Content Based Image Retrieval using Color, Texture and Shape features.

P. S. Hiremath , Jagadeesh Pujari
Dept. of P.G. Studies and Research in Computer Science,
Gulbarga University,
Gulbarga, Karnataka, India
hiremathps@yahoo.co.in, jaggudp@yahoo.com

ABSTRACT
Color, texture and shape information have been the primitive image descriptors in content based image retrieval systems. This paper presents a novel framework for combining all the three i.e. color, texture and shape information, and achieve higher retrieval efficiency. The image is partitioned into non-overlapping tiles of equal size. The color moments and moments on gabor filter responses of these tiles serve as local descriptors of color and texture respectively. This local information is captured for two resolutions and two grid layouts that provide different details of the same image. An integrated matching scheme, based on most similar highest priority (MSHP) principle and the adjacency matrix of a bipartite graph formed using the tiles of query and target image, is provided for matching the images. Shape information is captured in terms of edge images computed using Gradient Vector Flow fields. Invariant moments are then used to record the shape features. The combination of the color, texture and shape features provide a robust feature set for image retrieval. The experimental results demonstrate the efficacy of the method.

KEY WORDS
Multiresolution grid, Integrated matching, Local descriptors, Gradient vector flow field.

1. Introduction.

Content-based image retrieval (CBIR) [1,2,3,4,5,6,7,8] is a technique used for extracting similar images from an image database. The most challenging aspect of CBIR is to bridge the gap between low-level feature layout and high-level semantic concepts.

Color, texture and shape features have been used for describing image content. Different CBIR systems have adopted different techniques. Few of the techniques have used global color and texture features [8,9,10] where as few others have used local color and texture features [2,3,4,5]. The latter approach segments the image into regions based on color and texture features. The regions are close to human perception and are used as the basic building blocks for feature computation and similarity measurement. These systems are called region based image retrieval (RBIR) systems and have proven to be more efficient in terms of retrieval performance. Few of the region based retrieval systems, e,g, [2], compare images based on individual region-to-region similarity. These systems provide users with rich options to extract regions of interest. But precise image segmentation has still been an open area of research. It is hard to find segmentation algorithms that conform to the human perception. For example, a horse may be segmented into a single region by an algorithm and the same algorithm might segment horse in another image into three regions. These segmentation issues hinder the user from specifying regions of interest especially in images without distinct objects. To ensure robustness against such inaccurate segmentations, the integrated region matching (IRM) algorithm [5] proposes an image-to-image similarity combining all the regions between the images. In this approach, every region is assigned significance worth its size in the image. A region is allowed to participate more than once in the matching process till its significance is met with. The significance of a region plays an important role in the image matching process. In either type of systems, segmentation close to human perception of objects is far from reality because the segmentation is based on color and texture. The problems of over segmentation or under segmentation will hamper the shape analysis process. The object shape has to be handled in an integral way in order to be close to human perception. Shape feature has been extensively used for retrieval systems [14,15].

Image retrieval based on visually significant points [16,17] is reported in literature. In [18], local color and texture features are computed on a window of regular geometrical shape surrounding the corner points. General purpose corner detectors [19] are also used for this purpose. In [20], fuzzy features are used to capture the shape information. Shape signatures are computed from blurred images and global invariant moments are computed as shape features. The retrieval performance is shown to be better than few of the RBIR systems such as those in [3,5,21].

