



# Chordigram image descriptor based on visual attention model for image retrieval

S. Sathiamoorthy<sup>a</sup>, A. Saravanan<sup>b,\*</sup>, R. Ponnusamy<sup>b</sup>

<sup>a</sup> Tamil Virtual Academy, Chennai, India

<sup>b</sup> Division of Computer and Information Science, Annamalai University, Annamalai Nagar, India

## ARTICLE INFO

### Keywords:

Chordigram image descriptor  
Edge map  
Saliency map  
Salient edges

## ABSTRACT

A novel shape-based image retrieval is presented in this study. The foreground and background contents of images are strongly concealed, so they are represented individually to reduce their influence on each other in the proposed approach. The Otsu method is employed for segmenting the foreground from the background, and the saliency map and edge map are then clearly identified. Saliency reduces the time cost for feature computation, so salient edges are computed for the foreground and background images based on the selective visual attention model. Autocorrelation-based chordigram image descriptors are computed separately for the foreground and background images, which are then combined in a hierarchical manner to form the proposed new descriptor. This approach avoids the concealment of foreground and background information, and the new descriptor is rich in geometric and its underlying texture, structure and spatial information. The proposed novel shape-based descriptor performs considerably better than conventional descriptors at content-based image retrieval. The proposed shape descriptor were extensively tested at image retrieval based on the Gardens Point Walking, St Lucia, University of Alberta Campus, Corel 10 k, and self-photographed image data sets. The precision and recall values were compared for the proposed and state-of-the-art-approaches when applied for shape-based image retrieval from these databases. The proposed shape descriptor provided satisfactory retrieval results in the experiments.

## 1. Introduction

At present, the number of digital images is increasing rapidly due to the use of various photography devices, such as webcams, mobile phones, and closed-circuit television (CCTV) cameras, and thus the sizes of image databases are growing greatly. Hence, storage, retrieval, and maintenance are important tasks for image databases. Image retrieval is broadly divided into two groups comprising (1) text-based image retrieval (TBIR) [1] and (2) content-based image retrieval (CBIR) [2–4]. TBIR was first introduced in the early 1970s and it uses manually annotated words to describe images, which is a difficult, time-consuming, and tedious task when the size of the image database is large, and this method is also subject to problems related to visual perception [5,6]. To resolve the issues related to TBIR, CBIR was introduced in 1992 by Kato [7] and research in this domain of computer vision has continued for more than three decades. The need for more effective CBIR systems with high accuracy and low time costs has stimulated the development of improved CBIR systems. CBIR allows the user to retrieve images more efficiently from image databases by employing feature extraction and matching.

Feature extraction is characterized by the utilization of the color, shape, and texture of images, where it must be able to differentiate among images from the same and other classes with very small differences [8–11]. Furthermore, the features must be robust to geometric changes such as rotation, scaling, and translation, and photometric changes including differences in illumination and occlusion. Images are characterized using (1) global and (2) local approaches. In global approaches, images are characterized by ignoring the local and spatial information in the picture elements. The global approaches are computationally efficient and robust to noise to some extent, but they are not satisfactory at handling issues such as variations in illumination and occlusion. However, all of the problems with global approaches can be addressed by using local approaches where features are computed based on local patches, regions, or selected key points.

Shape is an important component used in image recognition and matching. In the present study, we propose a method based on shape information, and thus we focus on previous methods based only on shape information in the following. Several shape characterization and matching approaches have been proposed in previous studies. In general,

\* Corresponding author.

E-mail addresses: [ks\\_sathia@yahoo.com](mailto:ks_sathia@yahoo.com) (S. Sathiamoorthy), [sjpramoth@gmail.com](mailto:sjpramoth@gmail.com) (A. Saravanan).

<https://doi.org/10.1016/j.array.2020.100027>

Received 14 July 2019; Received in revised form 20 April 2020; Accepted 22 April 2020

Available online 1 May 2020

2590-0056/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

shape features are computed with: (1) boundary-or contour-based approaches, and (2) region-based approaches, where the former characterize the details of the shapes in images using the contours of an object and the latter use all of the pixels in a region [12–15]. We focus our discussion on boundary-based shape characterization methods.

Boundary-based shape characterization using chain code histogram [15,16] provides compact representations, translation invariant, preserving all of the morphological information. Shape signatures [17,18] such as cumulative angle, centroidal profile and chord length, and global shape descriptors such as the area, circularity, eccentricity, bending energy, convexity, major axis orientation, ratio of the principal axis, circular variance, and elliptic variance, and shape invariants were also employed in previous studies [19–21]. The contour point distribution histogram [22] comprises the distribution of points on an object contour under polar coordinates and it is a suitable approach for describing shapes with closed contours but not for images with multiple connected regions. Boundary moments were considered in some previous studies [23,24]. The curvature scale space approach employs a corner point detector to search for the curvature maxima or inflection points on the edges detected using the Canny approach [25]. Elastic matching involves an optimization problem based on the pixel to pixel correspondence between two images and it is robust to geometric deformation [26,27]. Eigen values are also invariant under rigid motion and scaling [28]. A shape-based non-redundant local binary pattern was presented by Yao and Chen [29]. The local maximum edge binary pattern suggested by Murala et al. [8] is computed based on the local differences among the center and eight neighborhood pixels, and it is combined with the Gabor transform to ensure its effectiveness. The local edge pattern proposed by Yao and Chen [30] uses the Sobel edge to compute the local edge pattern for segmentation and the local edge pattern for retrieval. Murala and Wu proposed peak valley edge patterns [31], where they obtained the local mesh peak valley edge pattern by including the first order derivative in the local mesh patterns [32]. The edge histogram descriptor reported by Jain and Vailya involves the distribution of the orientations of edgels, and the edges are computed using the Canny approach [33]. Another histogram method based on the edge distribution at the local level uses the Sobel operator for edge detection [34], although it obtains sub-standard retrieval results. Thus, a method was developed that uses the absolute locations of edge and the global composition to enhance the retrieval rate [35]. Jiebo et al. [36] reported a color edge co-occurrence histogram that calculates the distribution of the separation between pairs of color edges. A method based on the color distributions on directional and non-directional edges was suggested by Shim and Choi [37]. The block variation of local correlation coefficients method presented by Chen et al. identifies the edges and valleys in an image, which are characterized by first order moments [38,39], and a method based on the distributions of the edges and valleys was then proposed [40]. The edge orientation autocorrelogram method [41] computes the correlations among the edgels based on their orientation, and this method is robust to differences in illumination, viewpoint, translation, and small amounts of rotation. An enhanced version based on the edge histogram descriptor and edge orientation autocorrelogram methods [42,43] employs extremely minute and fine edges using a framework based on the full range autoregressive model for grayscale and color images, where the edges of color images are computed in the HSV space in order to avoid missing minute and fine edges due to changes in spectral and chromatic details.

Recently, Toshev et al. [44] proposed a chordigram image descriptor (CID) that computes the geometric details for a selected set of edgels obtained by segmentation. This method then employs geometric details comprising the distances among pairs of edgels, orientations of pairs of edgels, and the degree of the angle connecting pairs of edgels and the horizontal axis. However, the boundaries computed by segmentation might include fake edgels and this can affect the accuracy of results. Further computing the geometric details for every pair of edgels greatly increases the time cost. A subsequently developed method computes the

intensity and distance among each pair of edgels [45]. Moreover, a binary coherent edge descriptor was presented that characterizes the coordinates and orientation of each edgel, and the length of an edge that passes through the edgels [46].

