Chapter 2

Proposed Features Extraction Methods

The purpose of computing is insight, not pictures.
Lloyd N. Trefethen

2.1 Overall Structure of the Proposed Retrieval Method

The proposed content based image retrieval method based on a statistical framework consisting of three techniques (i) Haar wavelet based decomposition method (ii) multiresolution using statistical method and (iii) wavelet based orthogonal polynomial.

An overview of the proposed framework of the image retrieval method using combination of multiresolution color and texture features is depicted in Fig. 2.1. When a query image is given as input to the retrieval system, it is converted to an HSV color space whose components are $q_H, q_S, q_V$. Then the HSV component images are decomposed using wavelet transforms such Haar, Daubechies-4(Daub-4) and Wavelet packet transforms(WPT), into $W^{H}_{m,n}, W^{S}_{m,n}, W^{V}_{m,n}$, where $m \in \{LL, HL, LH, HH\}$ denotes subband orientation and $n$ denotes the wavelet decomposition level which are chosen as $n \in \{1,2,3,..\}$ to construct the proper feature vector.

In wavelet based image feature technique, the input image is decomposed using Haar wavelet transform. Features such as a spectrum of energy distribution and spatial relationship between the color pixels (covariance matrix) are extracted at optimum level of pyramid structure. Based on these features, feature vector is constructed. The system
determines the similarity between the feature vector of the query and the target images in an image database by Bhattacharyya similarity distance value. Then the system retrieves the $k$-relevant images according to the user needs. The details of the feature extraction technique, and similarity method are discussed in section 2.3.

In multiresolution based statistical feature technique, Daubechies-4 wavelet transform (DWT) adopted to decompose the images HSV color space into $W^H_{m,n}$, $W^S_{m,n}$ and $W^V_{m,n}$ components. The color feature $f_c$, the color autocorrelogram is extracted from $W^H_{m,n}$ and $W^S_{m,n}$ component images and texture features $f_t$, a set of statistical properties is extracted from $W^V_{m,n}$ component. Based on these features, the feature vector $f_q$ is constructed. Then the system computes the similarity between the feature vector of the query and target images by Minkowski-form distance method and retrieve the most relevant $k$-number of images from the image database. The details of the feature extraction technique, and similarity method are discussed in section 2.4.

In wavelet based orthogonal polynomial, the images in HSV color space input image is decomposed into $W^H_{m,n}$, $W^S_{m,n}$ and $W^V_{m,n}$ component images. Wavelet packet transform (WPT) is used to construct a 2-level multiresolution structure. The color feature $f_c$, the color autocorreloagram feature are extracted from $W^H_{m,n}$ and $W^S_{m,n}$ components, and texture features $f_t$, a set of statistical properties based on Gray-level co-occurrence matrix.
(GLCM), are extracted from $W_{m,n}$ component image. The system combines both color feature $f_c$ and texture feature $f_t$ to form the feature vector $f_q$. Then the system determines the similarity between the feature vectors of the query and each target images, and retrieves the $k$-number of relevant images from the image DB using the Manhattan distance method, which is illustrated in section 2.5
Fig. 2.1 Block diagram of the proposed CBIR frameworks
2.2 Image Decomposition Method

In real-time applications, though most of the image retrieval systems consider a number of features to obtain high retrieval rate, it requires more space and time complexity [RIT08]. This motivates us to develop a simple but efficient, novel technique for image retrieval, which is content-based wavelet multiresolution system. The Wavelet transform is used to construct an optimum level pyramid image structure [JAM07]. The optimum level means neither losing the dominant features in the image and nor the enormous number of feature vectors. To extract highly dominant features, we have computed spectrum of energy and spatial relationship between color pixels (covariance matrix) at the optimum level subbands of the pyramid structure.

Wavelet transform represents a function as a superposition of a family of basis functions called wavelets [MAL89]. A set of basis functions can be generated by translating and dilating the mother wavelet corresponding to a particular basis. The signal is passed through a low-pass filter (LPF) and a high-pass filter (HPF), and the outputs of the filter are decimated by two. Thus, wavelet transforms extracts information from the signal at different scales. In recent years, wavelet transforms has emerged as an effective tool for analyzing color and texture features as it decomposes the image into various subbands [DAU92, DAU90, TAO02] which is multi-scale oriented. For computational convenient, the given input color image with size $N \times N$ is converted to a sequence $X(n)$ where $n = N \times N$. The Discrete Haar wavelet transform (DHWT) is adopted to decompose the input image into one
approximation subband and three detail subbands, which result in the low-pass and high-pass filters respectively, which is shown in Fig. 2.2.1. Each subbands can be thought of as a smaller version of the image representing various image characters. The sub-images of the wavelet coefficients are called subbands and image decomposition on $N_j$ levels consist of $3 \times N_j + 1$. The concept of the three level decomposition structures is presented in Fig. 2.2.1. and Fig. 2.2.2.

