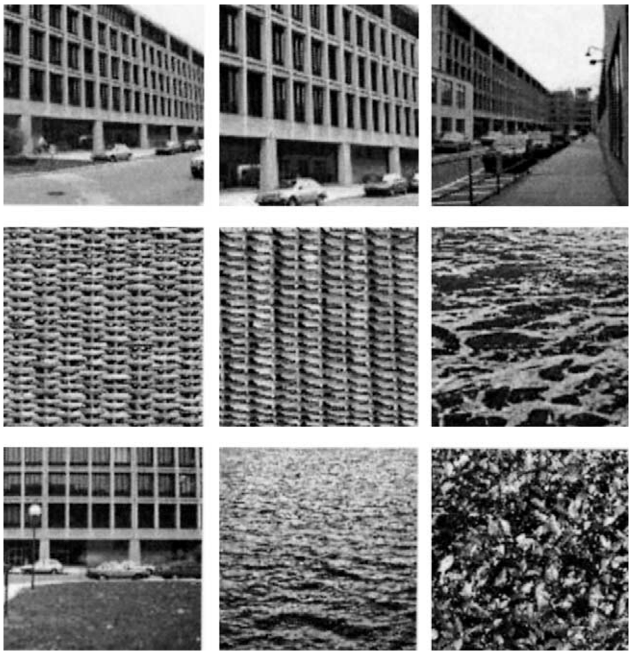

Fig. 5. Retrieval results for the building image using the color correlogram. The results are presented in the descending order of similarity from left to right and top to bottom. The best match is the query image itself, which is not shown here.

Fig. 6. Example of a sample query image.

In order to proceed in a step by step way, we first report the performance of our algorithm when we use only the texture information and disregard the color information. We first convert the query and all the database images into monochrome images. We derive MCMs for these gray-tone images and use it for retrieval purposes. Since these features do not contain the color information, the retrieval would be solely on the textural semantics of the image as captured by the scan motifs. Fig. 7 shows the retrieval results for the same query image in Fig. 6 using the feature vector derived from the gray level image. We observe that the corresponding performance, which is marginally better than the CCM based technique, is still far from being satisfactory. Though seven of the extracted images appear to be quite relevant, the order in which the relevant images appear is improper. For example, the image at rank 6 should be placed much ahead of the image at rank 3. This simple exercise demonstrates that the color information is also very important.

Now we perform the experiment to study what happens if the color and the (monochrome) texture informations are considered as separate entities. In order to implicitly introduce the global color distribution of the image into the feature vector already derived from the monochrome image, we define new similarity measure given by

$$d'' = \beta d' + (1 - \beta)d \tag{5}$$

where d'' is the new distance measure, d' is the distance measure derived from the MCM of the gray level image and d is the distance using the simple color histogram. The coefficient β $(0 \leq \beta \leq 1)$ decides the relative contributions of d and d' in the resulting similarity measure and it reflects the weight placed by the user in terms of dominance of either the color or the texture features. The optimum value of β could be found by the user using the technique of relevance feedback. We refrain from doing that in this study. Fig. 8 shows the retrieval results using the same query image but with the modified similarity measure. Now the retrieval results are definitely better than using only d' (texture) or d (color) as the measure. But we still retrieve about 2–3 irrelevant pictures from the database. This substantiate

Fig. 7. Retrieval results for the building image using the MCM of the monochrome image.

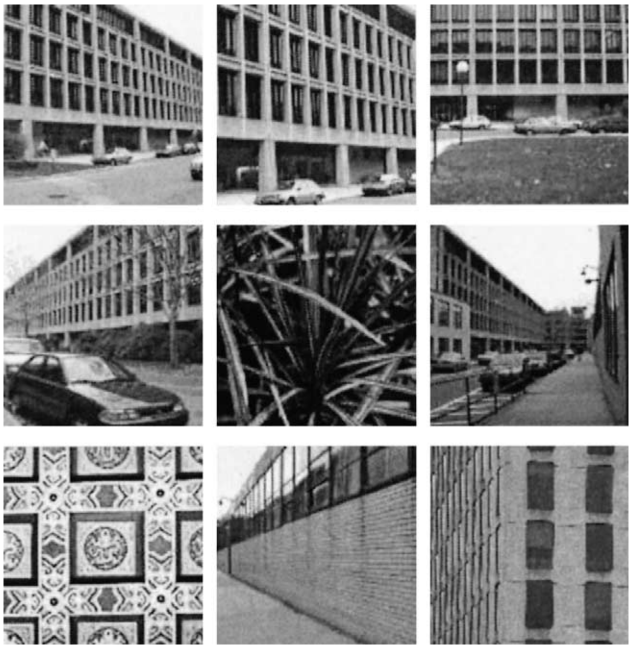

Fig. 8. Retrieval results by combining color and MCM as two independent features.

