Boosting Content Based Image Retrieval Performance Through Integration of Parametric & Nonparametric Approaches

Soumya Prakash Rana[a], Maitreyee Dey[a], Patrick Siarry[b]

[a] *London South Bank University, 103 Borough Rd, London SE1 0AA, United Kingdom*
[b] *University of Paris-Est Creteil, 61 Avenue du General de Gaulle, 94000 Creteil, Paris, France*

Abstract

The collection of digital images is growing at ever-increasing rate which rises the interest of mining the embedded information. The appropriate representation of an image is inconceivable by a single feature. Thus, the research addresses that point for content based image retrieval (CBIR) by fusing parametric color and shape features with nonparametric texture feature. The color moments, and moment invariants which are parametric methods and applied to describe color distribution and shapes of an image. The nonparametric ranklet transformation is performed to narrate the texture features. Experimentally these parametric and nonparametric features are integrated to propose a robust and effective algorithm. The proposed work is compared with seven existing techniques by determining statistical metrics across five image databases. Finally, a hypothesis test is carried out to establish the significance of the proposed work which, infers evaluated precision and recall values are true and accepted for the all image database.

Keywords: CBIR, Color Moments, Ranklet Transform, Nonparametric Statistics, Moment Invariants, Hypothesis Test

1. Introduction

With the explosion of digital technologies and greater storage capabilities, vast volumes of digital media now exist in various fields like, multimedia and spatial information system [1], medical image [2], time series data analysis [3], compression techniques [4]. When a required image is being located, normally employed methods are via keyword indexing or by simply browsing, which can be very time consuming and may not result in the exact image sought. This necessitate the development of an efficient algorithm for managing, indexing and searching these large image libraries. There are two types of image searching techniques, text-based image retrieval (TBIR) and content-based image retrieval (CBIR). TBIR relies on the manual search by keyword matching of existing image titles. The outcomes depend upon the human labelling which leads to irrelevant results and wastage of time [5] whereas, CBIR technique relies on low-level image features and reduces human labour drastically. However, these features are failed to represent a required image.

Email addresses: ranas9@lsbu.ac.uk (Soumya Prakash Rana), deym@lsbu.ac.uk (Maitreyee Dey), siarry@u-pec.fr (Patrick Siarry)

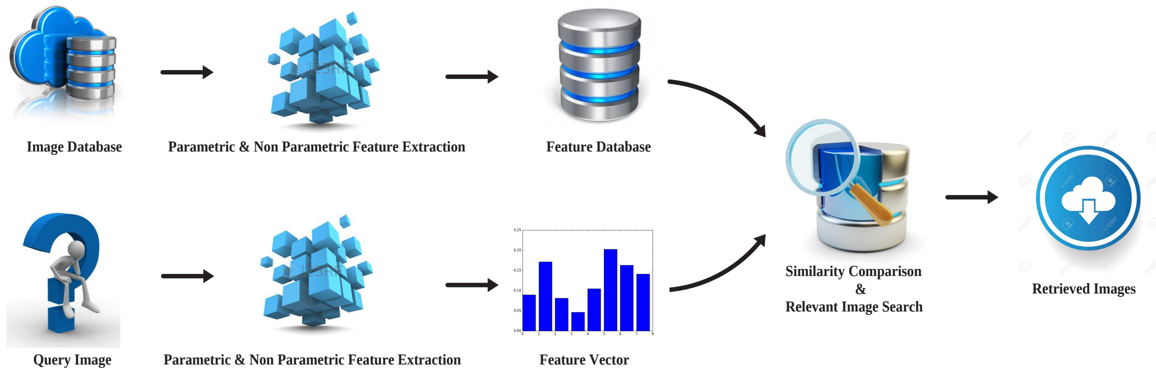

Figure 1: CBIR system structure.

CBIR is an image searching method from large image database depending upon extracted features by the visual descriptors. Here, user asks for an image and CBIR system searches for relevant images based on extracted features from stored images of the databases. This system structure of CBIR is described in Figure 1 where, user makes query, then visual content descriptor or feature extractor extracts low level features (e.g., color, texture, and shape) from the query image. Thus, CBIR is also known as feature based image retrieval (FBIR) [6, 7]. Then the distance is measured between query or example image with the feature vectors of stored images to find out the similarity and retrieve images on a suitable match. In some cases, researchers proposed region based image retrieval (RBIR), an extended version of CBIR where, images are segmented into different regions and features are extracted from each regions to represent an image. Unfortunately, the similarity measurement cost is very high in this case which, restricts RBIR from wide acceptance [8]. Therefore, the performance of a CBIR system heavily depends upon the low-level image processing or the extraction of fundamental image primitives. These primitives are derived to represent the images in such a way, that could express the query of a human mind properly through the numerical form. This is a challenging task over the decades for the researchers of CBIR field to express human concepts through features for mining the other similar images. Hence, the minimization of the gap of similarity between an example image and the retrieved images by the CBIR method is the main goal where, it can help user in various domains such as, image searching, browsing, remote sensing, crime prevention, publishing, medicine, architecture, historical research, etc.

1.1. Related Work

CBIR techniques most commonly employ visual color, texture and shape information. These existing feature extraction methods are relied on global and local features. Global features cover the whole image as visual content, whereas local feature based algorithms focus on key points or selected regions of a whole image. Several algorithms exist on global and local features extraction. Color is a wavelength-dependent perception [9] which has evolved as a widely used visual character for image retrieval and object recognition. Color histogram is a well known color descriptor which is invariant to orientation and scale. These properties make it a powerful image classification technique. Image retrievals based on this method were easy to implement and commonly used in CBIR fields. However, color histogram faces difficulties when, characterizing images with spatial structures. Subsequently, a number of color descriptors aimed at exploiting spatial information have been proposed. Those are mainly compact color moments, color coherence vector, and color autocorrelograms [10]. For example, in the MPEG-7 standard, color descriptors comprise histograms such as, color layout descriptor, dominant color descriptor, and scalable color descriptor [11, 12]. An indexed and encoded explanation of color information is utilized for improvement of CBIR results. It actually discards unimportant colors and keeps important color information to make indexed color histogram, which is further represented by golomb-rise (GR) encoding [13].

One of the pivotal characteristics of an image is texture which is frequently employed in CBIR systems. Texture analysis uses various algorithms e.g., gray level co-occurrence matrices (GLCM) [14], tamura texture feature [15], the markov random field (MRF) model [16], gabor filtering [17], and local binary patterns (LBP) [18]. Three types of texture descriptor are utilized in MPEG-7 standard; i.e., edge histogram descriptor, homogeneous texture descriptor, texture and browsing descriptor [11, 12]. Generally,these texture features are combined with color features to improve discrimination power and to enhance retrieval performance. In most of the cases, texture extraction algorithms combine color and gray-level texture features such as, multi-texton histogram [19], texton co-occurrences matrix [20], micro-structure description [21], and color edge co-occurrence histogram [22].

Shape features are also additionally applied in many research work like color and texture features. This simply works on the fact that humans can distinguish objects solely by their shapes. Classical methods to describe shape features include the use of moment invariants, fourier transform coefficients, edge curvature and arc length [23, 24]. MPEG-7 employs three shape descriptors for object-based image retrieval such as, 3D shape descriptor, region-based shapes derived from zernike moments, and the curvature scale space (CSS) descriptor [12]. In addition, local image feature extraction (LIFE) is also gaining the attention. One popular LIFE technique is scale invariant feature transform (SIFT) [25], which can tolerate some illumination change, perspective distortion, and transformation. It is also quite robust to occlusion issues. Another well known LIFE is the bag-of-visual words, which are derived from local features such as, key points and salient patches. This was actually proposed for object recognition and scene categorization [26]. Essentially, this algorithm is highly motivated by retrieval methods. This method has some limitations e.g., it heavily relies on computation power since it uses clustering techniques. Also, insufficient semantic information, text ambiguity, and long feature lengths makes it deficient. In practice, the classification accuracy of text words is far superior than visual words.

Apart from the individual use of color, texture, and shape, high number of CBIR works rely on multi-feature fusion to have better performance. Distribution of Color Ton (DCTon) [27] is a hybrid CBIR framework to extract the inter class features using color distribution with the help of dual tree complex wavelet transformation and singular value decomposition (SVD). It represents both color and texture information. Global correlation descriptor (GCD) [28] is proposed to extract color and texture features to enhance image retrieval performance. It has two sub-parts, global correlation vector (GCV) and directional global correlation vector (DGCV) which are using the advantages of histogram statistics and structure element correlation (SEC) to express color and texture features respectively. Multi-trend structure descriptor (MTSD) [29], describes color, shape, texture, and local spatial structure information to define an image in CBIR. It uses local structures of images to explore correlation between pixels. Color, edge orientation, and intensity mapping are considered to build the model. Srivastava and Khare [30] proposed a method using discrete wavelet transform (DWT), and local binary pattern (LBP) with legendre moments (LM) for refining the retrieval performance where, images are converted to gray scale for different levels of decomposition and then LBP extracts texture features from decomposed elements. A hybrid textual-visual relevance is used in [31] which, mines image tags, combines textual relevance and visual relevance for image retrieval. It is actually followed by correction of missing tags, capturing user's semantic cognition and finally a probability distribution on the permutations of tags are executed. Finally, instead of early fusion, a ranking aggregation strategy is acquired to sew up textual relevance and visual relevance seamlessly.