**Fig. 2.2.1** Three-level pyramid decomposition structure

**Fig. 2.2.2** Decomposition schemes (a). Wavelet transforms of an image – pyramid decomposition scheme (b) Pyramid decomposition.
In Fig. 2.2.2 (b) the three sub-images of wavelet coefficients are obtained at each level \( l \): \( \text{LLH} \), \( \text{HL} \) and \( \text{HH} \) and an approximation subimage \( l \): \( \text{LLL} \), where \( l \) represents the number of levels.

- **LL** – subband is representing the approximation of the original image.
- **LH** – subband is representing the changes of the image along horizontal directions.
- **HL** – subband is representing the changes of the image along vertical directions.
- **HH** – subband show the high frequency components of the image.

### 2.2.1 Pyramidal Structure

The Laplacian pyramidal algorithm has been incorporated to construct the pyramid image structure. Laplacian pyramidal algorithm introduced by Burt and Adelson [PET83] to the difference between two successive levels in a pyramid, defined itself in turn by repeatedly applying a low-pass (smoothing) filtering operation. After the filtering, only one sample out of two is kept. The number of pixels decreases by a factor 2 at each scale. The difference between images is obtained by expanding (or interpolating) one of the pair of images in the sequence associated with the pyramid. The convolution is carried out with the filter \( h \) by keeping one sample out of two in the two-dimensional case and the structure of pyramid is presented in Fig. 2.2.3

\[
c_{j+1}(m,n) = \sum_{k,l} h(k-2m,l-2n)c_j(k,l)
\]  
\[(2.2.1)\]
2.2.2 Construction of Optimum Level

In this section, the optimum level is defined from pyramid structure image [ADE84] which is discussed in next section 2.3. The optimum decomposition with L levels is allowed under the condition $N = 2^L$ where $N$ is the number of coefficients of the decomposed image. In the pyramid
structure, the apex level image is coined as optimum level which contains fewer dominant wavelet coefficients. This can be done very quickly due to the reduced image sizes. The best search at the optimum level can be refined using a fine-to-coarse approach. The number of levels of the pyramid image is determined by Eq. (2.2.2)

$$K_{\text{max}} \leq \log_2 \min(I_r, I_c)$$  \hspace{1cm} (2.2.2)

Where \((I_r, I_c)\) represent the size of the input and \(K\) represents the number of levels. The advantages of building a pyramid become quickly obvious: the \(K\)-th level of a pyramid has \(2^{2(K-1)}\) times fewer coefficients than the original image at level 0. The most of the literatures [MAR01] [SAI11] decompose the images upto the 5\(^{th}\) level to consider image features. In addition, the proposed method also obtains a better result at 5\(^{th}\) level, which is discussed in section 2.2.1. Since the 5\(^{th}\) level pyramid image is reduced as much as to 256 coefficients in the case of 256×256 size image. Thus this study considers the features at the 5\(^{th}\) level pyramid image.

### 2.3 Haar Wavelet Based Image Feature

A Haar wavelet is the simplest form in terms of computation. The Haar transform decomposes an image into two sub-images of half of its size. One subimage is a running average or trend; the other subimage is a running difference or fluctuation. The matrix representation of the Haar wavelet transform is shown in Fig. 2.3.1.
The input image signals $s_0, s_1...s_{N-1}$ contains $N$ coefficients; there are $N/2$ approximations and $N/2$ wavelet coefficients. The approximations are stored in half of the upper array $[a_0...a_3]$ and wavelet coefficients are stored in another half of the lower array $[c_0...c_3]$. This process is continued until the optimum level is obtained.

The procedure of decomposition process is presented in the algorithmic form as follows:

```
Procedure DecomposePixel( S[0...n])
Begin
    While n>1 do
        Begin
            n=n/2
            For i=0 to n-1 do
                a[i] = (S[2i] + S[2i+1])/2;
                c[n+i] = (S[2i] - S[2i+1])/2;
            End for
            S=a;
        End while
    End procedure
```

In the above procedure, the entries of $S$ represents 3-dimensional color components of RGB, each components are in the range of 0...255. The entire $N \times N$ image is decomposed till it obtains the optimum level. The following algorithm describes the decomposition procedure.
Procedure PyramidImage(I(0…n,0…n) of color)
For r=0 to n do
    DecomposePixel(I[r…n])
End for
For c=0 to n do
    DecomposePixel(I[n…c])
End for
End procedure