Fig. 9. Retrieval results for the same query image using the MCM of color image but without correcting for possible translational shift.

our understanding that the above problem is due to the fact that both these features, the color and the texture, are intrinsically related in an image and we should not be considering them as independent features.

Now that we understand the importance of considering both the color and the textural information together, our algorithm as discussed in Section 3 precisely achieves that. Fig. 9 shows the retrieval results when the image is

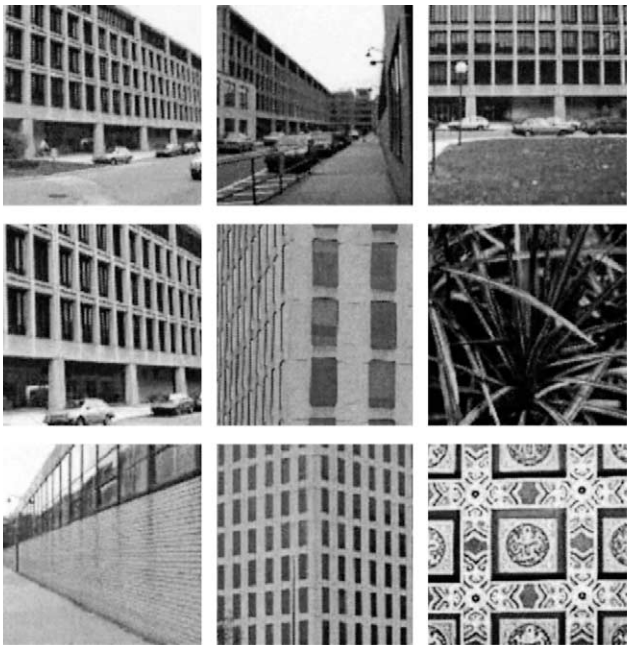

Fig. 10. Retrieval results for the proposed method after making the feature invariant to translation.

Fig. 11. Retrieval results when each pixel of the image is replaced by optimum motif traversing the 2×2 grid around that pixel.

scanned by the six primitive Peano scan motifs in all the three different color planes, but without making the feature translation invariant. We can see that there are about 7–8 images very relevant to the query image. But the order in which they appear is not proper, indicating the need for further improvement. The poor performance of the method is due to the fact that MCMs are not invariant to translational shift over a 2×2 grid as explained in Section 2.3. Hence, we must verify the performance by making the similarity measure translation invariant. Fig. 10 shows the corresponding results for the modified algorithm to get rid of the problem due to translation. We observe that all the relevant images which we had retrieved earlier are again retrieved but now they are retrieved in a more orderly fashion. Fig. 11 shows the results when image was transformed with the motif traversing the grid around each pixel of the image instead of being defined over non-overlapping 2×2 blocks. The result demonstrates that the feature captures translation invariant information and gives more relevant images. But this feature was experimentally found to be a bit sensitive to scaling (image magnification) and therefore the order in which the retrieved images appear is not that appropriate. Similar results were also observed with other queries. The experiments were also performed on a different database and similar observations are obtained.

5. Conclusion

In this paper, we have presented a method for image retrieval using the MCM of images. The MCM was shown to combine the information related to both color and texture. Through the MCM we capture the third order neighborhood statistics of the image, unlike the CCM which encodes only the first order statistics. Since the MCM is sensitive to translational effects, we have proposed how to make the retrieval translation invariant. The proposed method is very efficient with regard to both computational and storage requirements. One needs to store only a vector of length 36 per color place, and the computation of the similarity measure is also very inexpensive. The method is invariant to any monotonic mapping of individual color planes, such as gain adjustment, contrast stretching, and histogram equalization.

In future, the MCM representation could be extended to a multiresolution framework to make the method scale invariant. Further, the granularity of the texture can be captured by comparing the respective MCMs at different scales. A binary tree based data structure could be used to represent the image at multiple resolutions. We are currently investigating these issues.

Acknowledgements

The authors are very thankful to the reviewer for very constructive suggestions. The second author is also thankful to the funding support from the University of Paris-Sud in the form of a visiting professorship and to the *Swarnajayanti Fellowship* scheme of DST.

References

[1] A. DelBimbo, V. Castelli, S. Chang, C. Li (Eds.), Special issue: content-based access for image and video libraries, Journal for Computer Visions and Image Understanding, August, Academic Press, New York, 1999.