Wang et al. proposed a new variant of CID that collects the distributions of the chord details for patches in images. The image is divided into a number of non-overlapping rectangular patches to reduce the influence of lighting. The CID method is robust to edge detector and it reduces the computational cost by employing predominant edgels. Statistical tests are conducted to identify the predominant edgels in each patch [47] and for every pair of predominant edgels in each patch, it is necessary to compute the distance among every pair of predominant edgels, the orientations of each predominant edgel in a pair, and the degree of the angle between a line segment among pair predominant edgels and the horizontal axis, before combining these geometric details in a local edgel chordigram (LEC) [47]. The ordered collection of LECs for all patch images comprises the CID. The CID is robust to differences in illumination, translation, and in-plane rotation, but it is affected considerably by noise. Thus, the patches with noise are eliminated during the matching operation by avoiding higher values in the similarity results obtained between the corresponding patches in the query and target images. CID is an appropriate method for place recognition with illumination changes, while the time cost is low and it can avoid fake edgels because edge detectors are used for edge identification instead of segmentation.

In a previous study, we enhanced the efficiency of this method by computing the CID using an autocorrelation function to obtain the autocorrelation-based CID (ACID) [48]. The ACID exploits the spatial correlation among identical predominant edgels at distance  $d$ , the orientation details for each predominant edgel in a pair of identical predominant edgels at distance  $d$ , and the degree of the angle along the line segment between a pair of identical predominant edgels and the horizontal axis. Our method neglects the length between a pair of predominant edgels because the length is always 1 in our approach. We demonstrated that ACID performs better than the conventional CID.

All of the previous approaches mentioned above compute the shape details for either a whole image or objects segmented from an image. However, previous studies have shown that the background and foreground details in images are concealed by both the global and local features [49], thereby resulting in poor retrieval performance because the user may be focused on objects in the background or foreground, or both. However, pinpointing the interests of users such as the background or foreground or even a specific object in the foreground is a challenging problem for the current CBIR approaches. Furthermore, separating the objects in the foreground and comparing them with the corresponding object in a target image is still a difficult issue for the existing CBIR approaches. At present, these problems are resolved using a relevance feedback approach where users are permitted to choose the images from the retrieval results obtained by a query and the selected images are then jointly employed to refine the query image until it corresponds subjectively to a user's needs in a particular search, and thus this process continues until the user is satisfied with the results [50]. Recently, machine learning has been combined with the relevance feedback approach to enhance the retrieval rate, but this approach also fails because of the low number of training images and the unwillingness of users to participate in the relevance feedback approach for a lengthy period of time [50].

Recently, Feng et al. [50] presented a CBIR where the salient edges and salient regions in an image are used as the retrieval targets because they coincide with the interests of users. This method uses the selective visual attention model to exploit the salient edges and salient regions in image, where the salient edge histogram (SEH) and salient adjacency graphs are computed and used jointly for CBIR to reduce the computation costs incurred to obtain both the local and global level image features. Another study [49] showed that only considering the features in the foreground can result in poor performance when the images have rich contents in the background, and thus the distinctive features are

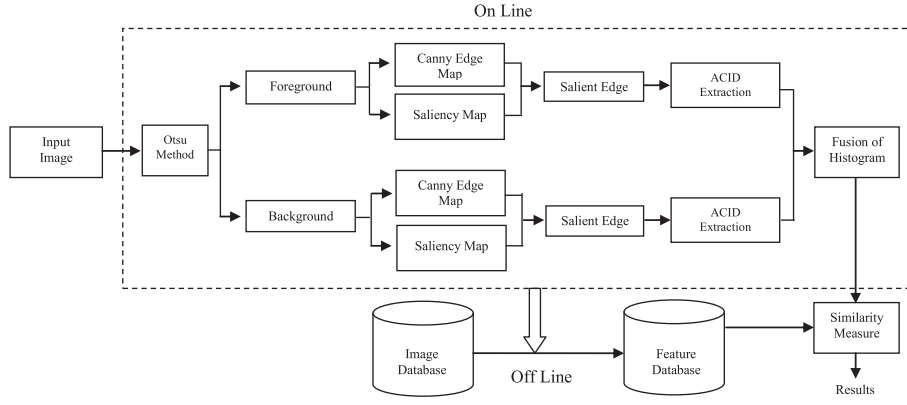


Fig. 1. Architecture of the proposed image retrieval approach.

computed in both the foreground and background in order to avoid them influencing each other, thereby obtaining hierarchical feature descriptions and achieving more accurate image matching.

In the method proposed in the present study, we computed the ACID based on the saliency edges for the foreground and background images, and obtain hierarchical feature descriptions by using the ACID to reduce the effects on each other of the foreground and background features. This method captures more geometric and its underlying texture, structure, and spatial details by considering a higher number of more responsive salient edges (25% of the total number of edgels) than the conventional method (15% of the total number of edgels). The proposed method selects more salient edges than the conventional approach in order to capture the rich underlying texture and structure information. By contrast, considering less salient edges will reduce the computational cost, but this method fails to capture much of the underlying texture and structure information among the salient edges. We experimentally evaluated the proposed approach by considering various subsets of salient edges where each subset varied in terms of the numbers of salient edges, and the response strengths of the salient edges were considered when selecting them for a subset. The experimental results demonstrated that considering 25% of the salient edges for extracting the rich underlying texture, structure and spatial information could obtain more accurate results, whereas reducing the number of salient edges significantly reduced the cost but it yielded less accurate retrieval results. Our proposed method employs the Otsu algorithm [49] to segregate the image into foreground and background details, while the Canny operator is used for edge detection and the selective visual attention model [50] to exploit the salient edges. We comprehensively tested the proposed approach based on benchmark databases and the results were compared with those obtained using CID [47], ACID [48], and SEH [50]. The proposed approach obtained more accurate results than CBIR. The proposed retrieval approach is an enhanced version of our previously reported method [48].

The remainder of this paper is organized as follows. In Section 2, we describe the approaches incorporated in the proposed CBIR. The experimental results and discussion are presented in Section 3. Finally, we give our conclusions in Section 4.

## 2. Feature extraction techniques

In the following, we provide overviews of the selective visual attention model, CID and ACID techniques, proposed CBIR, and feature descriptor matching method. The architecture of the proposed retrieval approach is illustrated in Fig. 1.

### 2.1. Selective visual attention model

Recently, Feng et al. [50] enhanced the saliency model and defined it

as a linear combination of the intensity contrasts in the Gaussian image pyramid. They reported that their approach can obtain more precise details from an image. Thus, let us assume that the image  $I$  is in the RGB color space and a mask with a size of  $3 \times 3$  is centered on a given pixel  $p$  in an image with  $M \times N$  dimensions in order to compute the saliency value at pixel  $p$  based on its adjacent neighbors as follows [50]:

$$SV(p) = \sum_{i=1}^L \sum_{p \in L \times L} \omega_c S_c^i(p, q) + \omega_o S_o^i(p, q) \quad (1)$$

where the number of levels ( $L$ ) in the pyramid is three,  $l$  represents the  $l$ th level in the pyramid image, and  $\omega_c$ ,  $\omega_o$ ,  $S_c(p, q)$ , and  $S_o(p, q)$  represent the weight coefficients, color intensity, and orientation contrasts between  $p$  and  $q$ , respectively. A Gabor filter with four orientations ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ) and four color channels comprising R, G, B, and Y [50] is used to compute the orientation and intensity contrast information. According to Feng et al. [50], the contrast at each level of the pyramid is computed and combined in a linear manner to frame the final saliency map, before a Gaussian filter with a standard deviation of 1 is applied to remove the noise.