1.2. Scope & contributions

Though, the aforementioned CBIR literature is quite strong still it needs improved computation due to the recent advancement. This includes availability of vast image database, fast computation for quick and relevant image retrieval. Thus, the proposed CBIR study is motivated by the fact of appropriate image retrieval for user search. Existing research is focused on: parametric features, nonparametric features, fusion of different parametric features, and fusion of different nonparametric features. The parametric approaches derive features by making assumption of pixel distribution where, nonparametric approach doesn't make assumption of pixel distribution and determines ranking from available numerical values [32]. Hence, a CBIR system which, employs only parametric color features for image retrieval faces problem to characterize an image containing more significant texture and shape information. In the same way, CBIR systems which only use nonparametric shape extraction method to describe an image are failed to describe color and texture information. Therefore, a hybrid CBIR method is proposed here that intends to depict an image through all three kinds of low level image features (color, texture, and shape) with minimal information lose which would be beneficial for this domain. This work is integrated color moments (parametric method), ranklet transformation (nonparametric method), and moment invariants (parametric method) to develop an efficient and robust CRM method. The contributions to the work using CRM are as follows:

I. Color features are determined through the CIE Lab color space which approximates the human vision better than RGB color model. Then, the distribution of colors in an image is measured by color moments for characterization of color features. The main focus of this research is the ranklet transformation which is nonparametric statistical method for texture analysis based on nonlinear rank based filtering technique, provides alternative series of measurements that requires very limited assumptions to be made about the data points. This is invariant in nature towards transformations (brightness, contrast changes and gamma correction). Also, moment invariants are executed for shape analysis which, is weighted average (moment) of the image pixel's intensities of object in an image. Subsequently, these three parametric and nonparametric features are concatenated for making a better feature vector of an image which is able to represent all low-level characteristics present there.

II. After, enhancing the feature extraction process, different similarity measurements are evaluated to find out suitable similarity method and enlarge the number of relevant images for a search. Four distinct distance algorithms: Chi-squared test or X^2 statistics, Manhattan, Euclidean, and Canberra distance are investigated and found Euclidean distance outperforms among them with better precision and recall. The Euclidean measure is executed with the enhanced features to retrieve images from five different image databases by applying three different frame sizes (10, 12, 15).

III. The proposed CRM model is also compared with seven different existing CBIR techniques: color difference histogram (CDH) [33], edge histogram descriptor (EHD) [34], multi-texton histogram (MTH) [19], color auto correlogram (CAC) [10], distribution of color ton (DCTon) [27], golomb-rice (GR) coding based indexed histogram [13], and local binary pattern (LBP) based method with the combination of discrete wavelet transformation (DWT) and legendre moments (LM) which is denoted as (LDM) [30].

IV. Statistical measurements are performed to validate the work and show the performance of the proposed CRM are superior than other compared methods. In addition, a hypothesis test is performed and a thorough analysis is accomplished to exhibit the potentiality of this study.

This article is organized as follows: Section 2 gives the detailed description about the proposed CRM method which includes the feature extraction procedure (Section 2.1), similarity measurement (Section 2.2), performance evaluation (Section 2.3), and hypothesis test (Section 2.4). Afterthat, experiment and result analysis are discussed in Section 3 with parameters are fitted in the proposed model (Section 3.1), data details (Section 3.2), retrieval performance (Section 3.3), and result validation (Section 3.4). Finally, conclusion is presented in Section 4.

2. Proposed Method

Proposed work is divided into two phases: online and offline along with four sub-stages such as, feature extraction, similarity measure, performance evaluation, and hypothesis test. Figure 2 shows the flow of this work and connection between each parts. Initially, the color, texture, and shape features are extracted from all images of a database to create the database of feature vectors using CRM method and stored in offline phase. In the online phase, when user makes a query image to retrieve relevant images then, CRM algorithm is performed to derive features to form feature vector of the query image. Then, the feature vector of the query image is compared with the stored feature vectors by the help of similarity tolerance to fetch the relevant images for the user. Thereafter, the statistical metrics are determined over the retrieved images to check the ability of the proposed CRM method for searching relevant images. In addition, the hypothesis test is performed to assure the ability of the work.

2.1. Feature Extraction

The feature extraction technique of CRM are detailed in this section. Initially, this process is started with an image and applied for all stored images as well as the query for deriving feature values which are informative and non-redundant for better human interpretation.

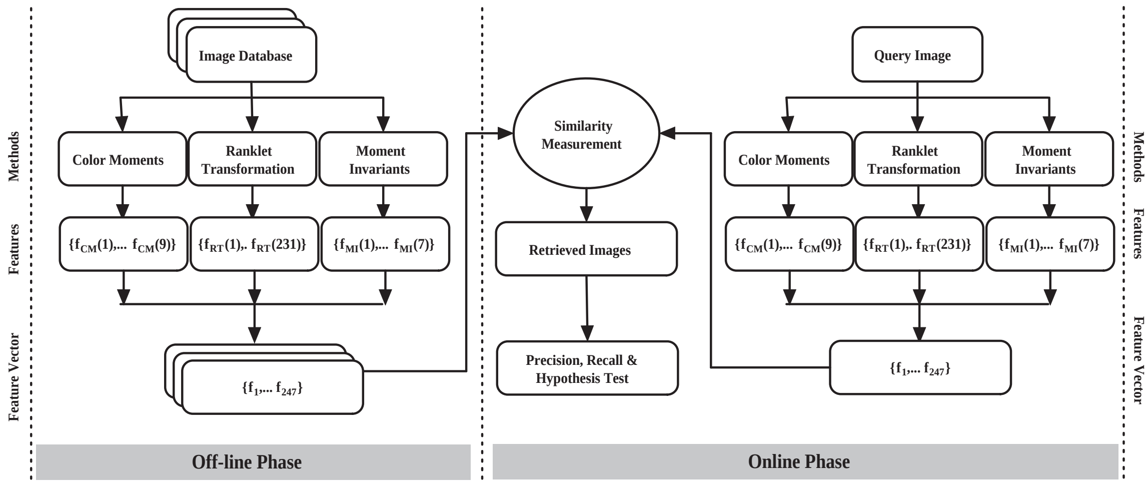

Figure 2: Flow diagram of proposed CBIR system.

2.1.1. Color Moments

Color is extremely used feature for image retrieval techniques [35]. Before selecting an appropriate color description, selection of color space is important and needs to choose a color model for color feature extraction process. Generally databases consider images in RGB format. However, this format contains highly redundant and correlated intensity values which degrade the efficiency of an algorithm. Thus, the proposed algorithm adapted CIE Lab color space for pulling out the color feature because of these pros and cons of RGB color model. It is a 3-axis color system with dimension L for lightness, a and, b for the color dimensions. The Lab color space includes all the colors of spectrum, as well as the colors of outside human perception. This is the most exact means of representing color and this is device independent color space. Scale and rotation invariant are most effective properties of this method where, first three color moments i.e., mean, variance, skewness are used as features in images retrieval. These are proved as effective and efficient features to represent color distribution of an image [36]. Color moments are defined below in (1) to (3),

$$\alpha_i = \frac{1}{N}\sum_{j=1}^{n} f_{ij} \tag{1}$$

$$\beta_i = (\frac{1}{N}\sum_{j=1}^{n}(f_{ij} - \alpha_i)^2)^{\frac{1}{2}} \tag{2}$$

$$\gamma_i = (\frac{1}{N}\sum_{j=1}^{n}(f_{ij} - \alpha_i)^3)^{\frac{1}{3}} \tag{3}$$

where, α_i, β_i, γ_i indicate mean, variance, and skewness, respectively; f_{ij} is the value of i-th color component of the image pixel j, and N is the number of pixel in the image. Color moments are calculated for each color channel (L, a, b) of an image. Therefore, 9 distinct color features are generated from color moments for a single image. The execution steps are listed in Algorithm 1.