[2] Y. Rui, T.S. Huang, S. Chang, Image retrieval: current techniques, promising directions and open issues, Journal of Visual Communication and Image Representation 10 (1999) 39–62.

[3] B. Johansson, A survey on: content based search in image databases, Technical Report, Computer Vision Lab, University of Linkoping, Sweden, 2002, http://www.isy.liu.se/cv/Projects/SuveyOnCBIR.

[4] N.S. Chang, K.S. Fu, Query by pictorial example, IEEE Transactions on Software Engineering 6 (6) (1980) 519–524.

[5] R. Nevatia, K. Price, Locating structures in aerial images, IEEE Transactions on PAM1 4 (5) (1982) 476–484.

[6] W. Nilback, et al., The QBIC Project: quering images by content using color, texture and shape, in: SPIE Proceedings on Storage and Retrieval for Image and Video Databases (1993) 173–187.

[7] A. Pentland, et al., Photobook: tools for content based image retrieval, International Journal of Computer Vision 18 (3) (1996) 233–254.

[8] J.R. Smith, S.F. Chang, VisualSEEk: a fully automated CBIR query system, in: Proceedings of Multimedia Conference, Boston, MA, November (1996).

[9] T. Gevers, A. Smeulders, PicToSeek: combining color and shape invariant features for image retrieval, IEEE Transactions on Image Processing 9 (1) (2000) 102–119.

[10] M.J. Swain, D.H. Ballard, Color indexing, International Journal of Computer Vision 7 (1) (1991).

[11] R.W. Picard, F. Liu, A new ordering for image similarity, in: Proceedings of IEEE Conference on ASSP, April (1994) 129–132.

[12] K. Hirata, T. Kato, A rough sketch-based image information retrieval, NEC Research and Development 34 (2) (1993) 263–273.

[13] A.K. Jain, Y. Zhong, S. Lakshmanan, Object matching using deformable templates, IEEE Transactions on PAMI 18 (3) (1996) 267–278.

[14] C.L. Jordan, T. Ebrahimi, M. Kunt, Progressive content-based shape compression for retrieval of binary images, Computer Vision and Image Understanding 71 (2) (1998) 198–212.

[15] K. Tieu, P. Viola, Boosting image retrieval, in: IEEE Computer Society Conference on CVPR, South Carolina, June (2000) 228–234.

[16] G. Pass, R. Zahib, J. Miller, Comparing images using color coherent vector, in: The Fourth ACM International Multimedia Conference, November (1996) 18–22.

[17] J. Huang, R. Kumar, M. Mitra, W. Zhu, W. Zahib, Image indexing using color correlogram, in: IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, June (1997) 762–768.

[18] J. Huang, R. Kumar, M. Mitra, Combined supervised learning with color correlogram for content based image retrieval, in: The Fifth ACM International Multimedia Conference, Seattle, WA (1997) 325–334.

[19] P. Chang, J. Krumm, Object recognition with color co-occurrence histogram, in: Proceedings of the IEEE International Conference on CVPR, Fort Collins, CO (1997) 498–504.

[20] C. Jacobs, A. Finkelstein, D. Salesin, Fast multiresolution image querying, in: Proceedings of the SIGGRAPH, November (1999) 277–285.

[21] S. Kumar, G. Seetharaman, Hierarchical histograms for image retrieval, in: Proceedings of the IEEE International Conference on Image Processing, Vancouver, September 2000 (2000).

[22] M. Mandal, T. Aboulnasr, S. Panchanathan, Fast wavelet histogram technique for image indexing, Computer Vision and Image Processing 75 (1) (1999) 99–110.

[23] M. Lew, D. Denteneer, Similarity measure, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (9) (1999) 871–879.

[24] S. Yogesh, S. Chaudhuri, A generalized color-histogram based scheme for image indexing, in: Workshop on Vision, Models and Visualization (VMV'99), Erlangen, Germany, November (1999).

[25] N. Jhanwar, S. Chaudhuri, Region based image retrieval, in: Proceedings of the Seventh National Conference on Communications, IIT Kanpur, India, January (2000) 320–324.

[26] J. Wu, A.D. Narasimhalu, Fuzzy content based retrieval in image database, Information Processing and Management 34 (5) (1998) 513–534.

[27] S. Santini, R. Jain, Similarity measure, IEEE Transactions on Pattern Analysis and Machine Intelligence, September (1999) 871–883.