### 2.2. Salient edge detection

A previous study showed that all of the edges identified by edge detectors are not valuable for characterizing an image [50]. Hence, several approaches have been reported for computing the salient edges. The term saliency is used to describe the difference between a pixel and those in its adjacent neighborhood [50]. Recently, Feng et al. [50] suggested a novel approach based on the visual attention model where the saliency of an edge is measured using its length and the saliency values around it.

According to Feng et al. [50], we employ the Canny operator to detect the edges as:

$$E^I = \{e_1^I, e_2^I, \dots, e_N^I\}, \quad (2)$$

where  $N$  denotes the total number of segments in an image  $I$ . The salient edges are computed as follows [50]:

$$SE(e_i^I) = \omega_L L(e_i^I) + \omega_S SA(e_i^I), i = 1, \dots, N, \quad (3)$$

where  $L(e_i^I)$  represents the length of edge  $e_i^I$ ,  $SA(e_i^I)$  represents the average saliency value for edge  $e_i^I$  based on the saliency map, and  $\omega_L$  and  $\omega_S$  represents the weights for  $L(e_i^I)$  and  $SA(e_i^I)$ , respectively, which are set to 0.3 and 0.7 [50]. The values of  $L(e_i^I)$  and  $SA(e_i^I)$  are normalized, and  $SA(e_i^I)$  is described as [50]:

**Table 1**

The representation of autocorrelation of identical predominant edgels at  $d = 1$ .

Edgel Valu	Distance (d)
	D = 1
255	0.019
254	0.075
253	0.102
.	.
.	.
.	.
.	.

$$SA(e_i^l) = \sum_{n=1}^{L(e_i^l)} \sum_{p \in \Theta_{p_n}^l} SV(p) / L(e_i^l), \quad (4)$$

Where  $\Theta_{p_n}$  represents a  $3 \times 3$  mask centered at pixel  $p_n^i$  and  $SV(p)$  is the saliency value of pixel  $p$ . After computing all of the saliency values for the edges, the following threshold operation is performed [50].

$$T_E = \max(SA(e_i^l)) / 4 \quad (5)$$

The final salient edge is then described as [50]:

$$\Theta_{SE} = \{e_i^l | SA(e_i^l) > T_E, i = 1, \dots, N\} \quad (6)$$

### 2.3. CID

A given image is divided into several non-overlapping rectangular patches numbered from 1, 2, 3, ..., N. Each patch is characterized using the LEC. To compute the LEC, the edges are identified using any edge operator and the prominent edgels are identified for each patch by applying a statistical hypothesis test, before the local geometric features are computed based on the prominent edgels. The LECs obtained for all patches are then collected in an order to construct a CID. The chord details for an image patch are computed based on each pair of prominent edgels coordinated at  $p$  and  $q$  as follows [47]:

$$C_{ipq} = (\ell_{pq}, \phi_{pq}, \theta_p, \theta_q), \quad (7)$$

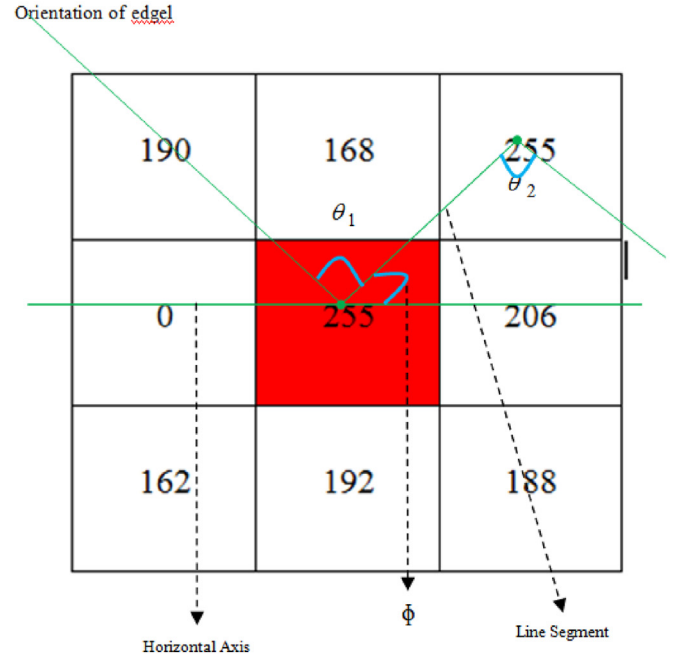
where  $\ell_{pq}, \phi_{pq}, \theta_p$ , and  $\theta_q$  denote the distance among predominant edgels  $p$  and  $q$ , the angle between the line connecting  $p$  and  $q$  and horizontal plane, and the degrees of orientation for predominant edgels  $p$  and  $q$  about the normal directions, respectively. The values of  $\ell_{pq}$  range from 0 to the diameter of an image patch, and  $\phi_{pq}, \theta_p$ , and  $\theta_q$  range from 0 to  $2\pi$ . The values of  $\ell_{pq}$  are discretized in the logarithmic space and they are divided into  $\eta_d$  bins where  $d$  is the distance, which is fixed to 4 [47].  $\phi_{pq}, \theta_p$ , and  $\theta_q$  are each quantized into eight bins [47]. Thus, CID is a four-dimensional histogram that is normalized as described previously [47], and it encompasses the chord details for every pair of prominent edgels in an image patch.

### 2.4. ACID

We propose a novel characterization approach for an image using ACID, which represents an improved version of the approach described by Wang et al. [47]. ACID computes the spatial correlations among the identical predominant edgels and explores the variations in the correlations at a distance  $d$  for each pair of the predominant edgels in a sub-image with a size of  $3 \times 3$  (one is at the center of the sub-image and other is an identical edgel at distance  $d$  from the center of the mask). We also compute the orientation of each predominant edgel and the degree of the angle along the line segment between the pair of predominant edgels and the horizontal plane. The autocorrelations among the identified identical predominant edgels are depicted in a table and indexed

228	218	0	0	0	0
0	255	196	189	0	0
0	255	255	198	202	0
0	0	255	190	168	255
0	235	184	0	255	206
0	244	0	162	192	188

**a**



**b**

**Fig. 2.** (a) Sample image of size  $6 \times 6$ , ( $3 \times 3$  mask is moved in non-overlapping manner. Center of  $3 \times 3$  mask with the salient edgel (value = 255) is marked in red color and its identical salient edgels at distance 1 is marked in blue color) 2 (b). Computation of orientation of edgels with value 255 and angle between the horizontal axis and line segment of edgels 255 and 255 at distance 1.

based on the edgel and distance values. An entry in the table denotes the probability of finding a predominant edgel with value  $i$  at a correlation distance  $d$  from a predominant edgel with value  $i$ . Thus, ACID considers the spatial correlation among the identical predominant edgels and it fixes the correlation distance to 1 because the lowest correlation distance provides the detailed local properties of an image [38]. Table 1 depicts



the proposed spatial correlations among identical predominant edgels at distance 1, where the first and second columns show the edgels with value  $i$  and the probability of finding an edgel with value  $i$  at correlation distance 1 from an edgel with value  $i$ , respectively. The autocorrelations among identical predominant edgels are described as follows [48].