Algorithm 1 Pseudo code for color moments

Require: all the color values are in $CIE\ L*a*b$ color space
  $CIE\ L*a*b\ image\ function = f(x,\ y)$
  for all $Color\ components = L,\ a,\ b$ **do**
    for all $x = 1\ to\ Number\ of\ rows\ in\ image$ **do**
      for all $y = 1\ to\ Number\ of\ columns\ in\ image$ **do**
        $Calculate\ \alpha,\ \beta,\ \gamma$ $\Leftarrow$ $using\ the\ Eq.\ 1,\ 2,\ 3$
      end for
    end for
  end for
  Return, $Feature\ vector$

2.1.2. Ranklet Transformation

Texture analysis of an image is the salient part of image description. A texture is a repeating appearance of particular pattern or intensities in an image which, draws the details about spatial alignment of these elements. There are two ways to analyze image texture i.e., structured and statistical approach. Structured approach considers the repetitive relationship of primitive texels whereas, texture is being considered as quantitative measure of the intensity arrangement in statistical approach. Here, proposed CRM method used ranklet transformation [37] to extract texture features. This is a non-parametric, multi-resolution and orientation selective algorithms that adopts the wavelet style. Ranklet coefficients are calculated for different resolution and orientation based on non-parametric statistics which deals with relative order of pixels instead of their intensity values. These rank based nonlinear filtering drops high spatial frequencies associated with noise, shifts the mean intensity in the direction of skewness, and preserves the shape of edges where no new intensity values are generated during this phase [38]. Here, texture feature extraction consists of two steps, one preprocessing step using ranklet transformation and another one is texture description using statistical descriptor.

(*a*) *Filtering with Ranklets*

In practice, ranklet transformation works with gray scale images, so images are converted into gray scale for this part. Let, each resolution contains N number of gray scale pixels. First, it is broken into two equal halves, one subset is T (treatment region) and the rest one is C (control region), these pairs of subsets are being defined differently depending on the orientation considered. Figure 3, describes T and C for three different orientations where, T_V and C_V are for vertical, T_H and C_H are for horizontal, and T_D and C_D are for diagonal orientation.

Then, non-parametric analysis is performed for each resolution and orientation. Initially, pixels of T and C regions satisfy the condition, that the gray scale value of pixels $p_m \in T$ is always higher than that of $p_n \in C$. Therefore, the ranklet coefficient is calculated based on the relative rank of pixels instead of their gray-scale intensity value,

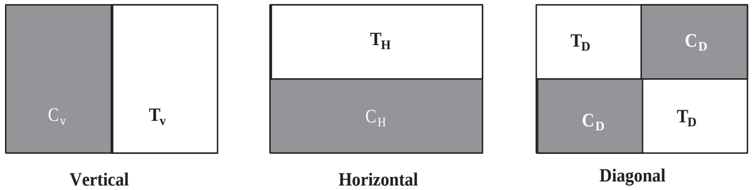

Figure 3: Different orientations of ranklet transformation.

$$R_j = \frac{\sum_{p \in T_j} \Pi(p) - \frac{N}{4}(\frac{N}{2} + 1)}{\frac{N^2}{8}} - 1 \qquad j = V, H, D \tag{4}$$

In (4), the summation of rank values $\Pi(p)$ in T_j is denoted as $\sum_{p \in T_j} \Pi(p)$. If more squares present in T_j that results higher intensity value than the pixels in C_j. Then the value of ranklet coefficient R_j is inclined to +1. Contrarily, it is inclined to -1 for more number of square crops in C_j and will have higher intensity pixel values than in T_j region. Also, R_V, R_H, and R_D are inclined to 0, if a square contains no vertical, horizontal, and diagonal value variation respectively. By employing this procedure, three ranklet images (RI), i.e., for vertical, horizontal, and diagonal orientations are decomposed from an image (I). Pixel values of these RI are represented by the ranklet coefficients R_V, R_H, and R_D which are determined from the square window of that specific resolution.

(b) Statistical Descriptor

In the proposed CRM method, each image is filtered with a set of multi-scale ranklets. There are seven different filters are chosen for window W, where each side contains 4, 6, 8, 10, 12, 14, and 16 pixels and ranklet orientations (such as, vertical, horizontal, diagonal) are computed for each of these resolutions. Subsequently, it takes the absolute value of filter responses and quantizes them into 21 equally spaced bins over the interval [-1, 1].

Equations (5) and (6) are implemented to calculate the ranklet histogram ($rhist$) and ranklet co-occurrence matrix ($rcom_{d,\theta}$) respectively, in order to calculate texture moments.

$$rhist(i) = \frac{n(i)}{\sum_{j=1}^{21} n(j)} \qquad i = 1, ..., 21 \tag{5}$$

where $n(i)$ is the number of the i-th quantized ranklet coefficients in the RI by taking values, $rvalue(i)$=(-1,-0.9,...,0,...,+0.9,+1).

$$rcom_{d,\theta} = \frac{n_{d,\theta}(i,j)}{\sum_{l=1}^{21} \sum_{k=1}^{21} n_{d,\theta}(l,k)} \qquad i, j = 1, ..., 21 \tag{6}$$

The $rcom_{d,\theta}$ converts the transitions probability between each pair of coefficients. Also, $n_{d,\theta}(i,j)$ is the frequency of ranklet values or co-occurrences which is quantized in bin i, j at a distance d pixels along θ direction.

The first two texture features are derived from the ranklet histogram $rhist$.

1. Mean Convergence (mc)

$$mc = \sum_{i=1}^{21} \frac{|rvalue(i).rhist(i) - \alpha|}{\beta} \tag{7}$$

where, α and β are mean and standard deviation of ranklet coefficient respectively.

2. Code variance (cv)

$$cv = \sum_{i=1}^{21} (rvalue(i) - \alpha)^2.rhist(i) \tag{8}$$

The remaining texture features are derived from ranklet co-occurence matrix ($rcom_{d,\theta}$), where d is fixed to 1 and $\theta \in (0^o, 45^o, 90^o, 135^o)$.

3. Code entropy (ce)

$$ce = \sum_{i=1}^{21}\sum_{j=1}^{21} rcom(i,j).log_{10}(rcom(i,j)) \tag{9}$$

4. Uniformity (un)

$$un = \sum_{i=1}^{21}\sum_{j=1}^{21} rcom(i,j)^2 \tag{10}$$

5. First order element difference moments (fdm)

$$fdm = \sum_{i=1}^{21}\sum_{j=1}^{21} |i-j| rcom(i,j) \tag{11}$$

6. Second order element difference moments (sdm)

$$sdm = \sum_{i=1}^{21}\sum_{j=1}^{21} (i-j)^2 rcom(i,j) \tag{12}$$

7. First order inverse element difference moments ($fidm$)

$$fidm = \sum_{i=1}^{21}\sum_{j=1}^{21} \frac{1}{1+|i-j|}.rcom(i,j) \tag{13}$$

8. Second order inverse element difference moments ($sidm$)

$$sidm = \sum_{i=1}^{21}\sum_{j=1}^{21} \frac{1}{1+(i-j)^2}.rcom(i,j) \tag{14}$$

9. First ranklet co-occurrence matrix for energy distribution ($edrcm_1$)

$$edrcm_1 = \sum_{i=9}^{13}\sum_{j=9}^{13} rcom(i,j) \tag{15}$$

10. Second ranklet co-occurrence matrix for energy distribution ($edrcm_2$)

$$edrcm_2 = \sum_{i=7}^{15}\sum_{j=7}^{15} rcom(i,j) - edrcm_1 \tag{16}$$

11. Third ranklet co-occurrence matrix for energy distribution ($edrcm_3$)

$$edrcm_3 = \sum_{i=3}^{19}\sum_{j=3}^{19} rcom(i,j) - edrcm_1 - edrcm_2 \tag{17}$$

So, there are 231 features (from (7) to (16)) produced to represent the texture of each RI. A top-down structuring approach of this feature extraction technique is included in Algorithm 2.

Algorithm 2 Pseudo code for ranklet transformation

Require: all the pixel values are in $Gray\ Scale$
  $Gray\ scale\ image\ function = f(x, y)$
  for all $resolution = 4\ to\ 16$ **do**
    for all $x = 1\ to\ (Number\ of\ rows\ in\ image - (resolution - 1))$ **do**
      for all $y = 1\ to\ (Number\ of\ columns\ in\ image - (resolution - 1))$ **do**
        $Put\ rank\ for\ the\ values\ of\ resolution$
        $Break\ the\ image\ into\ T\ and\ C\ region$
        for all $Orientation = T_{Horizontal},\ T_{Vertical},\ T_{Diagonal}$ **do**
          $Calculate\ Ranklet\ Coefficient\ R\ for\ each\ orientation\ (from,\ Eq.\ 4)$
        end for
        Return $Three\ Ranklet\ Images\ For\ Each\ Resolution$
        for all $Ranklet\ Image = RI_{Horizontal},\ RI_{Vertical},\ RI_{Diagonal}$ **do**
          $Calculate\ Ranklet\ Histogram\ (rhist)\ (from\ Eq.\ 5)$
          $Calculate\ Ranklet\ Co\ Occurrence\ Matrix\ (rcom)\ (from\ Eq.\ 6)$
          Then,
          $Calculate\ 11\ texture\ features\ (using,\ Eq.\ 7\ to\ Eq.\ 17)$
        end for
      end for
    end for
  end for
  Return, $Feature\ vector$