2.3.1 Subband Characterization with Gaussian Distribution Function

The generalized Gaussian distribution function [WOU99] is employed to characterize the statistical behavior of transformed wavelet coefficients of each subbands. The generalized Gaussian function may be expressed as in Eq. (2.3.1)

\[ f_{\alpha,\beta}(x) = \frac{\beta}{2\alpha \Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|x|}{\alpha}\right)^\beta} \]  

where \( \alpha \) is the scale factor, related to the standard deviation of Gaussian function, \( \beta \) is a shape parameter, its value is 2 for Gaussian distribution and \( \Gamma(.) \) is the gamma function defined for \( z>0 \) by the following equation

\[ \Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt \]  

The proposed system incorporates only the LL subbands coefficients from the bottom level to apex level pyramid structure (level 0 to level 4). For a sample, we have used Lena image which is shown in Fig. 2.2.4, with the generalized Gaussian distribution function, and obtained result is shown in Fig. 2.3.2. Here it is observed that the decomposition of a pyramid image from the coarse level to apex level (level 0 to level 4) follows the generalized
Gaussian distribution function. Therefore, this study considers extracting the features at higher level instead of coarse level since the coefficients of the apex level subbands compactly follow the Gaussian distribution and it contains fewer wavelet coefficients.

![wavelet coefficients plots](image)

**Level 0 (LL Approximation)**
**Level 1 (LL Approximation)**
**Level 2 (LL Approximation)**
**Level 3 (LL Approximation)**
**Level 4 (LL Approximation)**
**Level 5 (LL Approximation)**

*Fig. 2.3.2 The goodness of fit of the generalized Gaussian representation of subbands of pyramid image*

### 2.3.2 Feature Extraction

The proposed system considers the features at an optimum level, which is discussed in the previous section. The optimum level color subband image is segregated into RGB colors. The color is the most basic quality of visual content and therefore it is possible to use color to describe and represent the subband image. Based on the RGB color intensity values, the statistical features such as spectrum of energy ($\mu$) and relationship
among the coefficients are computed using the expression given in Eq. (2.3.3) and (2.3.4).

\[
\text{Spectrum of energy} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} |X_{ij}|
\]  

(2.3.3)

where \(N \times N\) is the size of the optimum level, \(X_{ij}\) are wavelet coefficients, and the spatial relationship among coefficients in colors are

\[
\mu = E(x) = \begin{bmatrix} E(x_r) \\ E(x_g) \\ E(x_b) \end{bmatrix} = \begin{bmatrix} \mu_r \\ \mu_g \\ \mu_b \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{r,r} & \sigma_{r,g} & \sigma_{r,b} \\ \sigma_{g,r} & \sigma_{g,g} & \sigma_{g,b} \\ \sigma_{b,r} & \sigma_{b,g} & \sigma_{b,b} \end{bmatrix}
\]  

(2.3.4)

where \(\mu_r, \mu_g\) and \(\mu_b\) are the mean of red, green and blue components of the wavelet coefficients at the optimum level of the pyramid image. These mean values represent most features of the image while the covariance matrix \(\Sigma\) represents the interaction or spatial relationship among the pixels in the color image [CHU97, ERC12]. The diagonal and non-diagonal elements of the covariance matrix represent the strength of the relationship of the pixels within the color and between the colors respectively. Based on these features, a feature vector is formed which is shown in Eq. (2.3.5)

\[
f = [\mu, \Sigma]
\]  

(2.3.5)

where \(f\) is the feature vector. The entire process of the proposed system is presented in an algorithmic form:

**algorithm**

*Step 1: Read color image and convert into RGB color space.*
Step 2: Call DWT to decompose an image and construct an optimum level pyramid image.

Step 3: Compute the spectrum of energy and spatial relationship among the coefficients at the optimum level of the pyramid image using Eqs. (2.3.3) and (2.3.4).

Step 4: Construct feature vector using step-3 and store them in feature space.

Step 5: Find the similarity between feature vectors of query and target images using Bhattacharyya similarity and Cosine distance methods.

Step 6: The recall and precision methods are adopted to perform the accuracy of the proposed system.

Step 7: Stop.

2.3.3 Image Retrieval Method

This section discusses the Bhattacharyya distance measure and Orthogonal cosine similarity method in order to obtain better retrieval accuracy and fitness of wavelet coefficients in the proposed system. The subsequent section provides the performance measure in terms of precision and recall method.