Let  $I$  be an  $n \times n$  image and the predominant edgels in  $I$  are  $e_1, e_2, \dots, e_m$ . For an predominant edgel  $E = (x, y) \in I$ , let  $I(E)$  represent its edge value.

Let  $I_e \triangleq \{E | I(E) = e\}$ . Hence, the notation  $E \in I_e$  is synonymous with  $E \in I, I(E) = e$ . For convenience, we use the  $L_\infty$  norm to assess the distance among predominant edgels [48], i.e., for predominant edgels  $E_1 = (x_1, y_1), E_2 = (x_2, y_2)$ , we define  $|E_1 - E_2| \triangleq \max\{|x_1 - x_2|, |y_1 - y_2|\}$ . We denote the set  $\{1, 2, \dots, n\}$  by  $|n|$ .

Definitions. The histogram  $h$  of  $I$  is expressed for  $i \in |m|$  as follows.

$$h_{e_i}(I) \triangleq n^2 \Pr[E \in I_{e_i}] \quad (8)$$

For any predominant edgel in the image,  $h_{e_i}(I)/n^2$  is the probability that the value of predominant edgel is  $e_i$ .

Let a distance  $d \in |n|$  be fixed a priori. Then, the correlation of  $I$  is expressed for  $i, j \in |m|, k \in |d|$  as follows.

$$\gamma_{e_i, e_j}^{(k)} \triangleq \Pr_{E_1 \in I_{e_i}, E_2 \in I} [E_2 \in I_{e_j} | |E_1 - E_2| = k] \quad (9)$$

Given any predominant edgel of value  $e_i$  in the image,  $\gamma_{e_i, e_j}^{(k)}(I)$  is the probability that a predominant edgel at distance  $k$  from the given predominant edgel is of value  $e_j$ . The autocorrelation of  $I$  only exploits the spatial correlation among identical predominant edgels, and it is defined as [48]:

$$\alpha_c^{(k)} \triangleq \gamma_{e, e}^{(k)}(I) \quad (10)$$

To estimate the ACID, a  $3 \times 3$  non-overlapping mask is moved over an image from left to right and then from the topmost left corner of the image for each identified predominant edgel value. For instance, as shown in Fig. 2, when we move the  $3 \times 3$  mask over an image to compute the autocorrelation value of the predominant edgel with a value of 255, we obtain an identical predominant edgel with a value of 255 at the center of the first  $3 \times 3$  sub-image and it has two identical predominant edgels at distance 1. Thus, we estimate the autocorrelation using Eq. (10) and then estimate the orientation of each predominant edgel in every pair of predominant edgels (center and its identical predominant edgel) as  $\theta_1$  and  $\theta_2$  then compute the degree of the angle of the line segment between each pair of predominant edgels and the horizontal axis as  $\phi$ , as depicted in Fig. 2(b). Next, the  $3 \times 3$  mask is moved to the left and there is no predominant edgel with a value of 255 at the center of the mask. Thus, the mask is moved further. In the third  $3 \times 3$  sub-image, there is also no center edgel with a value of 255 but the fourth contains a predominant edgel in the center of the sub-image with a value of 255 and it has one identical predominant edgel at distance 1. Therefore, we estimate the autocorrelations among the identical predominant edgels for the fourth sub-image using Eq. (10), and we compute the values of  $\theta_1, \theta_2$ , and  $\phi$  [48]. This process continues for each predominant edgel value identified in the image. Finally, we obtain the autocorrelations of predominant edgels, as shown in Table 1, and the set of  $\theta_1, \theta_2$ , and  $\phi$  values are represented by separate histograms.

In our previous study [48], we demonstrated that ACID obtains a better retrieval rate because it exploits geometric and its underlying texture, structure, and spatial details and the computational cost is equivalent to that of the CID approach [47]. We also normalized the histogram for ACID to obtain more lighting variations [48].

Therefore, in the method proposed in the present study, we employ four-dimensional ACID for CBIR. In the proposed CBIR method, an image is segregated into foreground and background images, and ACID is then computed for the whole foreground and background images instead of computing it based on the patches in the whole image [47, 48]. Therefore,

the proposed approach avoids the concealment of details in the foreground and background, and the computational cost is also significantly reduced.

## 2.5. Proposed CBIR approach

Details are concealed in the foreground and background of an image, where each influences others to affect the performance of CBIR. Hence, in the proposed CBIR approach, the query images are divided into foreground and background details using the Otsu algorithm [49] in order to reduce the effects of the foreground and background on each other. The Canny operator is then applied to the foreground image to compute the edge map. Subsequently, the saliency map is computed for the foreground image using the selective visual attention model proposed by Feng et al. [50]. The saliency edges are computed using the edge and saliency maps for the foreground image. The Canny operator and the approach proposed by Feng et al. [50] are used to obtain the edge map and saliency map for the background images, which are then employed to compute the saliency edges. Based on a trial and error method, we showed that selecting 25% of the highly responsive salient edgels from all of the salient edgels in an image to compute the proposed feature obtains good performance.

Based on the saliency edgels selected in the foreground and background images, ACID is computed as follows (see Section 2.D).

1. Determine the spatial correlations among the identified identical edgels and explore the variations in the correlations with distance 1.
2. For each pair of salient edgels in a sub-image with a size of  $3 \times 3$  (one is at the center of the sub-image and other is an identical salient edgel at distance 1 from the center of the mask), ACID computes the orientation of each salient edgel.
3. Calculate the degree of the angle along the line segment between the pair of salient edgels and the horizontal plane.

Thus, the proposed ACID approach based on the selective visual attention model obtains a histogram with four dimensions. In the proposed CBIR, the ACIDs in the foreground and background are combined in a hierarchical manner to describe the image, and thus the proposed CBIR exploits both ACIDs to reduce the influence of the foreground and background images on each other in CBIR. The algorithm for the proposed CBIR approach is described as follows.

**Input:** Image

**Output:** Retrieved images

1. Divide the image into foreground and background images using the Otsu approach.
2. Obtain the edge map using the Canny operator for both the foreground and background images.
3. Obtain saliency maps for both the foreground and background images using the Gaussian image pyramid.
4. Compute saliency edges for both the foreground and background images.
5. Compute ACIDs for the foreground and background images using the strongly responsive saliency edges.
6. Combine the ACIDs for the foreground and background images in a hierarchical manner.
7. Measure the similarity between feature vectors in the database and query image.
8. Sort the similarity values between the query and all of the images in database.
9. Result.

