Content Based Image Retrieval Using Motif Cooccurence Matrix

N. Jhanwar[1], S. Chaudhuri[2], G. Seetharaman[3], B. Zavidovique[4]
[1] School of Computing, National University of Singapore, Singapore
[2] Department of Electrical Engineering, Indian Institute of Tech, Mumbai 400076
[3] The Center for Advance Computer Studies, University of Louisiana, Lafayette, LA 70504-4330
[4] Institut d'Electronique Fondamentale, University of Paris-Sud, 91405 Orsay FR

Abstract

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

Keywords: Content-based Retrieval, Peano Scan, Optimal Scan, Image Query.

1. Introduction

Images are being generated today at an ever increasing rate by a variety of sources. A content based image retrieval (CBIR) system is required to retrieve these images effectively and efficiently. Such a system helps the user to retrieve relevant images based on visual properties such as color, texture and pictorial entities such as the shape of an object in the picture.

The prime goal of the CBIR system is to construct meaningful descriptions of physical attributes from images. Physical features and mathematical features are two such typical descriptions. Many research efforts have been made to extract physical features such as color, texture, edge, structure or a combination of two or more. The majority of the proposed solutions are variations of the color histogram[1] initially proposed for object recognition. Since color histogram lacked spatial information these methods were liable to produce false positives specially when the database was large. An attempt to provide spatial information was made in color correlogram [2,3]. The color correlogram is a three dimensional table where the (i, j, k) entry specifies the probability of finding a pixel of color j at a distance k from a pixel i in the image. Color cooccurence histogram[4] is just a normalized version of the color correlogram.

Swain and Ballard[5] proposed a CBIR system called color indexing that is based on color distribution. This algorithm assumes that the spectral content is held constant since the color distribution over an image depends on illumination. Another attempt to mix one or more features was done by Tieu *et al.*[6]. They proposed a mechanism for computing a very large number of highly selective features and comparing these features for some relevant images and using only those selected features which capture similarity in the given relevant images for image retrieval. The disadvantage of these methods is that the image transform or feature transforms are not invertible. Moreover the feature space is not illumination invariant.

Researchers have also attempted retrieving images on the basis of fuzzy logic. Wu *et al.*[7] proposed an appropriate fuzzy feature space and described the query processing in this feature space and indexed images using a complex fuzzy feature vector. Santini *et al.*[8] also proposed a similarity measure based on fuzzy logic. They used a fuzzy feature contrast model to assess the properties of fuzzy judgment and used a fuzzy measure to deal with dependencies among properties.

An alternate scheme for CBIR is to use mathematical features, unlike physical or geometrical features discussed so far. Such mathematical features could be either parametric or non-parametric. An example of parametric features could include modeling the data as an auto-regressive process. For non-parametric features, the data size is often very large, and it is practical to represent an image in a lower dimensional feature space. Principal component analysis[9] is one such way to reduce the dimension mathematically and to identify similar properties in images.

We present a new technique where image is retrieved using the motif cooccurence matrix(MCM). The original image is divided into 2×2 grids. These grids are then replaced by a Peano scan motif which would traverse the grid in some optimal sense. Here the optimality of the Peano scan is with respect to the incremental difference in intensity along 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. 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.

This paper is organized into 5 sections. Section 2 discusses how to form 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 Peano 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.

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 rasterized 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

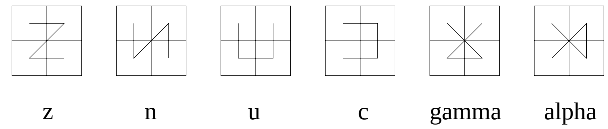


Figure 1: Primitive scans(called motifs) which would be used to traverse a 2×2 grid.

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 6 primitive scans, shown in figure 1 to scan the 2×2 grids of the image. The overall effect called motifs in this paper, of using the scan optimal to a certain grid to scan the image is to minimize abrupt variation in intensities along the scan resulting in compaction of the spectral energy into 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.

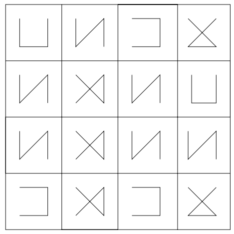
2.2 Motif Cooccurence Matrix

The previous subsection emphasized how we concentrate the spectrum of the image by minimizing the variation of pixels in a local neighborhood. Jhanwar *et al.*[10] tried to find the optimal scan with respect to the query image and scan all images using the scan optimal to the query image and compared the transformed image obtained in the frequency domain. Such an algorithm was useful but computationally expensive at run time. Instead of using the spectrum of such an image we can derive 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 purpose.

