

## Content Based Image Retrieval using Color Boosted Salient Points and Shape features of an image.

**Hiremath P. S**

*Department Of Computer Science,  
Gulbarga University,  
Gulbarga, Karnataka, India*

hiremathps@yahoo.co.in

**Jagadeesh Pujari**

*Department Of Computer Science,  
Gulbarga University,  
Gulbarga, Karnataka, India*

jaggudp@yahoo.com

---

### Abstract

Salient points are locations in an image where there is a significant variation with respect to a chosen image feature. Since the set of salient points in an image capture important local characteristics of that image, they can form the basis of a good image representation for content-based image retrieval (CBIR). Salient features are generally determined from the local differential structure of images. They focus on the shape saliency of the local neighborhood. Most of these detectors are luminance based which have the disadvantage that the distinctiveness of the local color information is completely ignored in determining salient image features. To fully exploit the possibilities of salient point detection in color images, color distinctiveness should be taken into account in addition to shape distinctiveness. This paper presents a method for salient points determination based on color saliency. The color and texture information around these points of interest serve as the local descriptors of the image. In addition, the 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 local color, texture and the global shape features provides a robust feature set for image retrieval. The experimental results demonstrate the efficacy of the method.

**Keywords:** *Color saliency, Local descriptors, Gradient vector flow field.*

---

### 1. INTRODUCTION

Content-based image retrieval (CBIR) [1,2,3,4] 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. In CBIR applications, the user typically provides an image (or a set of images) with no indication of which portion of the image is of interest. Thus a search in classical CBIR often relies upon a global view of the image. Localized CBIR [6] has been defined as a task where the user is only interested in a portion of the image and the rest is irrelevant.

To capture the local characteristics of an image, many CBIR systems either subdivide the image into fixed blocks [13,14], or more commonly partition the image into different meaningful regions by applying a segmentation algorithm [2,3,4]. In both the cases, each region of the image is represented as a feature vector of feature values extracted from the region. Other CBIR systems extract salient points (also known as interest points) [10,11,12], which are locations in an image where there is a significant variation with respect to a chosen image feature. With salient point methods, there is one feature vector created for each salient point. These representations enable a retrieval method to have a representation of different local regions of the image, and thus these images can be searched based on their local characteristics.

Usually the performance of a segmentation based method depends highly on the quality of the segmentation. Especially, a segmentation based representation usually measures features on a per-segment basis, and the average features of all pixels in a segment are often used as the features of that segment [3,4]. Therefore, this representation requires high quality segmentations because small areas of incorrect segmentation might make the representation very different from that of the real object. Moreover, the incorrect segmentation also hampers the shape analysis process. The object shape has to be handled in an integral way in order to be close to human perception. Shape has been extensively used for retrieval systems [8,9].

The salient point methods for retrieval assign features to a salient point based on the image features of all the pixels in a window around the salient point. Traditionally salient point detectors for CBIR often use the luminance component of the image for salient point computation, and thus, ignore the color information. The disadvantage of this method is that the salient points often gather at textured portions of the image or on the edges where the change of intensity is significant, so that many salient points capture the same portion of the image. This motivated us to develop a technique for image retrieval that uses color distinctiveness in determining the salient points and also that uses shape features in terms of the object edges. Different sized windows are used to capture the texture and color information around the salient points. Gradient Vector Flow (GVF) fields [7] 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 those in [3,4].

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 schematic block diagram of the proposed system based on color salient points is shown in Fig 1.

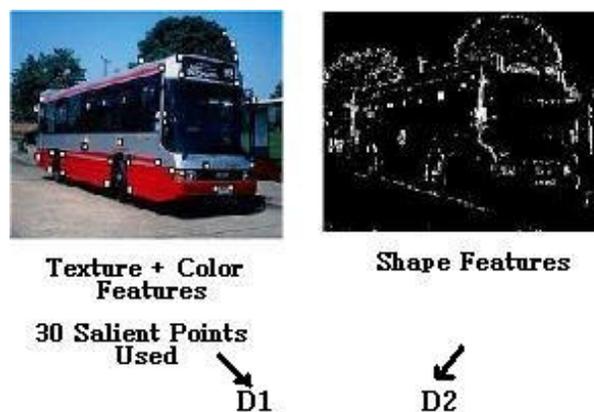


FIGURE 1: Overview of the proposed system.

## 2.1 COLOR DISTINCTIVENESS

The efficiency of salient point detection depends on the distinctiveness of the extracted salient points. At the salient points' positions, local neighborhoods are extracted and described by local image descriptors. The distinctiveness of the descriptor defines the conciseness of the representation and the discriminative power of the salient points. The distinctiveness of points of interest is measured by its information content. Most salient point detectors focus on two dimensional structures, such as corners, which are stable and distinctive at the same time. Color is also considered to play an important role in attributing image saliency. In our approach, the color saliency is based on the work reported in [16]. To achieve the color saliency, the color axes are rotated followed by a rescaling of the axis and, the oriented ellipsoids are transformed into spheres. Thus, the vectors of equal saliency are transformed into vectors of equal length. Fig. 2 shows the salient points detected using the Harris corner detector [12] and the proposed method. The salient points are circled in yellow and blue for Harris detector and the proposed method, respectively. The Harris detector detects points based on black and white events, while the proposed method uses color saliency to detect the events. It can be seen from the figure that the Harris detector detects salient points that typically cluster around textured areas, while the proposed method spreads them according to color saliency. In our experiments we have considered 30 salient points. The texture features were captured in a window of size 9 x 9 around every salient point and the color features in a window of size 3 x 3 around each salient point. The Fig. 3 illustrates this process. The procedure for computation of features is discussed in section 3.