[28] S. Akshoy, R. Haralick, Graph-theoretical clustering for image grouping and retrieval, in: Proceedings of the IEEE Conference on CVPR, Fort Collins, CO, June (1999).

[29] C. Nastar, M. Mitschke, C. Meilhac, Efficient query refinement for image retrieval, in: Proceedings of the IEEE Computer Society Conference on CVPR, Santa Barbara, CA, June (1998) 547–553.

[30] T.S. Chua, C.X. Chu, Color based pseudo object model for image retrieval with relevance feedback, in: Proceedings of the IEEE CVPR, Santa Barabara, CA, June (1998) 145–160.

[31] H. Raman, S. Venkatesh, C. Lam, Retrieval of images using similarity based queries, in: Indian Annual Conference on Computer Vision, New Delhi, December (1998) 351–357.

[32] Y. Wu, Q. Tian, T.S. Huang, Discriminant-EM algorithm with application to image retrieval, in: IEEE Computer Society Conference on CVPR, South Carolina, June (2000) 222–227.

[33] G. Aggarwal, S. Ghosal, P. Dubey, Efficient query modification for image retrieval, in: Proceedings of IEEE Computer Society Conference on CVPR, SC, vol. 2, June (2000) 255–260.

[34] T. Leung, J. Malik, Representing and recognizing the visual appearance of materials using 3-dimensional textons, in: International Journal on Computer Vision 43 (1) (2001) 29–44.
[35] O.G. Cula, K.J. Dana, Compact representation of bidirectional texture functions, in: Proceedings of Computer Vision and Pattern Recognition (2001) 1041–1047.
[36] S. Konishi, A.L. Yuille, Statistical cues for domain specific image segmentation with performance analysis, in: Proceedings of Computer Vision and Pattern Recognition (2000) 125–132.
[37] Y. Wu, S.C. Zhu, D. Murnford, Filter random-fields and maximum entropy (FRAME): towards a unified theory for texture modeling, in: IJCV 27 (2) (1998) 107–126.
[38] M. Varma, Texture classification: are filter banks necessary?, in: Proceedings of Computer Vision and Pattern Recognition (2003).
[39] J. Ponce, S. Lazebnik, C. Schmid, A sparse texture representation using affine-invariant regions, in: Proceedings of Computer Vision and Pattern (2003).
[40] N. Jhanwar, S. Chaudhuri, G. Seetharaman, B. Zavidovique, Content based image retrieval using optimal Peano scans, in: International Conference for Pattern Recognition, August (2002).
[41] G. Peano, Sur une courbe qui remplit toute une aire plaine, Mathematische Annalen 36 (1890) 157–160.
[42] D. Hilbert, Uber die stettige Abbildung einter linie auf ein Flachenstuck, Mathematical Annalen 38 (1891) 459–461.
[43] T. Bially, Space-filling curves: their generation and their application to bandwidth reduction, IEEE Transactions on Information Theory 5 (6) (1969) 658–664.
[44] L. Kanal, T.J. Harley, K. Abend, Space filling curves: classification of binary random patterns, IEEE Transactions on Information Theory 11 (4) (1965).
[45] A. Lempel, J. Ziv, Compression of two-dimensional data, IEEE Transactions on Information Theory 32 (1) (1986) 2–8.
[46] J. Quinqueton, M. Berthod, A locally adaptive Peano scanning algorithm, IEEE Transactions on PAMI, PAMI-3 (4) (1981) 403–412.
[47] P.T. Nguyen, J. Quinqueton, Space filling curves and texture analysis, IEEE Transactions on PAMI, PAMI-4 (4) (1982).
[48] R. Dafner, D. Cohen-Or, Y. Matias, Context based space filling curves, EUROGRAPH-ICS Journal 19 (3) (2000).
[49] G. Seetharaman, B. Zavidovique, Image processing in a tree of Peano coded images, in: Proceedings of the IEEE Workshop on Computer Architecture for Machine Perception, Cambridge, CA (1997).
[50] G. Seetharaman, B. Zavidovique, Z-trees: adaptive pyramid algorithms for image segmentation, in: Proceedings of the IEEE International Conference on Image Processing, ICIP98, Chicago, IL, October (1998).
[51] G. Shivaram, G. Seetharaman, Data compression of discrete sequence: a tree based approach using dynamic programming, IEEE Transactions on Image Processing (1997) (in review).
[52] A.R. Butz, Space filling curves and mathematical programming, Information and Control 12 (1968) 314–330.
[53] T. Menp, T. Ojala, M. Pietikinen, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, in: IEEE Transaction on Pattern Analysis and Machine Intelligence (2002) 971–987.