We derive a transformed image to accommodate the information about the motif used to scan the corresponding grid in the original image. The transformed image (i,j) entry therefore shows how the (i,j) grid was traversed in the original image minimizing the variation in pixel intensity in that grid. Since every entry in the transformed image represents a 2×2 grid in the original

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

(a)

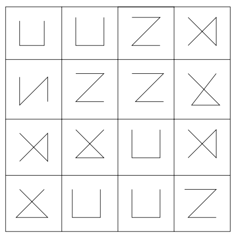


(b)

Figure 2: (a) An 8×8 image (b)Motif transformed image of the 8×8 image shown in (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

(a)



(b)

Figure 3: (a) The horizontally shifted image derived from figure 2 (b) The motif transformed image corresponding to the shifted image in (a).

image the size of the transformed image formed using a $N \times N$ image would be $N/2 \times N/2$. Figure 2(a) shows a 8×8 image and figure 2(b) shows the corresponding 4×4 transformed image. This transformed image is then used to calculate the MCM. A 3 dimensional motif cooccurence matrix is constructed using the transformed image whose (i,j,k) entry represent the probability of finding a motif i at a distance k from motif j. The intuition behind using a motif cooccurence matrix is that if there were a common object in the query and a database image the grids corresponding to the object in both the image would be optimally scanned using the same motif and the spatial relationship between the motifs would be conserved inside the common object in both cases, making 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 MCM highly suitable for CBIR problem.

2.3 Sensitivity Test

The feature vector which we had derived in the previous sub-section 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 transformed by one pixel which are being compared are different and therefore have a different relationship among them and may give an entirely different optimal motif which would traverse the grid. Figure 3(a) shows the same 8×8 image as in figure 2 but which is shifted horizontally by one pixel:Figure 3(b) shows the corresponding transformed image which is quite different from the transformed image derived from original image. Therefore a completely different MCM may be recognized for the shifted image reducing the effectiveness of MCM.

To diminish the effect of translation we derived four feature vectors(MCM) for every image in the database. The feature vectors corresponding to the original image, image shifted by one pixel horizontally, vertically and diagonally. The spatial relationship among the motifs of the query and the translated image would be conserved in one of the four derived images independent of the amount of translation. In general if the database image is translated x pixels horizontally and y pixels vertically in comparison to the query image then the database image would produce MCM similar to the query image if translated by another x mod 2 pixels horizontally and y mod 2 pixels vertically indicating that one of the four derived images would actually be able to minimize the effect of translation. Therefore when we compare two images to find out the distance between the two images, the query image is compared with all the 4 derived images. The minimum distance which we get by comparing the query with the four derived images is said to be the actual distance between the query image and the database image.

3 Image Retrieval

In section 2 we discussed what is a motif cooccurence matrix(MCM). Conceptually the MCM is quite similar to color cooccurence matrix(CCM). The primary difference between the MCM and a CCM is that where 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 figure 4(a). This grid could be traversed optimally using both the Z and N type scans as shown in figure 4(b). Therefore in terms of implementation where CCM encodes the relationship between unique pixel intensities at two location, MCM has to keep in account that there could be more then one motif which could traverse the grid optimally and should be able to track all the motif which could traverse the grid optimally. Even though the case shown in figure 4(a) is a pathological case but such a case would be very frequent in regions having a fairly homogeneous texture. This could be implemented using a link list associated with every grid whose entry contains the motifs which could traverse that grid optimally.

The information which the motif contains is how pixel intensities vary in a local neighborhood i.e. a grid. Consider two grids at a distance k from another grid which are used to update the probabilities in MCM. Consider both the grids could individually be traversed optimally using all the six scans. This implies that all the values local to the two grid are same. Therefore there is no variation in the intensities of pixels in these two grids. It implies that 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 wrong information since our feature would suggest that these two grids are actually traversed optimally by this particular motif which would be an error. Moreover the error is not equally distributed in our feature vector. The change in probabilities would be more for the pairs of motifs which occur less frequently in the image then the pairs which occur more frequently. Therefore such instances would be neglected for MCM calculation.