## 2.6. Feature matching method

Next, we present an effective approach for feature matching by computing the similarity between the query image and target image. Two

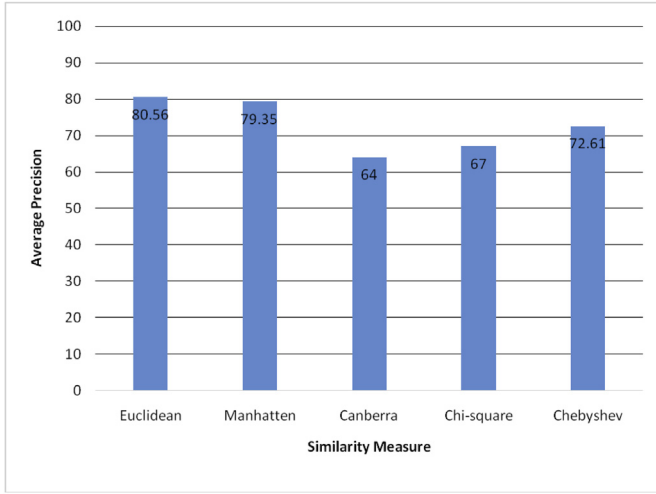


Fig. 3. Average retrieval accuracy attained by various distance measures for the proposed feature descriptor.

images are related to each other when the similarity between the two images is a small value. Various similarity metrics can be employed to compute the similarity of two images, but the Euclidean method is more familiar and it is used extensively for CBIR. In the proposed method, after computing the features locally in the form of four-dimensional histograms, the similarity between the query and target images is computed with the following equation [50–52]:

$$S(Q, T) = \sqrt{\sum_{i=0}^N (|Q_i - T_i|)^2}, \quad (11)$$

where  $T$  and  $Q$  denote the input image and target image feature descriptor, respectively, and  $N$  is the number of descriptors in the image feature vector. We also compared the performance of other distance measures [50–52] such as the Manhattan, Canberra, Chi-square, and Chebyshev metrics in terms of the average precision, and the results demonstrated that worst performance was obtained with the Canberra metric and the highest performance using the Euclidean distance metric. The Manhattan metric performed significantly worse than the Euclidean metric. Thus, the Euclidean distance is used in our proposed CBIR approach. Fig. 3 shows the average recall results with various distance measures.

### 3. Experimental results and discussion

In the experiments, we compared selected methods comprising CID [47], ACID [48], and SEH [50] with the proposed approach. We evaluated the performance of these approaches based on the Gardens Point Walking [47], University of Alberta (UA) Campus [47], St Lucia [47], Corel 10 k [54], and self-photographed image databases. The superior performance of the proposed approach was validated by evaluating the precision and recall, and based on comparisons with the other approaches with all five image databases. The precision defines the relationship among the total number of related images retrieved for a given input image and the total number of images retrieved from the database, which gradually decreases as we increase the number of retrieved images. The precision is defined as follows [48].

$$\text{Precision}(P) = \frac{\text{Total No. of correct images retrieved from database}}{\text{Total No. of images retrieved from the database}} \quad (12)$$

Another common measure used for computing the accuracy is the recall, which is defined as the probability of retrieving a correct related

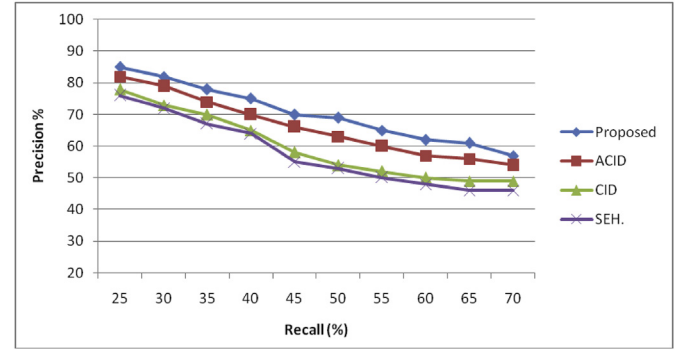


Fig. 4. Precision Vs recall of the proposed and existing approaches for Gardens Point Walking database.

image for a given query and it increases as the number of retrieved images increases. The recall is defined as follows [48].

$$\text{Recall}(R) = \frac{\text{Total No. of correct images retrieved from database}}{\text{Total No. of images relevant images in the database}} \quad (13)$$

The average precision is computed as [48]:

$$P_{\text{Avg}}^C = \frac{1}{N} \sum_{i=1}^N P, \quad (14)$$

and the average recall is computed as [48]:

$$R_{\text{Avg}}^C = \frac{1}{N} \sum_{i=1}^N R, \quad (15)$$

where  $c$  and  $N$  denote the class and total number of images in the database, respectively. Thus, the total precision and total recall are computed as follows.

$$P_{\text{total}} = \frac{1}{C} \sum_{i=1}^C P_{\text{Avg}}^C \quad (16)$$

$$R_{\text{total}} = \frac{1}{C} \sum_{i=1}^C R_{\text{Avg}}^C \quad (17)$$

To measure the effectiveness of the methods, a query image was selected from each benchmark database and the top 100 images retrieved by CID [47], ACID [48], SEH [50], and the proposed approach were considered. All of the images in the databases were used as queries to assess the retrieval performance. The state-of-the-art methods and the proposed technique were implemented on a PC with an Intel Pentium Core 2 Duo 2.10 GHz processor and 2 GB RAM.

The proposed approach was compared with the state-of-the-art descriptors based on the Gardens Point Walking database, which contains three groups of images recorded by pedestrian. The viewpoint variations were taken at day (by walking on both left and right) and night (footage on right). Each group contains 200 images captured twice during the day and night, and the sizes of the images are approximately  $960 \times 540$  pixels. The precision and recall curves shown in Fig. 4 demonstrate the performance of the different descriptors. Clearly, SEH obtained the worst performance among all of the descriptors and ACID performed better than CID and SEH. However, ACID based on the application of the selective visual attention model for the foreground and background images in a hierarchical manner obtained much higher accuracy than the CID, ACID, and SEH descriptors.

We then conducted an experiment based on the UA Campus database where we considered the sequences collected at three different times of

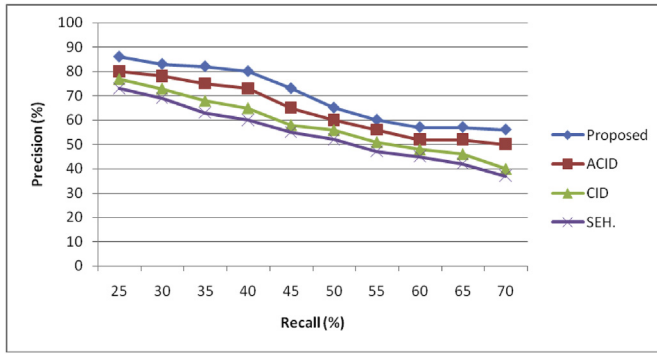


Fig. 5. Precision Vs recall of the proposed and existing approaches for UA database.