2.1.3. Moment Invariants

The earliest remarkable work on moments for image processing and pattern recognition was performed by Hu [39] and Alt [40], which is used as a shape feature extractor in the proposed CRM algorithm. It derives relative and absolute combinations of moments from binary images which are translation, rotation and scale invariant. If (x, y) is the co-ordinate of the pixel and $M_{p,q}$ is the 2D moment of the image function $f(x, y)$, then order of the moment is $(p + q)$ where, p and q are natural numbers. Image regular moments $M_{p,q}$ are defined as,

$$M_{p,q} = \int\int x^p y^q f(x, y) dx dy \tag{18}$$

Digital form of the above equation becomes,

$$M_{p,q} = \sum_x \sum_y x^p y^q f(x, y) \tag{19}$$

The image centroids are used to define the central moments for normalization and translation of the image plane. The centre of gravity of the image are calculated by (18).

$$x_c = \frac{M_{10}}{M_{00}} \qquad y_c = \frac{M_{01}}{M_{00}} \tag{20}$$

The central moments $(\omega_{p,q})$ of order $(p + q)$ for a shape of object R are defined as,

$$\omega_{p,q} = \sum_{(x,y)\in R} (x - x_c)^p (y - y_c)^q \tag{21}$$

where (x_c, y_c) is the center of that object. The ratio $\rho_{p,q}$ is determined to make the features scale invariant [41] as, $\rho_{p,q} = \frac{\omega_{p,q}}{\omega_{0,0}{}^{(p+q+2)/2}}$. A set of moments ($\xi_1$ to ξ_7) are calculated which are translation, rotation, and scale invariant and defined as follows,

$$\xi_1 = \omega_{2,0} + \omega_{0,2} \tag{22}$$

$$\xi_2 = (\omega_{2,0} - \omega_{0,2})^2 + 4\omega_{1,1}^2 \tag{23}$$

$$\xi_3 = (\omega_{3,0} - 3\omega_{1,2})^2 + (\omega_{0,3} - 3\omega_{2,1})^2 \tag{24}$$

$$\xi_4 = (\omega_{3,0} + \omega_{1,2})^2 + (\omega_{0,3} + \omega_{2,1})^2 \tag{25}$$

$$\begin{aligned} \xi_5 = {} & (\omega_{3,0} - 3\omega_{1,2})(\omega_{0,3} + \omega_{1,2})[(\omega_{0,3} + \omega_{1,2})^2 - 3(\omega_{0,3} + \omega_{2,1})^2] \\ & +(\omega_{0,3} - 3\omega_{2,1})(\omega_{0,3} + \omega_{2,1})[(\omega_{0,3} + \omega_{2,1})^2 - 3(\omega_{3,0} + \omega_{1,2})^2] \end{aligned} \tag{26}$$

$$\xi_6 = (\omega_{2,0} - \omega_{0,2})[(\omega_{3,0} + \omega_{1,2})^2 - (\omega_{0,3} + \omega_{2,1})^2] + 4\omega_{1,1}(\omega_{3,0} + \omega_{1,2})(\omega_{0,3} + \omega_{2,1}) \tag{27}$$

$$\xi_7 = (3\omega_{2,1} - \omega_{0,3})(\omega_{3,0} + \omega_{1,2})[(\omega_{3,0} - \omega_{1,2})^2 - 3(\omega_{0,3} + \omega_{2,1})^2] \tag{28}$$

Moment invariants have produced 7 invariant features with respect to translation, rotation and scale to describe the shape of an image. The structural conventions of this feature extraction mechanism are described in Algorithm 3.

Algorithm 3 Pseudo code for moment invariants

Require: Binary image format
  $Binary\ image\ function = f(x, y)$
  for all $x = 1\ to\ Number\ of\ rows$ **do**
    for all $y = 1\ to\ Number\ of\ columns$ **do**
      $Calculate\ the\ mass\ of\ the\ whole\ image, M_{00}\ (using\ Eq.\ 19)$
      $Calculate\ the\ mass\ of\ the\ whole\ image\ towards\ x\ axis, M_{10}\ (using\ Eq.\ 19)$
      $Calculate\ the\ mass\ of\ the\ whole\ image\ towards\ y\ axis, M_{01}\ (using\ Eq.\ 19)$
    end for
  end for
  Return,$M_{00},\ M_{10},\ M_{01}$
  $Calculate\ centre\ of\ gravity\ of\ the\ image,\ x_c,\ y_c\ (using\ Eq.\ 20)$
  $Calculate\ moment\ invariants, \xi_1\ to\ \xi_7\ (using,\ Eq.\ 22\ to\ Eq.\ 28)$
  Return, $Feature\ vector$

2.2. Distance Metric

Besides improved feature representation, good similarity measure (or distance metric) plays a crucial role for better retrieval performance. There are four similarity measures are investigated during the experiment, namely Chi-squared test or X^2 statistics [42], Manhattan distance or City block distance [43], Euclidean distance [44], and Canberra distance [45]. It is shown in Table 1 that, Euclidean distance metric yields better precision and recall than other distance measuring criteria. Therefore the Euclidean distance is chosen as similarity comparison method in the proposed CRM method. The similarity between two images with n dimensional feature vector is obtained using Euclidean distance. Let, Q be the feature vector of query image and S is the feature vector of a stored image. The Euclidean distance between Q and S is denoted as $D(Q, S)$,

$$D(Q, S) = \sqrt{\sum\nolimits_{i=0}^{n-1} (Q_i - S_i)^2} \tag{29}$$

where $Q = \{Q_0, Q_1, ..., Q_{n-1}\}$ and $S = \{S_0, S_1, ..., S_{n-1}\}$.

2.3. Performance Metrics

Precision and recall are well-known performance metrics in information retrieval. Precision or positive predictive value (PPV) specifies the retrieved outcomes which are relevant, whereas recall or sensitivity specifies relevant outcomes which are retrieved. Hence, high precision value means a system returns more relevant images than irrelevant ones. Conversely, high recall indicates the employed algorithm returns most of the relevant images. Here, precision (P_r) and recall (R_e) are defined as (30) and (31),

$$P_r = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ imgaes\ retrieved} \tag{30}$$

$$R_e = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ imgaes\ in\ the\ database} \tag{31}$$

2.4. Hypothesis Test

Precision and recall values are measured to show the potential performance of the proposed work. In this section the statistical hypothesis test, one sample t-test is discussed to determine the probability of the precision and recall sample mean truly being characteristic of the population or being a misinterpretation of the image population used in the experiment. The idea is to compare the average precision and recall obtained from the proposed method with the average precision and recall produced by existing methods over same image population. There are some assumptions to pursue the hypothetical test. Here, the experimental values are continuous interval variables and the sample of probability distribution of precision and recall values should be fitted in bell curve or normal distribution. As these quantitative values are independent, one sample t-test is appropriate for this analysis.

There are two kinds of hypothesis for this t-test i.e., null hypothesis and alternative hypothesis. The alternative hypothesis assumes the difference between an existing (precision or recall) mean and a proposed (precision or recall) mean is significant for same image population. On the other hand null hypothesis assumes the difference is insignificant. The goal is to measure any difference, regardless of direction, and a two-tailed hypothesis is used. If the direction of the difference between the evaluated sample mean and the comparison value matters, either an upper-tailed or lower-tailed hypothesis is used. The null hypothesis remains the same for each type of one sample t-test. The form of the hypothesis is formally defined below,

$$H_0 : \mu \leq \mu_0 \ vs. \ H_a : \mu > \mu_0 \tag{32}$$

Equation (32) shows the right-tailed test performed in this work to look at the potential improvements or increment in precision and recall value, where null hypothesis, alternative hypothesis, hypothesized precision, recall mean (existing precision, recall mean), evaluated precision, recall mean are denoted by H_0, H_a, μ_0, and μ respectively. The standard one sample t-test equation is stated in (33).

$$t = \frac{\mu - \mu_0}{S/\sqrt{n}} \tag{33}$$

Where, t-statistic value or t-value, precision and standard deviation of precision, recall mean and sample size are signified by t, S, and n respectively. Also, a desired significance level $\alpha = 0.05$ is assumed for accepting and rejecting the null hypothesis, where, t and p are compared to α for deciding the statistical significance.

3. Experiment and Result Analysis

In this study, three features (color,texture, and shape) are combined to propose CRM method. The generation of feature vector is explained in previous sections, where, 9, 231, and 7 number of features are extracted by color moments, ranklet transformation, and moment invariant respectively for representing an image (both the database and query images). The ultimate feature vector, F, is formed by sequentially concatenating (9+231+7=247) all these features as represented in (34).

$$F = \{f_{color}(1), ..., f_{color}(9), f_{texture}(1), ..., f_{texture}(231), f_{shape}(1), ..., f_{shape}(7)\} \tag{34}$$

Therefor, the performance of the proposed methodology is evaluated based on the results produced for each query image and explained in the following section.