The discussion above clearly indicates that, in CBIR, local features play a significant role in determining the similarity of images along with the shape information of the objects. Precise segmentation is not only difficult to achieve but is also not so critical in object shape determination. A windowed search over location and scale is shown more effective in object-based image retrieval than methods based on inaccurate segmentation [22]. The objective of this paper is to develop a technique which captures local color and texture descriptors in a coarse segmentation framework of grids, and has a shape descriptor in terms invariant moments computed on the edge image. The image is partitioned into equal sized non-overlapping tiles. The features computed on these tiles serve as local descriptors of color and texture. This grid framework is

0-7695-3059-1/07 $25.00 © 2007 IEEE
DOI 10.1109/ADCOM.2007.21

extended across resolutions so as to capture different image details within the same sized tiles. An integrated matching procedure based on adjacency matrix of a bipartite graph between the image tiles is provided, similar to the one discussed in [5], yielding image similarity. A two level grid framework is used for color and texture analysis. Gradient Vector Flow (GVF) fields [13] are used to compute the edge image, which will capture the object shape information. GVF fields give excellent results in determining the object boundaries irrespective of the concavities involved. Invariant moments are used to serve as shape features. The combination of these features forms a robust feature set in retrieving applications. The experimental results are compared with [3,5,20,21].

The section 2 outlines the system overview and proposed method. The section 3 deals with experimental setup. The section 4 presents results. The section 5 presents conclusions.

2. System overview and proposed method.

The Fig. 1 shows the system overview.

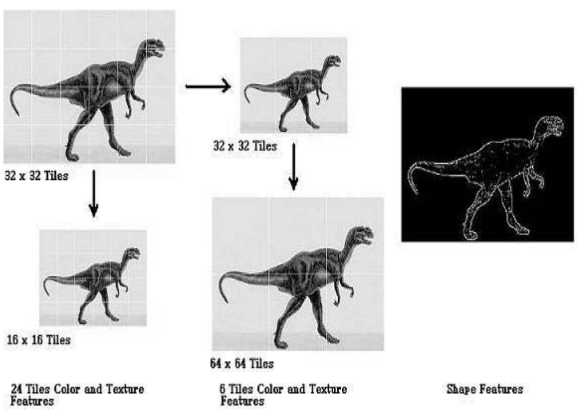

Fig. 1. System overview.

The proposed method is described below:

2.1. Grid.

An image is partitioned into 24 (4 x 6 or 6 x 4) non overlapping tiles as shown in Fig1. These tiles will serve as local color and texture descriptors for the image. Gabor features are used for texture similarity. With the Corel dataset used for experimentation (comprising of images of size either 256 x 384 or 384 x 256), with 6 x 4 (or 4 x 6) partitioning, the size of individual tile will be 64 x 64. The choice of smaller sized tiles than 64 x 64 leads to degradation in the performance. Most of the texture analysis techniques make use of 64 x 64 blocks. This tiling structure is extended to second level decomposition of the image. The image is decomposed into size M/2 x N/2, where M and N are number of rows and columns in the original image respectively. With a 64 x 64 tile size, the number of tiles resulting at this resolution is 6 as shown in Fig 1. This allows us to capture different image information across resolutions. For robustness, we have also included the tile features resulting from the same grid structure (i.e. 24 tiles at resolution 2 and 6 tiles at resolution 1) as shown in Fig. 1. The computation of features is discussed in section 3. Going beyond second level of decomposition added no significant information. So, a two level structure is used.

2.2. Integrated image matching.

An integrated image matching procedure similar to the one used in [5] is proposed. The matching of images at different resolutions is done independently as shown in Fig. 1. Since at any given level of decomposition the number of tiles remains the same for all the images (i.e. either 24 at first level of decomposition or 6 at second level of decomposition), all the tiles will have equal significance. In [23] a similar tiled approach is proposed, but the matching is done by comparing tiles of query image with tiles of target image in the corresponding positions. In our method, a tile from query image is allowed to be matched to any tile in the target image. However, a tile may participate in the matching process only once. A bipartite graph of tiles for the query image and the target image is built as shown in Fig. 2. The labeled edges of the bipartite graph indicate the distances between tiles. A minimum cost matching is done for this graph. Since, this process involves too many comparisons, the method has to be implemented efficiently. To this effect, we have designed an algorithm for finding the minimum cost matching based on most similar highest priority (MSHP) principle using the adjacency matrix of the bipartite graph. In this, the distance matrix is computed as an adjacency matrix. The minimum distance d_{ij} of this matrix is found between tiles i of query and j of target. The distance is recorded and the row corresponding to tile i and column corresponding to tile j, are blocked (replaced by some high value, say 999). This will prevent tile i of query image and tile j of target image from further participating in the matching process. The distances, between i and other tiles of target image and, the distances between j and other tiles of query image, are ignored (because every tile is allowed to participate in the matching process only once). This process is repeated till every tile finds a matching. The process is demonstrated in Fig. 3 using an example for 4 tiles. The complexity of this matching procedure is reduced from $O(n^2)$ to $O(n)$, where n is the number of tiles involved. The integrated minimum cost match distance between images is now defined as:

$$D_{qt} = \sum_{i=1,n} \sum_{j=1,n} d_{ij} \ ,$$

where d_{ij} is the best-match distance between tile i of query image q and tile j of target image t and D_{qt} is the distance between images q and t.

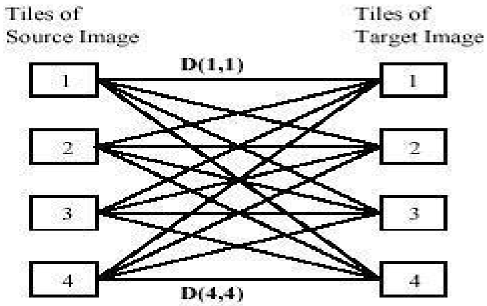

Fig. 2. Bipartite graph showing 4 tiles of both the images.

4.87	2.56	14.88	3.71
1.67	9.45	2.43	3.39
5.54	9.28	13.29	4.78
2.67	18.67	25.33	7.81

(a)

999	2.56	14.88	3.71
999	999	999	999
999	9.28	13.29	4.78
999	18.67	25.33	7.81

(b)

999	999	999	999
999	999	999	999
999	999	13.29	4.78
999	999	25.33	7.81

(c)

999	999	999	999
999	999	999	999
999	999	999	999
999	999	25.33	999

(d)

D = 1.67 + 2.56 + 4.78 + 25.33 = 34.34

Fig. 3. Image similarity computation based on MSHP principle, (a) first pair of matched tiles i=2,j=1 (b) second pair of matched tiles i=1, j=2 (c) third pair of matched tiles i=3, j=4 (d) fourth pair of matched tiles i=4,j=3, yielding the integrated minimum cost match distance 34.34.

2.3. Shape.

Shape information is captured in terms of the edge image of the gray scale equivalent of every image in the database. We have used gradient vector flow (GVF) fields to obtain the edge image [13].

Gradient Vector Flow:

Snakes, or active contours, are used extensively in computer vision and image processing applications, particularly to locate object boundaries. Problems associated with their poor convergence to boundary concavities, however, have limited their utility. Gradient vector flow (GVF) is a static external force used in active contour method. GVF is computed as a diffusion of the gradient vectors of a gray-level or binary edge map derived from the images. It differs fundamentally from traditional snake external forces in that it cannot be written as the negative gradient of a potential function, and the corresponding snake is formulated directly from a force balance condition rather than a variational formulation.

The GVF uses a force balance condition given by

$$F_{\text{int}} + F_{ext}^{(p)} = 0 \ ,$$

where F_{int} is the internal force and $F_{ext}^{(p)}$ is the external force.

The external force field $F_{ext}^{(p)} = V(x, y)$ is referred to as the GVF field. The GVF field $V(x, y)$ is a vector field given by $V(x,y) = [u(x,y), v(x,y)]$ that minimizes the energy functional

$$\varepsilon = \iint \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |V - \nabla f|^2 \, dxdy$$

This variational formulation follows a standard principle, that of making the results smooth when there is no data. In particular, when $|\nabla f|$ is small, the energy is dominated by the sum of squares of the partial derivatives of the vector field, yielding a slowly varying field. On the other hand, when $|\nabla f|$ is large, the second term dominates the integrand, and is minimized by setting $V = |\nabla f|$. This produces the desired effect of keeping V nearly equal to the gradient of the edge map when it is large, but forcing the field to be slowly-varying in homogeneous regions. The parameter μ is a regularization parameter governing the tradeoff between the first term and the second term in the integrand.