## 2.2 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 [7].

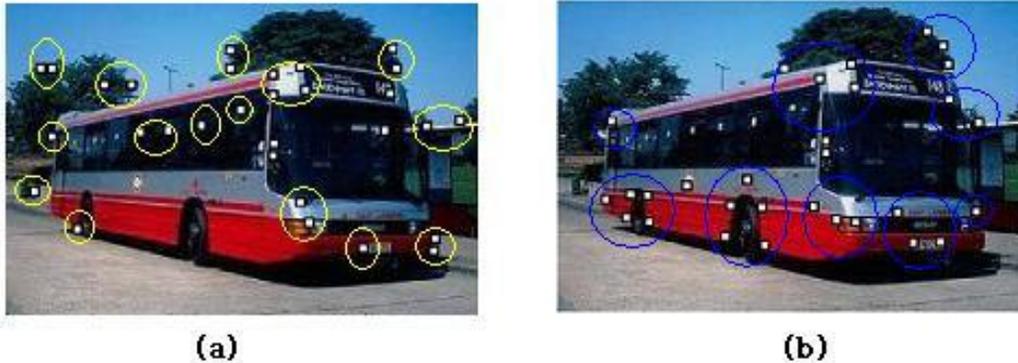


FIGURE 2: (a) Harris Corner Detector. (b) Proposed Method.

### 2.2.1 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 extensively used in active contour method. GVF is computed as a diffusion of the gradient vectors of a grey-level or binary edge map derived from the images. The GVF uses a force balance condition given by

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

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

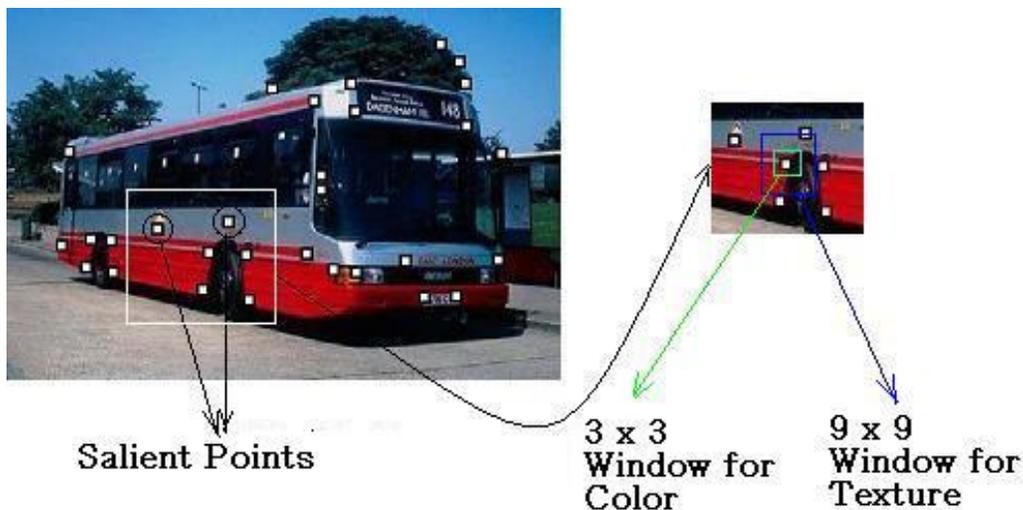
The external force field  $F_{ext}^{(p)} = V(x, y)$  is referred to as the GVF field. The GVF field  $V(x, y)$  is a vector field given by  $V(x, y) = [u(x, y), v(x, y)]$  that minimizes the energy functional

$$\mathcal{E} = \iint \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |V - \nabla f|^2 dx dy$$

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  $\mu$  is a regularization parameter governing the tradeoff between the first term and the second term in the integrand.

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 grey scale 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  $\sigma$  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.



**FIGURE 3:** Feature Computation Process

### 3. EXPERIMENTAL SETUP

(a) Data set: Wang's [5] dataset consists of 1000 Corel images with ground truth. The image set comprises 100 images in each of 10 categories.

(b) Feature set: Texture and color features are explained below.

**Texture:** Gabor filter responses are used as texture features. 6 orientations and 4 scales are considered for this purpose [17]. A window of size 9 x 9 around every salient point is considered to capture the local texture features. First and second order statistical moments of the 24 filter responses, on the L component of the CIE-Lab space for the image, serve as the texture features. The responses were normalized over the entire image database.

**Color:** The first and second order statistical moments of the color bands, a and b, in the CIE-Lab color space of the image, are computed around every salient point within a window of size 3 x 3, as color features.