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

$$d_k(I_1, I_2) = \sum_{i=1}^{6} \sum_{j=1}^{6} \alpha(k,i,j) |M^{I_1}(k,i,j) - M^{I_2}(k,i,j)| \tag{1}$$

where I_1, I_2 are the two images, M is the MCM, k is the color plane being considered and α[k][i][j] represent a dynamic weight. The weight should be such that all the entries corresponding to the MCM should contribute equally for similarity measure. This suggest that the weight should be dependent on the individual value of the two MCM components which we are comparing. The weight which was used in our study was

$$\alpha(k,i,j) = \frac{1}{M^{I_1}(k,i,j) + M^{I_2}(k,i,j) + \beta} \tag{2}$$

where β is a small number used to avoid possible calculation complexities.

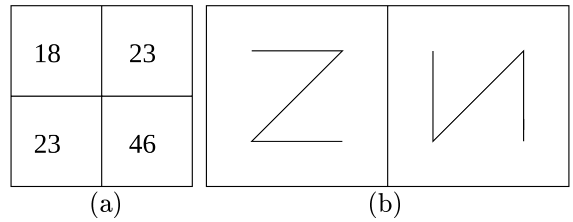


Figure 4: (a) A 2×2 grid under consideration (b) The scans which would optimally scan the 2×2 grid in (a).

Figure 5: A sample query image.

Modifying equation 2 to consider the similarity over all the color planes and accommodating the recommendation given in order to make the feature translation invariant the distance $\bar{D}$ between two images a and b could be given as

$$\bar{D}(a,b) = min(D(a,b_0), D(a,b_1), D(a,b_2), D(a,b_3)) \tag{3}$$

where $b_{0,1,2,3}$ are the four derived images from the original image b and

$$D(a,b) = \sum_{k=1}^{3} d_k(a,b) \tag{4}$$

4 Results

As discussed in the introduction section, there are a number of CBIR schemes in the currently available literature and they all report a varied amount of success in 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 also compare the performance of the proposed method with that of other techniques to demonstrate the superiority of our proposal. However, it is not possible to compare the performance with all existing schemes and hence we restrict the comparison to that with some of the most popular ones.

The Vistex database from the MIT Media Lab has been used in this study to evaluate the performance of the proposed CBIR system. The database covered a wide range of categories. 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. Figure 6 shows the retrieval results corresponding to the query image shown in figure 5 using the CCM which captures just the first order statistics in a local neighborhood. The retrieval accuracy is still far from being satisfactory as one retrieves only five images correctly out of the top ten matches.

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 grey-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. Figure 7 shows the retrieval results for the same query image in figure 5 using the feature vector derived from the grey level image. We observe that the corresponding performance 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 5 should be placed much ahead of rank2. This simple exercise demonstrates that the color information is also very important.

In order to implicitly introduce the global color distribution of the image into the feature vector derived from the monochrome image, we define another similarity measure given by

$$M'' = \alpha M' + (1 - \alpha)M \quad (5)$$

where M'' is the new similarity measure, M' is the similarity measure derived from the MCM of the grey level image and M is the measure using the color histogram. The coefficient α $(0 <= \alpha <= 1)$ decides the relative contributions of M and M' in the resulting similarity measure and it reflects the weight placed by the user in terms of dominance of either color or texture features. The optimum value of α could be found by the user using the technique of relevance feedback. We refrain from doing that in this study. Figure 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 M (texture) or M' (color) as the measure. But we still retrieve three irrelevant pictures from the database. We understand 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 together, our algorithm as discussed in the previous section precisely achieves that. Figure 9 shows the retrieval results when the image is scanned by the 6 primitive Peano scan motifs in all the three different color planes, but without making the feature translation invariant. We can see that there are eight images very relevant to the query image. But the order in which they appear is not proper, indicating that some pruning of the feature is necessary. The feature pruning is required because the MCMs are not invariant to translational shift over a 2×2 grid as explained during the discussion on sensitivity test in section 2. Hence we must verify the performance by translating the image over a 2×2 grid. Figure 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 retrieve earlier are again retrieved but now they are retrieved in a more orderly fashion. Similar results were also observed with other queries.