3.1. Parameter Fitting

The proposed model is a nonlinear combination of particular parameters which are fitted by the method of successive approximations. These parameters are threshold of distance in similarity or dissimilarity measurement, and significance level in one sample t-test. Threshold for distance is assumed as 0.001 which means the distance between a query image and a stored image of 0.001 or less is considered as a relevant image for that query. In case of one sample t-test, the conventional 0.05 significance level is considered which, indicates the probability value (p-value) of less than or equal to the significance level 0.05 would reject the null hypothesis or accept otherwise. Therefore, above mentioned parameters should be considered for validating the CRM model.

3.2. Data Details

Five different image databases, Simplicity, Corel-5K, Corel-10K, Caltech-101, and MSR (Microsoft Research) are employed for the experiment. The Simplicity database contains 1000 images with two types of resolution, either 384×256 or 256×384, and 10 classes in JPEG format. Large image varieties are present in the Corel databases e.g., animals, outdoor sports, natural scenes etc. The Corel-5K database contains total 5000 images with 50 different image categories like, fireworks, trees, waves, glasses, etc., and Corel-10K contains 10000 images including 100 different categories like, doors, buildings, sunsets, etc. Each category contains 100 images of size 192×128 or 128×192 in JPEG format for both of the Corel subsets. Image database Caltech-101 includes the pictures of objects with 101 categories. Each category contains 40 to 800 images (mostly, 50). The size of each image is 300×200 pixels. Microsoft Research of Cambridge is created an image database 'MSR' for machine vision algorithms is also utilize here. It contains 4313 high resolution images with 23 categories. The resolution of each image is either 640×480 or 480×640.

3.3. Retrieval Performance

Different distance measurement are investigated first and analyzed the outcomes for choosing a suitable similarity measure for CRM. There are four similarity measures are applied in CRM to acquire relevant images and the performance metrics are determined and listed in Table 1.

Table 1: Precision and recall obtained using proposed CRM method with various distance measures.

Distance Metrics	Performance Metrics	Simplicity	Corel - 10K	Corel - 5K	Caltech - 101	MSR
Manhattan	Precision	0.3949	0.3817	0.3960	0.3766	0.3990
	Recall	0.0473	0.0458	0.0480	0.0451	0.0480
Chi Square	Precision	0.3941	0.2952	0.3820	0.2844	0.3241
	Recall	0.0473	0.0354	0.0460	0.0341	0.0389
Canberra	Precision	0.4268	0.3142	0.4140	0.3046	0.3858
	Recall	0.0512	0.0377	0.0500	0.0365	0.0463
Euclidean	Precision	0.6760	0.6744	0.6796	0.6450	0.6480
	Recall	0.0681	0.0577	0.0747	0.0645	0.0674

From Table 1 it is shown that the Chi-squared distance resulted the lowest precision compared to the other distance metrics investigated. This may be due to the fact that the samples used in the experiment are unpaired or independent data from large sample images [42], means the set of images are taken from separate individuals. The Manhattan distance function determines a grid like path distance between two data points. Hence, distance between two integers is the sum of the differences in their corresponding components. Investigation using Manhattan distance achieved improved precision values (minimum precision in Caltech dataset is 0.3766, and maximum precision in MSR dataset is 0.3990) than Chi-squared test.

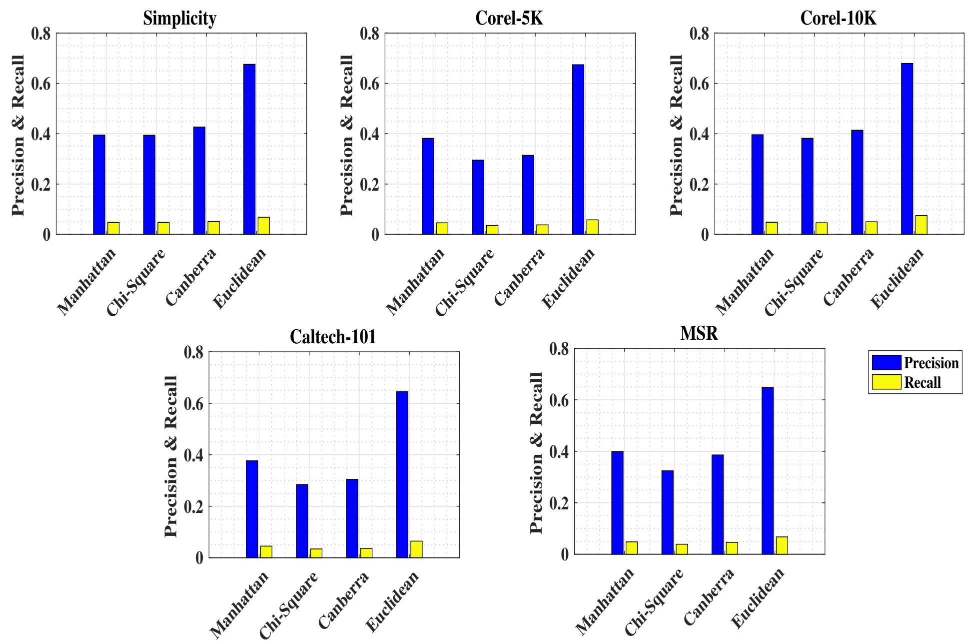

Figure 4: Performance of the proposed algorithm over different distance metrics.

The outcomes of Canberra distance is slightly better than Manhattan distance, yet the result is not satisfactory, because it could produce better result when data are scattered around origin. Here, the distinction of the absolute difference between the variables of the two images are divided by the sum of the absolute variable values prior to summing. Euclidean distance obtained best results among the compared distance metrics. It is obtained the minimum distance between two sample images on a plane. Euclidean distance measures feature wise difference of query and database images by squaring and adding the difference which, effectively increases the divergence between them. This method also produced highest recall compared to others, which indicates the ability of this to return more relevant results than others. The precision and recall of the proposed CRM method are plotted in Figure 4 for different distance metrics. This, clearly shows that, the Euclidean distance metric delivers the better result than other metrics. Therefore, the Euclidean distance is considered for the proposed CRM method.

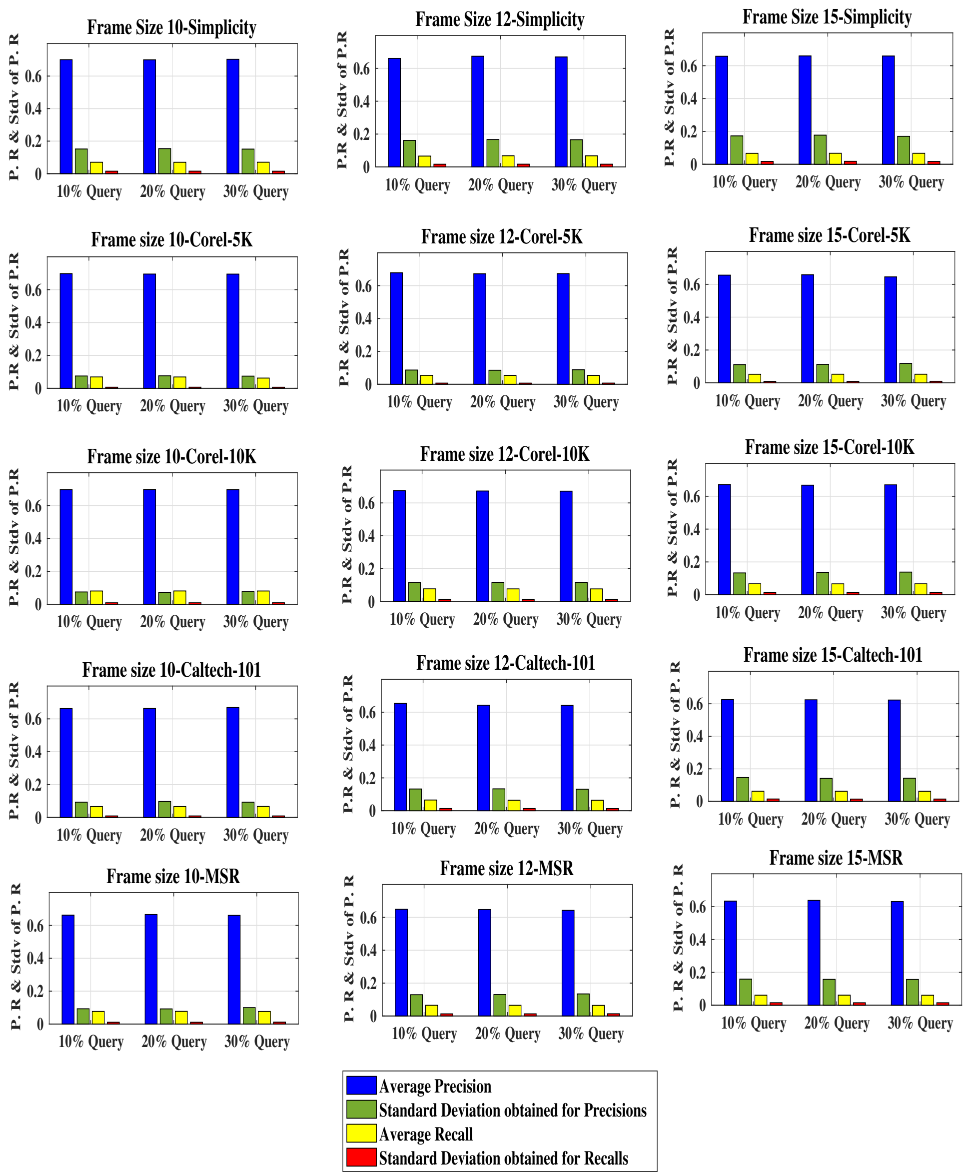

Figure 5: Average precision & recall obtained by different frame size.