A total of 52 features are computed for each salient point. A total of 30 salient points are considered. Going beyond 30 salient points, did not yield considerable improvement in retrieval result worth the computational overheads involved.

The overall similarity distance  $D_j$  for the  $j^{\text{th}}$  image in the database is obtained by linearly combining the similarity distance of each individual feature:

$$d_j = \sum w_i s_j(f_i)$$

, with  $S_j(f_i) = (x_i - q_i)^T(x_i - q_i)$ ,  $j = 1, \dots, N$  and  $i = 1, \dots, M$ , where  $N$  is the total number of images in the database and  $M$  the total number of color and texture features. The low level feature weights  $w_i$  for color and texture are set to be equal.

**Shape:** Translation, rotation, and scale invariant one-dimensional normalized contour sequence moments are computed on the edge image [15]. The gray level edge images of the R, G and B individual planes are taken and the shape descriptors are computed as follows:

$$F_1 = \frac{(\mu_2)^{1/2}}{m_1}, \quad F_2 = \frac{\mu_3}{(\mu_2)^{3/2}}, \quad F_3 = \frac{\mu_4}{(\mu_2)^2}, \quad F_4 = \frac{\mu_5}{\mu_2}$$

where

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

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

A total of 12 features result from the above computations. In addition, moment invariant to translation, rotation and scale is computed on R, G and B planes individually considering all the pixels [15]. The transformations are summarized as below:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad \text{where } \gamma = \frac{p+q}{2} + 1 \quad (\text{Central moments}) \quad \phi = \eta_{20} + \eta_{02}, \quad (\text{Moment invariant})$$

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

Canberra distance measure is used for similarity comparison in all the cases. It allows the feature set to be in unnormalized form and 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  $d$ . The distance between two images is computed as  $D = D_1 + D_2$  where  $D_1$  is the distance computed using color and texture information around the salient points and  $D_2$  is the distance resulting from shape comparison..

#### 4. EXPERIMENTAL RESULT

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

The results are benchmarked with standard systems namely, SIMPLcity and FIRM, using the same database as in [3,4]. The quantitative measure defined is average precision as explained below:

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

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  $t$  ( $1 \leq t \leq 10$ ) is given by

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

The results are tabulated in Table 1. The results of retrieval obtained using the Harris corner detector are also provided for the sake of comparison. In most of the categories our proposed method has performed at par or better than other systems. The results are considerably improved by considering color saliency in salient point detection as compared to grey scale salient points detected by Harris corner detector.

#### 6. CONCLUSION

We have proposed a novel method for image retrieval using color, texture and shape features. Salient points based on color saliency are computed on the images. Texture and color features are extracted from fixed sized windows around these salient points to serve as local descriptors. Gradient vector flow fields are used to extract edge images of objects. Invariant moments are used to describe the shape features. A combination of these local color, texture and global shape features provides a robust set of features for image retrieval. The experiments using the Corel dataset demonstrate the efficacy of this method.

Class	SIMPLicity [4]	FIRM [3]	Proposed Method	
			With salient points detected by Harris Corner Detector	With color salient points
Africa	.48	.47	.40	.48
Beaches	.32	.35	.31	.34
Building	.35	.35	.32	.33
Bus	.36	.60	.44	.52
Dinosaur	.95	.95	.92	.95
Elephant	.38	.25	.28	.40
Flower	.42	.65	.58	.60
Horses	.72	.65	.68	.70
Mountain	.35	.30	.32	.36
Food	.38	.48	.44	.46

**TABLE 1:** Comparison of average precision obtained by proposed method with other standard retrieval systems SIMPLicity and FIRM.

## 6. 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 & 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] . 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.
- [5] <http://wang.ist.psu.edu>
- [6] R. Rahamani, S. Goldman, H. Zhang, J. Krettek and J. Fritts, "Localized content-based image retrieval", ACM workshop on Multimedia Image Retrieval, pp. 227-236, 2005.
- [7] 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.
- [8] 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.
- [9] A.K.Jain and Vailalya., "Image retrieval using color and shape", pattern recognition, vol. 29, pp. 1233-1244, 1996.
- [10] D.Lowe, "Distinctive image features from scale invariant keypoints", International Journal of Computer vision, vol. 2(6),pp.91-110,2004.

- [11] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors", *International Journal of Computer Vision*, vol. 1(60), pp. 63-86, 2004.
- [12] C. Harris and M. Stephens, "A combined corner and edge detectors", 4th Alvey Vision Conference, pp. 147-151, 1988.
- [13] Q. Tian, Y. Wu and T. Huang, "Combine user defined region-of-interest and spatial layout for image retrieval", *ICIP 2000*.
- [14] P. Howarth and S. Reger, "Robust texture features for still-image retrieval", *IEE. Proceedings of Visual Image Signal Processing*, Vol. 152, No. 6, December 2005.
- [15] Dengsheng Zhang, Guojun Lu, "Review of shape representation and description techniques", *Pattern Recognition* Vol. 37, pp 1-19, 2004.
- [16] J. van de Weijer, Th. Gevers, J-M Geusebroek, "Boosting Color Saliency in Image Feature Detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27 (4), April 2005.
- [17] 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