The GVF field gives excellent results on concavities supporting the edge pixels with opposite pair of forces, obeying force balance condition, in one of the four directions (horizontal, vertical and diagonal) unlike the traditional external forces which support either in the horizontal or vertical directions only. The algorithm for edge image computation is given below:

Algorithm: (edge image computation)

1. Read the image and convert it to gray scale.
2. Blur the image using a Gaussian filter.
3. Compute the gradient map of the blurred image.
4. Compute GVF. (100 iterations and $\mu = 0.2$)
5. Filter out only strong edge responses using $k\sigma$, where σ is the standard deviation of the GVF. (k – value used is 2.5)
6. Converge onto edge pixels satisfying the force balance condition yielding edge image.

3. Experimental Setup.

(a) Data set: Wang's [11] dataset comprising of 1000 Corel images with ground truth. The image set comprises 100 images in each of 10 categories. The images are of the size 256 x 384 or 384 x 256.

(b) Feature set: The feature set comprises color, texture and shape descriptors computed as follows:

Texture: Texture feature set comprises first and second order statistical moments of Gabor filter responses [26]. The 4 scales and six orientations are considered yielding a total of 24 filter responses. The L component of the image in CIE-Lab space is considered for this purpose.

Color: The first and second order moments of the color bands, a and b, of the image in CIE-Lab color space, are computed as color features.

A total of 52 features are computed for every image tile (per resolution).

Shape : Translation, rotation, and scale invariant one-dimensional normalized contour sequence moments are computed on the edge image [24,25]. The gray level edge images of the R, G and B individual planes are taken and the shape descriptors are computed as follows:

$$F_1 = \frac{(\mu_2)^{1/2}}{m_1},$$

$$F_2 = \frac{\mu_3}{(\mu_2)^{3/2}},$$

$$F_3 = \frac{\mu_4}{(\mu_2)^{2}},$$

$$F_4 = \overline{\mu_5},$$

where $m_r = \frac{1}{N}\sum_{i=1}^{N}[z(i)]^r$,

$$\mu_r = \frac{1}{N}\sum_{i=1}^{N}[z(i) - m_1]^r \text{ and } \overline{\mu_r} = \frac{\mu_r}{(\mu_2)^{r/2}}.$$

The $z(i)$ is the set of Euclidian distances between centroid and all N boundary pixels of the digitized shape.

A total of 12 features result from the above computations. In addition, moment invariant to translation, rotation and scale is taken on R, G and B planes individually considering all the pixels [24]. The transformations are summarized as below:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}},$$

where $\gamma = \frac{p+q}{2} + 1$, (Central moments)

$\phi = \eta_{20} + \eta_{02}$, (Moment invariant)

The above computations will yield additional 3 features amounting to a total of 15 features.

The distance between two images is computed as D = D_1+ D_2 + D_3, where D_1 and D_2 are the distance computed by integrated matching scheme at two resolutions and D_3 is the distance resulting from shape comparison.

Canberra distance measure is used for similarity comparison in all the cases. It allows the feature set to be in unnormalized form. The Canberra distance measure is given by:

$$CanbDist(x, y) = \sum_{i=1}^{d} \frac{|x_i - y_i|}{|x_i| + |y_i|},$$

where x and y are the feature vectors of database and query image, respectively, of dimension d.

4. Experimental Results.

The experiments were carried out as explained in the sections 2 and 3.