Also, the frame size is varied by keeping distance measure (Euclidean) unchanged for investigation purpose. Three frame sizes 10, 12, and 15 are employed to investigate the variation of precision and recall when, 10%, 20%, and 30% images are selected randomly as query image from an image database. Figure 5 shows the average precision and recall are measured for different frame size. Along with the average precision and recall, standard deviation is also calculated for each frame to determine the precision and recall variation. It is found that, precision and recall are high at frame size 10, but decreased gradually for frame 12 and 15 because the misidentifications of one frame are carry forwarded by the other frames.

Table 2 encompasses the results obtained from CRM and seven recent CBIR algorithms. EHD expresses the spatial distribution of local edges for images with substantial texture which is non-homogeneous. The working procedure of this algorithm is simple. It creates histogram of edge directions for fixed size blocks. In practice, it also describes shapes depending on the edge field of object boundaries but, it is sensitive to object or scene distortions [34]. Huang et al. proposed color correlograms for image indexing and retrieval by a joint probabilistic style. Color correlogram is modified as color auto correlograms (CAC) to manage the high dimensionality of feature vector. Huang's used the concept of color quantization in RGB plane to present CAC [10] which generates a 256-dimensional feature vector and apprehends spatial correlation of color intensities. It provides maximum precision 0.4977 and recall 0.0597. The obtained precision and recall are mapped in Figure 6 for comparing the performance of implemented algorithms visually. The plot also includes the outcomes for different image database.

Table 2: Comparison of proposed CRM method with existing algorithms.

Methods	Measurements	Simplicity	Corel-5K	Corel-10K	Caltech-101	MSR
EHD	Precision	0.3721	0.4087	0.3431	0.3731	0.3817
	Recall	0.0446	0.0490	0.0411	0.0447	0.0458
CAC	Precision	0.4753	0.4977	0.3950	0.3862	0.3951
	Recall	0.0570	0.0597	0.0474	0.0463	0.0474
MTH	Precision	0.4852	0.5011	0.4157	0.4064	0.4157
	Recall	0.0582	0.0601	0.0498	0.0487	0.0498
CDH	Precision	0.5454	0.5874	0.4735	0.4629	0.4735
	Recall	0.0654	0.0704	0.0568	0.0555	0.0568
DCTon	Precision	0.5514	0.4878	0.4953	0.4350	0.4800
	Recall	0.0686	0.0616	0.0689	0.0641	0.0635
GR	Precision	0.6631	0.6319	0.6202	0.6500	0.6489
	Recall	0.0630	0.0490	0.0422	0.0580	0.0410
LDM	Precision	0.6523	0.6176	0.6437	0.6498	0.5608
	Recall	0.0618	0.0567	0.0585	0.0625	0.0540
CRM	Precision	0.6760	0.6744	0.6796	0.6450	0.6480
	Recall	0.0681	0.0577	0.0747	0.0645	0.0674

MTH illustrates spatial correlation of color intensities and edge orientation by employing texton analysis [19]. MTH considers four types of texton which may not depict the complete content of texton elements, because it fails to consider the perceptual uniform color difference. It produced 0.4852 precision and 0.0582 recall in Simplicity, 0.5011 precision and 0.0601 recall in Corel-5K, 0.4157 precision and 0.0498 recall in Corel-10K, etc. CDH [33] measures the color difference between two pixels by using Lab color space, and generates the color edge histogram of an image. However, color differences cannot be measured in RGB color space that is close to human color perception, and it generates a high dimensional feature vector. In the case of CDH, increment of equalization number of color and edge orientation may not always enhance the description power. Therefore, results may not always be satisfactory. DCTon [27], is a hybrid CBIR method which is used to compare the performance of proposed framework. DCTon extracts the inter class features using distribution of color ton with association of dual tree complex wavelet transformation and singular value decomposition (SVD). It is a strong feature descriptor for images which, represents color and texture information. This method produces high recall value, which signifies that, it retrieves more relevant image instances. The highest precision and recall obtained by this method are 0.5514 (in Simplicity database) and 0.0689 (in Corel-10K database) respectively. This method is limited by high

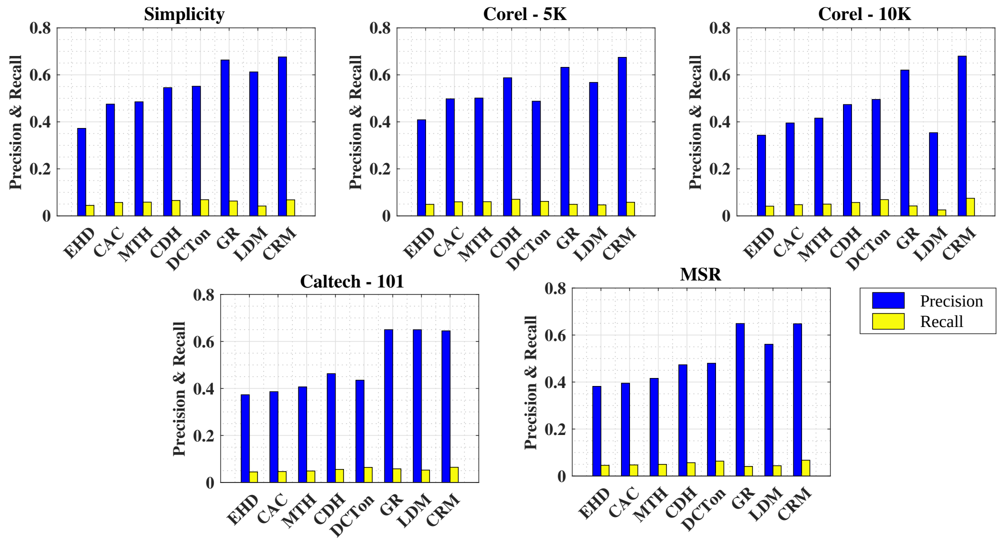

Figure 6: Performance comparison of proposed method with other existing algorithms.

computational complexity and scattering property of light. In [13], an indexed and encoded representation of color information is used to retrieve images from large databases. In first stage, the insignificant colors are discarded and important color information are kept to make an indexed color histogram. Then, the GR encoding is used to represent the indexed color histogram. This indexed histogram with GR encoding based CBIR method achieves good precision 0.6631 in Simplicity database. In case of Corel-5K and Corel-10K, it retrieved low amount of relevant images which results low precisions of 0.6319 and 0.6202 respectively. Also, it delivers low recall (0.0422 and 0.0410) in Corel-10K and MSR databases, which indicates the number of relevant images over total images is low. This GR encoding based method only considers the color information of images but in practice the images contain highly complex color, shape, texture information, and semantic content. Also, the histogram ignores shape and texture which, causes the problem to distinguish different objects having same color, e.g. black dog and black elephant. The indexing method of the histogram uses limited number of simultaneous colors per image which, is unable to describe the complex color as well as other semantic information. Though, the second phase uses a lossless GR encoding, still it again employs predicted pixel value for optimizing the representation which may obscure actual meaning of an image. Srivastava and Khare [30] proposed an algorithm LDM by combining three different feature extraction techniques. Initially, the images are converted to gray scale for different levels of decomposition and LBP extracts texture information from these decomposed elements. Finally, an orthogonal transformation LM is used to represent the images. It achieved maximum precision 0.6523 in Simplicity image database and recall 0.0625 in Caltech-101 dataset. The multi level decomposition of images causes intensive computational complexity, but the method shows a good set of average precision and recall. In some cases it cannot discriminate among the images. The MSR dataset contains images with complex color. However, this method uses gray scale which is a bottleneck for describing the color distribution. It may be a reason for not to produce good result in MSR. The result shows that, the method is retrieved few number of relevant images over the frames as well as from each database which, causes a low precision and recall fraction. Experiment also indicates that, this method unable to discriminate among different views of an object.

Figure 7: Picture in left top most is the query and remaining images are the retrieved ones.