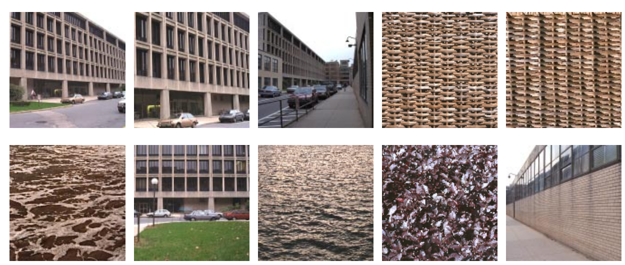


Figure 6: Retrieval results for the building image using the color correlogram.

5 Conclusion

We presented a technique for content based image retrieval which effectively retrieved images using MCM. MCM are computationally inexpensive but were sensitive to translation. A heuristic method defined as sensitivity test was useful in diminishing the effect of limited translation. We also found that MCM is useful in encoding the dependency of both color and texture. In future MCM could be extend in multiresolution since the effect of translation diminishes into lower resolution. One can

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

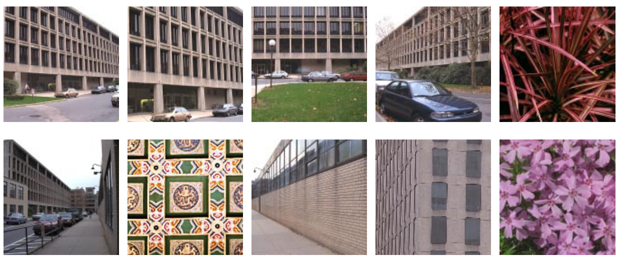


Figure 8: Retrieval results by combining color and MCM as two independent features.

Figure 9: Retrieval results for the building image using the MCM of color image but without correcting for the possible translational shift.

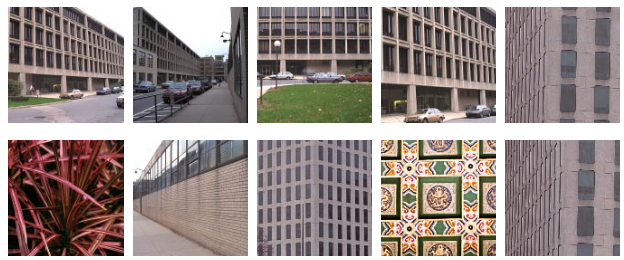


Figure 10: Retrieval results for the proposed method after making the feature invariant to translation.

try to find out the optimal resolution where the translation effect is minimized and use feature corresponding to these layer for retrieval purpose thereby eliminating the need for sensitivity test.

References

[1] Greg Pass, Raman Zahib and Justin Miller, "Comparing images using color coherent vector", *Proc. Fourth ACM International Multimedia Conference*, pp. 65-73, Boston, Massachusetts, November 18-22, 1996.

[2] Jing Huang, Ravi Kumar, Mandar Mitra, W.J. Zhu and R.Zahib " Image Indexing using color correlogram", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 762-768, San Juan, Puerto Rico, June 1997.

[3] Jing Huang, Ravi Kumar and Mandar Mitra "Combining Supervised learning with color correlogram for content based image retrieval", *Proc. Fifth ACM International Multimedia Conference*, pp. 325-334, 1997, Seattle, Washington.

[4] Peng Chang and John Krumm "Object Recognition with color cooccurence histogram", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 498-504, 1997, Fort collins, Collorda, USA.

[5] M.Swain and D.Ballard "Color Indexing", *International Journal on Computer Vision*, pp. 11-33, 7(1), Nov. 1991.

[6] Kinh Tieu and Paul Viola "Boosting Image Retrieval", *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp. 228-234, June 2000, vol. 1, South Carolina.

[7] Jian Kang Wu and A.Desai Narasimhalu "Fuzzy content based retrieval in Image database", *Information Processing and Management*, pp. 513-534, volume 34, Issue 5,sep 1998.

[8] Santini Simome, R. Jain "Similarity Measure", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pp. 871-883, September 1999.

[9] H.Raman, S.Venkatesh and C.P.Lam "Retrieval of Images using similarity based Queries", *Proc. Indian Conference on Vision, Graphics and Image Processing*, pp. 351-357, December 1998, New Delhi, India.

[10] N. Jhanwar, S. Chaudhuri, G. Seetharaman and B. Zavidovique "Content Based Image Retrieval Using Optimal Peano Scans." *The Sixteenth International Conference on Pattern Recognition*, August 11-15, 2002. Quebec City, Canada.