The results are benchmarked with standard systems using the same database as in [3,5,20,21]. The quantitative measure defined is average precision as explained below:

$$p(i) = \frac{1}{100} \sum_{1 \le j \le 1000, r(i,j) \le 100, ID(j) = ID(i)} 1,$$

where $p(i)$ is precision of query image i, $ID(i)$ and $ID(j)$ are category ID of image i and j respectively, which are in the range of 1 to 10. The $r(i, j)$ is the rank of image j (i.e. position of image j in the retrieved images for query image i , an integer between 1 and 1000).
This value is the percentile of images belonging to the category of image i in the first 100 retrieved images.
The average precision p_t for category $t\ (1 \le t \le 10)$ is given by

$$p_t = \frac{1}{100} \sum_{1 \le i \le 1000, ID(i) = t} p(i)$$

The comparison of experimental results of proposed method with other standard retrieval systems reported in the literature [3,5,20,21] is presented in Table 1. The SIMPLIcity and FIRM are both segmentation based methods. Since in these methods, textured and non textured regions are treated differently with different feature sets, their results are claimed to be better than histogram based method [21]. Further, edge based system [20] is at par or at times better than SIMPLIcity [5] and FIRM [3]. But, in most of the categories our proposed method has performed at par or at times even better than these systems. Fig. 4 shows the sample retrieval results for all the ten categories. The first image is the query image.

The experiments were carried out on a Pentium IV, 1.8 GHz processor with 384 MB RAM using MATLAB.

Table 1. Comparison of average precision obtained by proposed method with other standard retrieval systems[3,5,20,21].

Class	SIMPLIcity [5]	Histogram Based [21]	FIRM [3]	Edge Based [20]	Proposed Method
Africa	.48	.30	.47	.45	.48
Beaches	.32	.30	.35	.35	.34
Building	.35	.25	.35	.35	.36
Bus	.36	.26	.60	.60	.61
Dinosaur	.95	.90	.95	.95	.95
Elephant	.38	.36	.25	.60	.48
Flower	.42	.40	.65	.65	.61
Horses	.72	.38	.65	.70	.74
Mountain	.35	.25	.30	.40	.42
Food	.38	.20	.48	.40	.50

5. Conclusions.

We have proposed a novel method for image retrieval using color, texture and shape features within a multiresolution multigrid framework. The images are partitioned into non-overlapping tiles. Texture and color features are extracted from these tiles at two different resolutions in two grid framework. Gabor filter responses are used as texture features and color moments serve as color features. An integrated matching scheme based on most significant highest priority (MSHP) principle, involving adjacency matrix of a bipartite graph constructed between image tiles, is implemented for image similarity. Gradient vector flow fields are used to extract shape of objects. Invariant moments are used to describe the shape features. A combination of these color, texture and shape features provides a robust feature set for image retrieval. The experiments using the Corel dataset demonstrate the efficacy of this method in comparison with the existing methods in the literature.

References

[1]. Ritendra Datta, Dhiraj Joshi, Jia Li and James Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age", Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, November 10-11, 2005, Hilton, Singapore.

[2]. C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," in *IEEE Trans. On PAMI*, vol. 24, No.8, pp. 1026-1038, 2002.

[3]. Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval," in *IEEE Trans. on PAMI*, vol. 24, No.9, pp. 1252-1267, 2002.

[4]. A. Natsev, R. Rastogi, and K. Shim, "WALRUS: A Similarity Retrieval Algorithm for Image Databases," in *Proc. ACM SIGMOD Int. Conf. Management of Data*, pp. 395–406, 1999.

[5]. J. Li, J.Z. Wang, and G. Wiederhold, "IRM: Integrated Region Matching for Image Retrieval," in *Proc. of the 8th ACM Int. Conf. on Multimedia*, pp. 147-156, Oct. 2000.

[6]. V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Region-based Image Retrieval Using an Object Ontology and Relevance Feedback," in *Eurasip Journal on Applied Signal Processing*, vol. 2004, No. 6, pp. 886-901, 2004.