The proposed CRM is a hybrid mechanism of three features which, increased the discrimination power between images. Although the total feature vector length is 247 which, can be considered a little high however, it reflects high discrimination power among the images. The proposed method achieved highest precision 0.6796 and recall 0.0747 in the Corel-10K database. The lowest precision and recall outcomes are 0.6450 and 0.0645 respectively in the Caltech-101 database. Two examples of the retrieval result are shown in Figure 7 and Figure 8 from MSR image database. The query image of Figure 7 is a door image which is nearly dark brown in color and rectangular in shape with texture. CRM retrieved 10 correct images

Figure 8: Picture in left top most is the query and remaining images are the retrieved ones.

out of 15. If the frame size of 10 is being considered then CRM retrieved all the relevant images but, in case of frame size 12 and 15, few irrelevant retrieval images are found. Thus, the precision 0.6666 is determined for the particular instance of frame size 15. Though, it is important that the 10 images are retrieved correctly irrespective of different color and texture whereas, 5 incorrectly retrieved images are somehow homologous with respect to color, texture, and shape. In case of Figure 8, the query image is an yellow flower from a green field where, CRM return 12 relevant images among 15. Somehow, all 15 retrieved images are similar in nature because, all are flower images e.g., image 13th in the frame is visually similar but white in color. This results precision of 0.8 which is quite high. Overall, it can be stated that proposed multi-feature fusion is effective towards improvement of CBIR performance.

3.4. Result Validation

The precision and recall measurements illustrate the accuracy of the proposed method based on relevant and irrelevant images are being returned for a query. The precision and recall data are systematically examined with the purpose of highlighting useful information and the results of this quantitative experimental work communicated via tables and graphs. The hypothesis test is conducted to determine the impact and quality of the work presented here. The CRM method used continuous interval variables throughout the experiment, therefore, a quantitative inferential analysis is possible based on the probability distribution of the precision and recall samples. A sample of probability distribution of precision and recall values for 30% query image from MSR and Caltech-101 image database are included in Figure 9, and the data distribution curves are of bell shape or normal distribution.

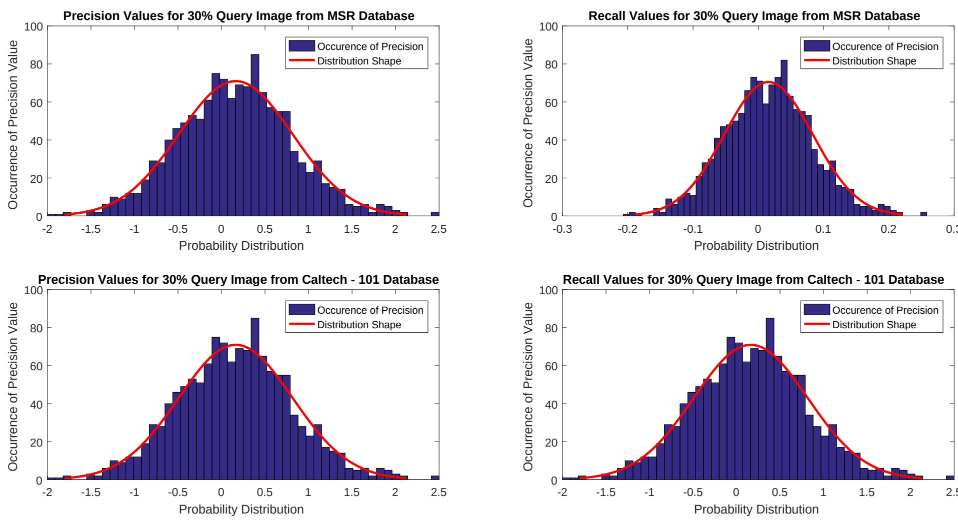

Figure 9: Distribution of precision, recall scores in MSR and Caltech-101 image database.

Therefore, a parametric statistical test is performed based on the shape of the curve. According to theoretical assumptions, normally distributed data could be tested via t-test or t-statistic. In the proposed work, images are independent in nature, so the one sample t-test is performed on determined precision and recall samples. The test is used to prove that, the means of precisions and recalls are determined from the same image databases. The t-values of precision and recall are calculated from five image databases using (33). Table 3 is included to present the p values obtained from different image database. The t-test is executed to established the obtained precision and recalls. These dataset are already being used by existing algorithms, mentioned earlier. Equations (32) and (33) are used to calculate the p and t values where, precision and recall means of existing techniques (listed in Table 2) are considered as hypothesized mean.

The precision and recall mean are obtained from CRM are considered as sample mean. Sample and hypothetical means are compared to determined t and p values. The p value less than significance level (α) and t value is conventionally considered as significant which, also decides the acceptance and rejection of null hypothesis (described in Section 2.4). It is found

Table 3: p values from precision and recalls over different database and existing algorithms.

Methods	t-test Parameters	Simplicity	Corel-5K	Corel-10K	Caltech-101	MSR
EHD & CRM	$P_{precision}$	0.00025	0.00045	0.00037	0.00014	0.00041
	P_{recall}	0.00017	0.00091	0.00050	0.00088	0.00086
CAC & CRM	$P_{precision}$	0.00047	0.00029	0.00059	0.00059	0.00017
	P_{recall}	0.00018	0.00036	0.00011	0.00065	0.00038
MTH & CRM	$P_{precision}$	0.00083	0.00017	0.00041	0.00093	0.00044
	P_{recall}	0.00023	0.00088	0.00090	0.00019	0.00021
CDH & CRM	$P_{precision}$	0.00072	0.00068	0.00032	0.00018	0.00031
	P_{recall}	0.00067	0.00024	0.00027	0.00060	0.00086
DCTon & CRM	$P_{precision}$	0.00092	0.00038	0.00011	0.00039	0.00035
	P_{recall}	0.00553	0.00045	0.02170	0.00794	0.00066
GR & CRM	$P_{precision}$	0.00417	0.00059	0.00075	0.00677	0.00836
	p_{recall}	0.00062	0.00085	0.00016	0.00022	0.00042
LDM & CRM	$p_{precision}$	0.02370	0.00028	0.00056	0.00752	0.00074
	p_{recall}	0.00012	0.00060	0.00017	0.00058	0.00067

that the resultant p values are statistically meaningful and validate the proposed findings. Subsequently, the p values are plotted against respective t values as shown in Figure 10. In Figure 10, it is clearly shown that, the t-values are large and p-values are comparatively very small which, is statistically very significant. Additionally, if p-values failed to meet the significant threshold level α (=0.05) results the rejection of null hypothesis mentioned in (33). The pictorial representation of null hypothesis acceptance and rejection are added in Figure 11. This figure also indicates decreased support for the null hypothesis because, the p-vales never meets the cut-off or threshold value at which statistical significance of 0.05 is claimed, also it ensures approximately 95% of confidence in the results.

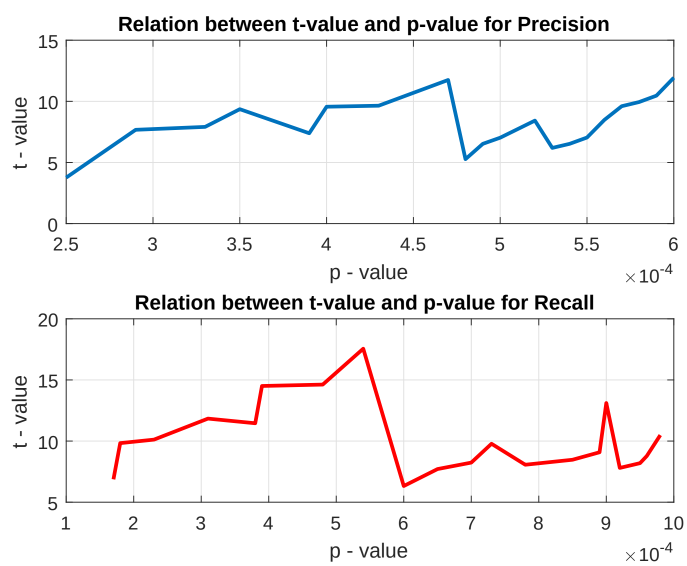

Figure 10: Relation between t-value and p-value.

3.5. Discussion

CBIR is a classical and difficult research domain because of the increasing image size, collection, and distinct image nature. Also, the technical aspects here to represent human

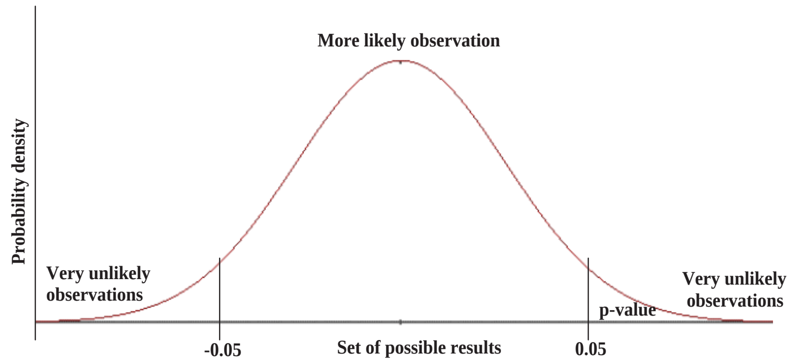

Figure 11: Null hypothesis acceptance and rejection.

perception makes this domain complex. Despite of the rich state-of-art, the real-time image retrieval is quite challenging due to the diverse image property. For example, Figure 7 includes the 15 retrieved images for single query image of door. The different doors have different properties due to the image contrast or color or the time when the picture was taken or the camera resolution or many more. Visually, it can be seen that, shape and texture are dominant feature here because, user needs the doors of any color and doors are always same in size. In Figure 8, user needs a yellow flower which has its own shape. Hence, the color and shape are dominant in this case. So, the proposed study tried to provide an efficient and hybrid CBIR approach that, doesn't lose any significant property of an image and retrieve relevant images as much as possible. Thus, the selection of different feature extraction algorithms is vital step towards the CBIR process. Color moment, ranklet transformation, and moment invariant are selected to make an efficient CBIR tool by using their individual capabilities. Also, the other aspects of CBIR such as, similarity measure and performance analysis over different frame size (10, 12, and 15) are explored to tune the performance properly where, Euclidean measure delivers optimal performance for all the relevant image search. The proposed method is executed over five different image dataset and compared with recent CBIR techniques. This comparison assure the performance of this method is prevailing among other method across all the five database images. Eventually, the one sample t-test is executed to validate the work theoretically and practically.