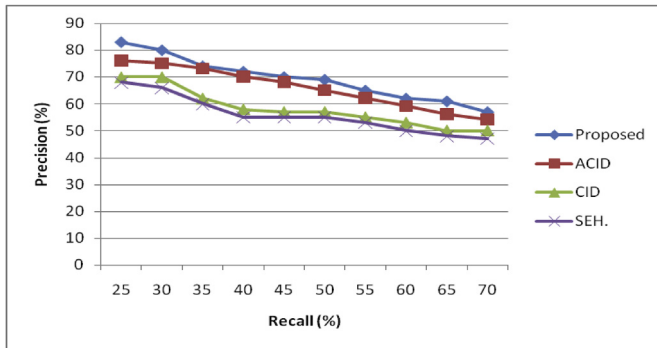


Fig. 6. Precision Vs recall of the proposed and existing approaches for St.Lucia database.

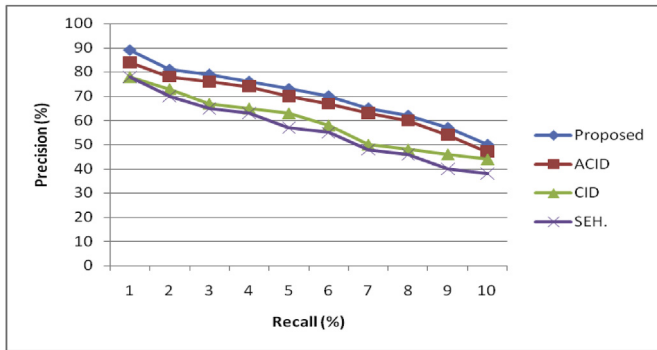


Fig. 7. Precision Vs recall of the proposed and existing approaches for Corel 10 k database.

day, i.e., at 06:20, 16:40, and 22:15, where each sequence contained 630 frames [47]. The results obtained by applying CID, ACID, SEH, and the proposed ACID based on the selective visual attention model to the foreground and background images in a hierarchical manner were relatively similar to the results obtained with the Gardens Point Walking database. The proposed approach performed better than ACID, while the performance of CID was intermediate and that of SEH was again the worst. The corresponding precision and recall plots are depicted in Fig. 5.

Next, we compared the performance of the different descriptors based on the St. Lucia database. The precision and recall curves obtained for the proposed and existing approaches are depicted in Fig. 6. The St Lucia database contains images acquired at five different times between the early morning and late afternoon during the day, and during the day after two weeks, where it comprises 10 groups of images. In the experiments, the worst performance was obtained using SEH and the best performance

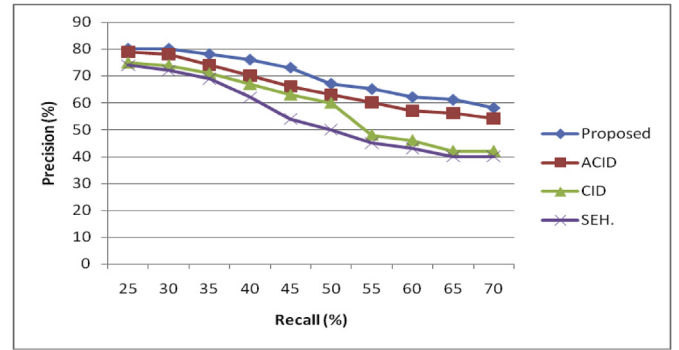


Fig. 8. Precision Vs recall of the proposed and existing approaches for Self-photographed image database.

Table 2

Retrieval performance of the proposed, ACID, CID and SEH.

Datasets	Proposed	ACID	CID	SEH
Gardens Point Walking	74.23	72.10	68.93	67.89
UA Campus	72.47	70.88	62.91	60.78
St. Lucia	69.32	65.83	58.25	55.71
Corel 10 k	71.03	67.31	59.20	56.13
Self Photographed Images	70.00	65.70	58.80	54.90

with the proposed method, while CID obtained intermediate performance and ACID performed well according to the precision and recall plots. The proposed approach is an enhanced version of our previously presented method [48] combined with that described by Wei et al. [49], and it performed better than the existing approaches with the St.Lucia database.

The subsequent experiment was conducted based on the Corel 10 k database, which is has been utilized by many researchers for CBIR assessments. The Corel 10 k database contains 10000 images in 100 classes, including crab, rhino, panda, cup, tiger, door, fitness, bob, dish, and flags, and each class comprises 100 images. The sizes of the images are approximately  $192 \times 128$  or  $128 \times 192$  pixels. The precision and recall curves are shown in Fig. 7. Clearly, SEH obtained the worst performance, while ACID performed better than CID and SEH, but the highest performance was achieved using the proposed approach.

We also tested the performance of the proposed method, CID, ACID, and SEH based on a self-photographed images database and the corresponding results in terms of the precision and recall curves are illustrated in Fig. 8. The self-photographed images were acquired at various times during the day and they comprised 1140 images with sizes of approximately  $1280 \times 720$  and  $1040 \times 780$  pixels. The results clearly demonstrated that the proposed descriptor obtained better retrieval performance than all of the other descriptors.

The results demonstrated that the proposed approach was more robust to differences in illumination and occlusion compared with CID and SEH. The proposed approach and ACID exhibited similar robustness to differences in illumination and occlusion. However, the proposed approach obtained better retrieval performance than CID, ACID, and SEH. The low retrieval accuracy of SEH is due to computation at the global level and a failure to capture the geometric details of salient edges. CID obtained good accuracy but it ignores the underlying texture, structure and spatial information of dominant edges, which is important for CBIR. Thus, the retrieval accuracy was reduced with CID. ACID considers the geometric and its underlying texture, structure and spatial details of dominant edges, but foreground and background details are concealed in images and the method employed to determine the dominant edges for computing the ACID leads to poor performance. These problems with the state-of-the-art methods are addressed in the proposed approach, and thus it obtains significantly better retrieval performance.



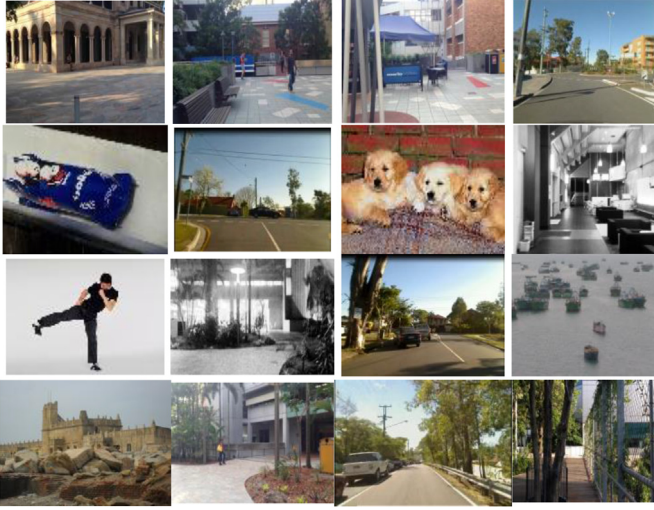


Fig. 9. Few sample images of experimental databases.

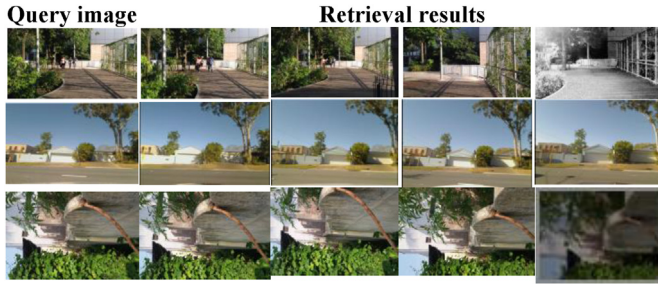


Fig. 10. For instance, top 4 retrieval results for Gardens point walking, St. Lucia and self photographed image databases.