[7]. W.Y. Ma and B.S. Manjunath, "NETRA: A Toolbox for Navigating Large Image Databases," in *Proc. IEEE Int. Conf. on Image Processing*, vol. I, Santa Barbara, CA, pp. 568–571, Oct. 1997.

[8]. W. Niblack *et al.*, "The QBIC Project: Querying Images by Content Using Color, Texture, and Shape," in *Proc. SPIE*, vol. 1908, San Jose, CA, pp. 173–187, Feb. 1993.

[9]. A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Content-based Manipulation of Image Databases," in *Proc. SPIE Storage and Retrieval for Image and Video Databases II*, San Jose, CA, pp. 34–47, Feb. 1994.

[10]. M. Stricker, and M. Orengo, "Similarity of Color Images," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, pp. 381-392, Feb. 1995.

[11] http://wang.ist.psu.edu/

[12]. P.S.Hiremath, Jagadeesh Pujari, *"Enhancement of Image Retrieval by Image Attention based on Region Layout Analysis and Relevance Feedback."* in Proc. ICACC 2007 International Conference, Madurai, India,pp. 417-420,9-10 Feb, 2007.

[13] Chenyang Xu, Jerry L Prince, "Snakes,Shapes, and Gradient Vector Flow", IEEE Transactions on Image Processing, Vol-7, No 3,PP 359-369, March 1998.

[14] T. Gevers and A.W.M. Smeuiders., "Combining color and shape invariant features for image retrieval", Image and Vision computing, vol.17(7),pp. 475-488 , 1999.

[15] A.K.Jain and Vailalya,, "Image retrieval using color and shape", pattern recognition, vol. 29, pp. 1233-1244, 1996.

[16] D.Lowe, "Distinctive image features from scale invariant keypoints", International Journal of Computer vision, vol. 2(6),pp.91-110,2004.

[17] K.Mikolajezyk and C.Schmid, "Scale and affine invariant interest point detectors", International Journal of Computer Vision, vol. 1(60),pp. 63-86, 2004.

[18] Etinne Loupias and Nieu Sebe, "Wavelet-based salient points: Applications to image retrieval using color and texture features", in Advances in visual Information systems, Proceedings of the 4th International Conference, VISUAL 2000, pp. 223-232, 2000.

[19] C. Harris and M. Stephens, "A combined corner and edge detectors", 4th Alvey Vision Conference, pp. 147-151, 1988.

[20] M.Banerjee, M,K,Kundu and P.K.Das, "Image Retrieval with Visually Prominent Features using Fuzzy set theoretic Evaluation", ICVGIP 2004, India, Dec 2004.

[21] Y. Rubner, L.J. Guibas, and C. Tomasi, "The earth mover's distance, multi-dimensional scaling, and color-based image retrieval", Proceedings of DARPA Image understanding Workshop, pp. 661-668, 1997.

[22] D.Hoiem, R. Sukhtankar, H. Schneiderman, and L.Huston, "Object-Based Image retrieval Using Statistical structure of images", Proc CVPR, 2004.

[23] P. Howarth and S. Ruger, "Robust texture features for still-image retrieval", IEE. Proceedings of Visual Image Signal Processing, Vol. 152, No. 6, December 2005.

[24] Dengsheng Zhang, Guojun Lu, "Review of shape representation and description techniques", Pattern Recognition Vol. 37,pp 1-19, 2004.

[25] M. Sonka, V. Halvac, R.Boyle, Image Processing, Analysis and Machine Vision, Chapman & Hall, London, UK, NJ, 1993.

[26] B.S.Manjunath and W.Y.Ma , "Texture Features for Browsing and Retrieval .of Image Data", IEEE transactions on PAAMI, Vol 18, No. 8, August 1996. Frontier Technologies, Feb 2007, Manipal, India.