4. Conclusion & Future work

The proposed CRM method combines three invariant features which is implemented over five different image databases and achieved significant precision values with respect to other existing techniques. Experimental result shows that, this is a good representation of images, and generates very high discrimination power among images. CRM uses the feature vector of length 247, which is comparatively high with respect to other existing techniques may cause of longer time complexity. Thus, reduction of feature vector length would be a future consideration as time is also a pivotal property of CBIR systems. Another issue that would be taken into account is the purification of obtained results by using relevance feedback (RF). The results that are initially returned for a given query would be visually filtered through human intervention and put the feedback to system for better results in the next iteration.

References

[1] C. H. Chuang, S. C. Cheng, C. C. Chang, Y. P. P. Chen, Model-based approach to spatial–temporal sampling of video clips for video object detection by classification, Journal of Visual Communication and Image Representation 25 (2014) 1018–1030.

[2] H. T. Wu, J. Huang, Y. Q. Shi, A reversible data hiding method with contrast enhancement for medical images, Journal of Visual Communication and Image Representation 31 (2015) 146–153.

[3] M. Gouiffès, B. Planes, C. Jacquemin, HTRI: High time range imaging, Journal of Visual Communication and Image Representation 24 (2013) 361–372.

[4] W. Xing Yuan, L. Fan Ping, W. Shu Guo, Fractal image compression based on spatial correlation and hybrid genetic algorithm, Journal of Visual Communication and Image Representation 20 (2009) 505–510.

[5] I. K. Sethi, I. L. Coman, D. Stan, Mining association rules between low-level image features and high-level concepts, in: Data Mining and Knowledge Discovery: Theory, Tools, and Technology III, International Society for Optics and Photonics, Orlando, FL, United States (2001), pp. 279–291.

[6] C. H. Lin, R. T. Chen, Y. K. Chan, A smart content-based image retrieval system based on color and texture feature, Image and Vision Computing 27 (2009) 658–665.

[7] Y. Liu, D. Zhang, G. Lu, W.-Y. Ma, A survey of content-based image retrieval with high-level semantics, Pattern recognition 40 (2007) 262–282.

[8] R. Weber, M. Mlivoncic, Efficient region-based image retrieval, in: Proceedings of the twelfth international conference on Information and knowledge management, New Orleans, LA, USA (2003), pp. 69–76.

[9] S. H. Schwartz, Visual Perception: A Clinical Orientation, 3rd edition, 2004.

[10] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, R. Zabih, Image indexing using color correlograms, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, United States (1997), pp. 762–768.

[11] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, A. Yamada, Color and texture descriptors, IEEE Transactions on Circuits and Systems for Video Technology 11 (2001) 703–715.

[12] B. S. Manjunath, P. Salembier, T. Sikora, Introduction to MPEG-7: Multimedia Content Description Interface, John Wiley & Sons, Inc., New York, NY, USA, 2002.

[13] S. Shaila, A. Vadivel, Indexing and encoding based image feature representation with bin overlapped similarity measure for cbir applications, Journal of Visual Communication and Image Representation 36 (2016) 40–55.

[14] R. M. Haralick, K. Shanmugam, Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics SMC-3 (1973) 610–621.

[15] H. Tamura, S. Mori, T. Yamawaki, Textural features corresponding to visual perception, IEEE Transactions on Systems, Man and Cybernetics 8 (1978) 460–473.

[16] G. R. Cross, A. K. Jain, Markov random field texture models, IEEE Transactions Pattern Analysis and Machine Intelligence 5 (1983) 25–39.

[17] B. S. Manjunath, W. Y. Ma, Texture features for browsing and retrieval of image data, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (1996) 837–842.

[18] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 971–987.

[19] G. H. Liu, L. Zhang, Y. K. Hou, Z. Y. Li, J. Y. Yang, Image retrieval based on multi-texton histogram, Pattern Recognition 43 (2010) 2380–2389.

[20] G. H. Liu, J. Y. Yang, Image retrieval based on the texton co-occurrence matrix, Pattern Recognition 41 (2008) 3521–3527.

[21] G. H. Liu, Z. Y. Li, L. Zhang, Y. Xu, Image retrieval based on micro-structure descriptor, Pattern Recognition 44 (2011) 2123 – 2133.

[22] J. Luo, D. J. Crandall, Color object detection using spatial-color joint probability functions., IEEE Transactions on Image Processing 15 (2006) 1443–1453.

[23] W. Burger, M. J. Burge, Principles of Digital Image Processing: Core Algorithms, Springer Publishing Company, Incorporated, 1st edition, 2009.

[24] R. C. Gonzalez, R. E. Woods, Digital Image Processing, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 3rd edition, 2006.

[25] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2004) 91–110.

[26] J. C. V. Gemert, C. J. Veenman, A. W. M. Smeulders, J. M. Geusebroek, Visual word ambiguity, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2010) 1271–1283.

[27] M. Rahimi, M. E. Moghaddam, A content-based image retrieval system based on color ton distribution descriptors, Signal, Image and Video Processing 9 (2015) 691–704.

[28] L. Feng, J. Wu, S. Liu, H. Zhang, Global correlation descriptor: a novel image representation for image retrieval, Journal of Visual Communication and Image Representation 33 (2015) 104–114.

[29] M. Zhao, H. Zhang, J. Sun, A novel image retrieval method based on multi-trend structure descriptor, Journal of Visual Communication and Image Representation 38 (2016) 73–81.

[30] P. Srivastava, A. Khare, Integration of wavelet transform, local binary patterns and moments for content-based image retrieval, Journal of Visual Communication and Image Representation 42 (2017) 78–103.

[31] C. Cui, P. Lin, X. Nie, Y. Yin, Q. Zhu, Hybrid textual-visual relevance learning for content-based image retrieval, Journal of Visual Communication and Image Representation (2017).

[32] V. Kim, L. Iaroslavskii, Rank algorithms for picture processing, Computer Vision graphics and image processing 35 (1986) 234–258.

[33] G. H. Liu, J. Y. Yang, Content-based image retrieval using color difference histogram, Pattern Recognition 46 (2013) 188–198.

[34] C. S. Won, D. K. Park, S.-J. Park, Efficient use of MPEG-7 edge histogram descriptor, ETRI journal 24 (2002) 23–30.

[35] J. D. Foley, A. V. Dam, S. K. Feiner, J. F. Hughes, Computer Graphics: Principles and Practice (2nd Ed.), Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1990.

[36] M. A. Stricker, M. Orengo, Similarity of color images, in: Storage and Retrieval for Image and Video Databases III, International Society for Optics and Photonics, San Jose, CA, United States (1995), pp. 381–393.

[37] Texture classification using invariant ranklet features, Pattern Recognition Letters 29 (2008) 1980–1986.

[38] R. Hodgson, D. Bailey, M. Naylor, A. Ng, S. McNeill, Properties, implementations and applications of rank filters, Image and Vision Computing 3 (1985) 3–14.

[39] M. K. Hu, Visual problem recognition by moment invariants, IRE Transactions on Information Theory IT-8 (1962) 179–187.

[40] F. L. Alt, Digital pattern recognition by moments, Journal of the ACM (JACM) 9 (1962) 240–258.

[41] A. K. Jain, F. Farrokhnia, Unsupervised texture segmentation using gabor filters, Pattern recognition 24 (1991) 1167–1186.

[42] N. K. G. Gosall, G. Singh, Doctor's Guide to Critical Appraisal (3 Ed.), Knutsford: PasTest, 2012.

[43] D. L. Donoho, For most large underdetermined systems of linear equations the minimal $l1$-norm solution is also the sparsest solution, Communications on Pure and Applied Mathematics 59 (2006) 797–829.

[44] W. Ma, B. S. Manjunath, Netra: A toolbox for navigating large image databases, Multimedia systems 7 (1999) 184–198.

[45] G. N. Lance, W. T. Williams, Computer programs for hierarchical polythetic classification ("similarity analysis"), The Computer Journal 9 (1966) 60–64.