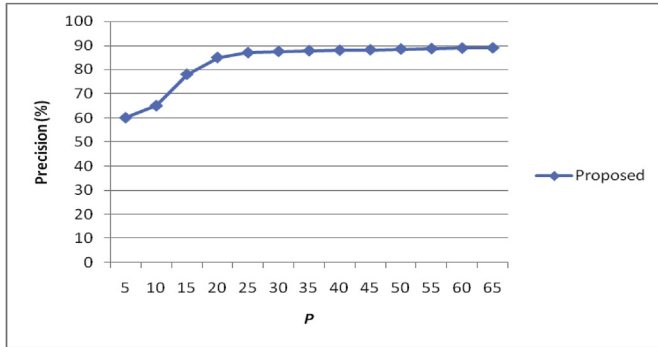


Fig. 11. Average precision Vs various P% of salient edges.

Table 2 show the average precision and recall values using the proposed descriptor and the ACID, CID, and SEH approaches based on the Gardens Point Walking, St Lucia, UA Campus, Corel 10 k, and self-photographed images databases. Examples of images in all of the benchmark databases are shown in Fig. 9. In addition, examples of the retrieval results obtained by the proposed approach based on the Gardens Point Walking, St Lucia, and self-photographed images databases are presented in Fig. 10.

In the experiments, we also computed the proposed feature using various subsets of salient edges, where each subset (S) varied in terms of the number of salient edges. In particular, if N salient edges are present in an image, then the subset S contains  $P \times N$  salient edges, where P is the proportion used to determine the size of the subset. In the experiments, we started with  $P = 5\%$ , i.e., 5% of the stronger responsive salient

Table 3

Computational complexity of proposed, ACID, CID and SEH..

Methods	Time in seconds
SEH	47
CID	56
ACID	45
Proposed	52

edges were selected initially, and we then increased P by 5% incrementally. For each value of P, we computed the proposed feature and measured the performance in terms of the accuracy and time cost. The best performance was obtained when  $P = 25\%$ . In the experiments, we changed the value from  $P = 5\%$ – $100\%$  and the accuracy increased gradually to  $P = 25\%$ , but there were no further changes in the performance with higher values when we increased P, although the computational cost increased, as shown in Fig. 11. Therefore, the results clearly showed that the proposed approach could extract rich geometric and its underlying texture, structure and spatial information when  $P = 25\%$ .

Any descriptor must perform effectively and the computational cost should also be low. Therefore, the central processing unit (CPU) times required by the proposed method, CID, SEH, and ACID were computed by extracting the features offline (to create a feature database) and online (retrieval). The average CPU times required by the proposed method and the CID, SEH, and ACID descriptors are shown in Table 3. SEH required the least CPU time whereas the proposed approach consumed the most CPU time, and it was slightly higher than that by ACID and slightly lower than that by CID. However, the fairly high time cost is acceptable because of the high accuracy of the proposed method. Thus, the proposed approach exhibits efficient retrieval performance and it is robust to differences in illumination and occlusion.

#### 4. Conclusion

In this study, we developed a novel shape-based approach for image retrieval. In contrast to the previously proposed CID descriptor and its variants, the proposed descriptor splits the image into foreground and background images using the Otsu approach, before computing the saliency and edge maps for the foreground and background images. Next, the salient edges computed using the saliency and edge maps are employed to obtain the ACID by using the salient edges in the foreground and background images in order to avoid missing concealed information, which would influence the characterization of both the foreground and background images. Furthermore, the proposed approach considers a higher amount of salient edges (25%) for feature computation to capture more of the rich underlying texture, structure and spatial information than the ACID and CID descriptors. Therefore, the information extracted by the proposed approach is more accurate and it is more robust to changes in illumination and occlusion than ACID and CID. However, the computational cost of the proposed approach is significantly higher than that of ACID, although this difference is negligible due to the higher retrieval accuracy. Experiments conducted based on five databases confirmed that the proposed descriptor can efficiently characterize the shape details and obtain better retrieval results compared with the state-of-the-art techniques. In future research, weights can be assigned to the contour-based details determined in the foreground and background images during the feature matching phase based on the saliency details identified.

#### Credit author statement

**S.Sathiamoorthy:** Conceptualization, methodology, writing - original draft, writing - review and editing, supervision. **A.Saravanan:** Formal analysis, software and investigation, data curation. **R.Ponnusamy:** Resources, software and investigation, data curation, validation.



## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] Ma H, Zhu J, Lyu MRT, King I. Bridging the semantic gap between image contents and tags. *IEEE Trans Multimed* 2010;12(5):462–73.
- [2] Feng S, Xu D, Yang X. Attention-driven salient edge (s) and region (s) extraction with application to CBIR. *Signal Process* 2010;90(1):1–15.
- [3] Wang M, Ye Z-L, Wang Y, Wang S-X. Dominant sets clustering for image retrieval. *Signal Process* 2008;88(11):2843–9.
- [4] Murala S, Maheshwari R, Balasubramanian R. Local tetra patterns: a new feature descriptor for content-based image retrieval. *IEEE Trans Image Process* 2012;21(5):2874–86.
- [5] Long F, Zhang H, Feng DD. Fundamentals of content-based image retrieval. In: *Multimedia information retrieval and management*. Springer; 2003. p. 1–26.
- [6] Zhang X, Zhao X, Li Z, Xia J, Jain R, Chao W. Social image tagging using graph-based reinforcement on multi-type interrelated objects. *Signal Process* 2013;93(8):2178–89.
- [7] Androustos P, Kushi A, Plataniotis KN, Venetsanopoulos AN. Aggregation of color and shape features for hybrid query generation in content based visual information retrieval. *Signal Process* 2005;85(2):385–93.
- [8] Subrahmanyam M, Maheshwari R, Balasubramanian R. Local maximum edge binary patterns: a new descriptor for image retrieval and object tracking. *Signal Process* 2012;92(6):1467–79.
- [9] Zhang J, Ye L. Local aggregation function learning based on support vector machines. *Signal Process* 2009;89(11):2291–5.
- [10] Qian Y, Hui R, Gao X. 3D CBIR with sparse coding for image-guided neurosurgery. *Signal Process* 2013;93(6):1673–83.
- [11] Hu MK. Visual pattern recognition by moment invariants. *IRE Trans. Info. Theory* 1962;8:179–87.
- [12] Lu G, Sajjanhar A. Region based shape representation and similarity measure suitable for content-based image retrieval. *Multimed Syst* 1999;7(2):165–74.
- [13] Zhang D, Lu G. Shape-based image retrieval using generic Fourier descriptor. *Signal Process Image Commun* 2002;17(10):825–48.
- [14] Kim W-Y, Kim Y-S. A region based shape descriptor using Zernike moments. *Signal Process Image Commun* 2000;16:95–102.
- [15] Freeman, H., Saghi, A., November 7–10, 1978. Generalized chain codes for planar curves, in: *Proceedings of the fourth international joint conference on pattern recog.* Kyoto, Japan, pp. 701–703.
- [16] Iivarinen J, Visa A. Shape recognition of irregular objects. In: *Casasent DP, editor. Intelligent robots and computer vision XV: algorithms, techniques, active vision, and materials handling*. Proc; 1996. p. 25–32. SPIE 2904.
- [17] Otterloo PJV. A contour-Oriented Approach to shape analysis. Englewood Cliffs, NJ: Prentice-Hall International (UK) Ltd; 1991. p. 90–108.
- [18] Zhang DS, Lu G, January 22–25. A comparative study of Fourier descriptors for shape representation and retrieval. In: *Proceedings of the fifth asian conference on computer vision*. Melbourne, Australia: ACCV02; 2002. p. 646–51.
- [19] Yong J, Bowie Walker, J. An analysis technique for biological shape. *Comput Graph Image Process* 1974;25:357–70.
- [20] Peura M, Iivarinen J, May. Efficiency of simple shape descriptors. In: *Proceedings of the third inter. Workshop on visual form*. Italy: Capri; 1997. p. 443–51.
- [21] Huang C-L, Huang D-H. A content-based image retrieval system. *Image Vis Comput* 1998;16:149–63.
- [22] Shu X, Wu X-J. A novel contour descriptor for 2D shape matching and its application to image retrieval. *Image Vis Comput* 2011;29(4):286–94.
- [23] Sonka M, Hlavac V, Boyle R. *Image processing, analysis and machine vision*. London, UK, NJ: Chapman & Hall; 1993. p. 193–242.
- [24] Gonzalez RC, Woods RE. *Digital image processing*. Reading, MA: Addison-Wesley; 1992. p. 502–3.
- [25] Mokhtarian F, Abbasi F, Kittler J. Efficient and robust retrieval by shape content through curvature scale space. *Int. workshop on Image Databases and Multi-Media Search*; 1997. p. 51–8.
- [26] Del Bimbo A, Pala P. Visual image retrieval by elastic matching of user sketches. *IEEE Trans Pattern Anal Mach Intell* 1997;19(2):121–32.
- [27] Attalla E, Siy P. Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. *Pattern Recognit* 2005;38:2229–41.
- [28] Squire DM, Caelli TM. Invariance signature: characterizing contours by their departures from invariance. *Comput Vis Image Understand* 2000;77:284–316.
- [29] Nguyen DT, Ogunbona PO, Li W. A novel shape-based non-redundant local binary pattern descriptor for object detection. *Pattern Recogn* May 2013;46(5):1485–500.
- [30] Yao CH, Chen SY. Retrieval of translated, rotated and scaled color textures. *Pattern Recogn* 2003;36(4):913–29.
- [31] Murala S, Wu QM. Peak valley edge patterns: a new descriptor for biomedical image indexing and retrieval. In: *IEEE conference on computer vision and pattern recognition workshops*. CVPRW; 2013. p. 444–9.
- [32] Murala S, Wu QJ. MRI and CT image indexing and retrieval using local mesh peak valley edge patterns. *Signal Process. Image Commun* 2014;29(3):400–9.
- [33] Jain AK, Vailaya A. Image retrieval using color and shape. *Pattern Recogn* 1996;29(8):1233–44.
- [34] Cieplinski L, Kim M, Ohm J-R, Pickering M, Yamada A, editors. *Text of ISO/IEC 15938-3/PCD information technology -multimedia content description interface-Part 3: visual*. Final committee draft; 2001. ISO/IEC/JTC1/SC29/WG11 (MPEG), document no. N4062.
- [35] Chee SunWon DongKwonPark, Soo JunPark. EfficientUseofMPEG7 edge histogram descriptor ; 2002.
- [36] Luo J, Crandall D. Color object detection using spatial-color joint probability functions. *IEEE Trans Image Process* 2006;15(6):1443–53.
- [37] Shim Seong-O, Choi Tae-Sun. Edge color histogram for image retrievalvol. 3. Rochester,NY, USA: Proc. Int.Conf.onImageProc.; 2002. p. 957–60. <https://doi.org/10.1109/ICIP.2002.1039133>.
- [38] Chun YD, Seo SY, Kim NC. Image retrieval using BDIP and BVLC moments. *IEEE Trans Circ Syst Video Technol Sep.* 2003;13:951–7.
- [39] Chun YD, Kim NC, Jang IH. Content-based image retrieval using multiresolution color and texture features. *IEEE Trans Multimed* 2008;10(6):1073–84.
- [40] Sathiamoorthy S, Kamarasan M. A novel approach for image retrieval using BDIP and BVLC. *Int. Journal of Innovative Research in Computer and Communication Engineering* 2014;2(9):5897–902.
- [41] Mahmoudi F, Shanbehzadeh J, Eftekhari AM, Soltanian-Zadeh H. Image retrieval based on shape similarity by edge orientation autocorrelation. *Pattern Recogn* 2003;36:1725–36.
- [42] Seetharaman K, Sathiamoorthy S. An improved edge direction histogram and edge orientation autocorrelation for an efficient color image retrieval. In: 2013 international conference on advanced computing and comm. Systems. India: Coimbatore; 2013. p. 1–4. <https://doi.org/10.1109/ICACCS.2013.6938725>.
- [43] Seetharaman K, Sathiamoorthy S. A unified learning framework for content based medical image retrieval using a statistical model. *Journal of King Saud University - Computer and Information Sciences* 2016;28(1):110–24.
- [44] Toshev B, Taskar, K, Daniilidis. Shape-based object detection via boundary structure segmentation. *Int J Comput Vis* 2012;99(2):123–46.
- [45] Kovalev S, Volmer. Color co-occurrence descriptors for querying-by example. In: *Multimedia modeling, 1998. MMM'98. Proceedings*. 1998. IEEE; 1998. p. 32–8.
- [46] Zitnick CL. Binary coherent edge descriptors. In: *European conference on computer vision*. Springer; 2010. p. 170–82.
- [47] Wang Xiaolong, Zhang Hong, GuohuaPeng. A chordigram image descriptor using local edgels. *J Vis Commun Image Represent* 2017;49:129–40.
- [48] Saravanan A, Sathiamoorthy S. Autocorrelation based chordigram image descriptor for image retrieval. In: 4th international conference on communication and electronics systems (ICCES 2019). Coimbatore, India: PPG Institute of Technology; 2019. p. 1990–6. <https://doi.org/10.1109/ICCES45898.2019.9002528>. July 17–19.
- [49] Song Wei, Zhang Yubing, Liu Fei, Chai Zhilei, Ding Feng, Qian Xuezhong, Park Soon Cheol. Taking advantage of multi-regions-based diagonal texture structure descriptor for image retrieval. *Expert Syst Appl* 2017. <https://doi.org/10.1016/j.eswa.2017.12.006>.
- [50] Feng S, et al. Attention-driven salient edge(s) and region(s) extraction with application to CBIR. *Signal Process* 2009. <https://doi.org/10.1016/j.sigpro.2009.05.017>.
- [51] Sharma Pooja. Improved shape matching and retrieval using robust histograms of spatially distributed points and angular radial transform. *Optik* 2017;145:346–64.
- [52] Liu Guang-Hai, Yang Jing-Yu, et al. Content-based image retrieval using computational visual attention model. *Pattern Recogn* 2015;48(8):2554–66.