2.3.3.1 Bhattacharyya Measure

The Bhattacharyya measure (BM) [BHA43] is employed to measure the distance between the query and target images. The features are not only sufficient for retrieving the image from the database, but the distance based method also ensures the enhancement of retrieval accuracy. The BM
statistically measures the distance between the images and it has the symmetric and rotation invariant properties [THA97]. The rotation invariant ensures that the magnitude of mean (energy) and covariance (spatial relationship between pixels) is independent of any planar rotation of a pattern around its center of mass. The query and target image(s) are assumed to be Gaussian distributions [SEE07a, SEE07b] as discussed in section 2.3.1 and the optimum level color image is modeled into RGB space. Thus, the RGB image is assumed to be a multivariate distribution. The feature vector \( \mu \) and \( \Sigma \) are computed in section 2.3.2 and is applied in Eq. (2.3.6) to measure the similarity between the query and target images in the database.

\[
D(q, p) = \frac{1}{8} \left[ \mu_q - \mu_p \right]^T \left[ \frac{\Sigma_q + \Sigma_p}{2} \right]^{-1} \left[ \mu_q - \mu_p \right] + \frac{1}{2} \ln \left( \frac{2}{\sqrt{\det(\Sigma_q) \det(\Sigma_p)}} \right) \tag{2.3.6}
\]

The BM method is employed at various levels (L_0 to L_5) of query and target images and the obtained experimental results are presented in Table 2.3.1. Here, it is observed that there is no significant difference among the levels (produces same distance value at all levels), which is demonstrated graphically in Fig. 2.3.3. Hence, it is evident that the features obtained at Level 5 are enough for retrieving the target images. The time and space complexities can be reduced considerably at this level, because it contains fewer dominant wavelet coefficients. Moreover, fewer addition and division operations are involved in deriving the optimum level image.
Table 2.3.1 The distance between the query and target images using Bhattacharyya method

<table>
<thead>
<tr>
<th>Image</th>
<th>L₀ 256 x 256</th>
<th>L₁ 128 x 128</th>
<th>L₂ 64 x 64</th>
<th>L₃ 32 x 32</th>
<th>L₄ 16 x 16</th>
<th>L₅ 8 x 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title0000 Vs. Title0001</td>
<td>0.1146</td>
<td>0.1154</td>
<td>0.1163</td>
<td>0.1181</td>
<td>0.1193</td>
<td>0.1196</td>
</tr>
<tr>
<td>Wood Vs. Wood</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Titles0003 Vs. Titles0004</td>
<td>0.3275</td>
<td>0.3299</td>
<td>0.3525</td>
<td>0.3644</td>
<td>0.3753</td>
<td>0.3782</td>
</tr>
<tr>
<td>Flower Vs. Squirrel</td>
<td>0.4503</td>
<td>0.4772</td>
<td>0.4868</td>
<td>0.4761</td>
<td>0.4738</td>
<td>0.4965</td>
</tr>
<tr>
<td>Decose0006 Vs. Brickpaint000</td>
<td>0.7277</td>
<td>0.7318</td>
<td>0.7571</td>
<td>0.7636</td>
<td>0.7789</td>
<td>0.7849</td>
</tr>
<tr>
<td>Lena Vs. Lena</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Prisonwindow.000 Vs. Sand.0008</td>
<td>1.3061</td>
<td>1.334</td>
<td>1.4192</td>
<td>1.4653</td>
<td>1.4823</td>
<td>1.4901</td>
</tr>
<tr>
<td>Valleywater.0002 Vs. Groundwatercity.000</td>
<td>2.0000</td>
<td>2.1251</td>
<td>2.2300</td>
<td>2.2411</td>
<td>2.2609</td>
<td>2.2754</td>
</tr>
</tbody>
</table>

Fig. 2.3.3 Graphical representation between query image and target image database.

2.3.3.2 Cosine Distance Method

Furthermore, to emphasis the efficiency of the proposed method, orthogonality test is employed on the query and target images by using Cosine distance method as defined in Eq. (2.3.7). Orthogonal property assumes that the query and target images are independent of each other. In two-dimensional space, the Cosine distance compares the feature vector of two images and returns the Cosine of the angle between them. The angle θ
can be described as a function of \(q\) and \(p\), where \(q\) and \(p\) are formed from the coefficient of Discrete Haar Wavelet Transform (DHWT) of the optimum level image and the obtained results are ...

\[
\cos q p
\]

Table 2.3.2

<table>
<thead>
<tr>
<th>Image</th>
<th>Radian</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title0000 vs. Title0001</td>
<td>0.9476</td>
<td>16.4954</td>
</tr>
<tr>
<td>Wood vs. Wood</td>
<td>1.0000</td>
<td>0</td>
</tr>
<tr>
<td>Titles0003 vs. Titles0004</td>
<td>0.9785</td>
<td>11.8830</td>
</tr>
<tr>
<td>Flower vs. Squirrel</td>
<td>0.9870</td>
<td>9.2372</td>
</tr>
<tr>
<td>Decose0006 vs. Brickpaint000</td>
<td>0.9745</td>
<td>12.9638</td>
</tr>
<tr>
<td>Lena vs. Lena</td>
<td>1.0000</td>
<td>0</td>
</tr>
<tr>
<td>Prisonwindow.000 vs. Sand.0005</td>
<td>0.9305</td>
<td>17.4839</td>
</tr>
<tr>
<td>Valleywater.0002 Vs. Groundwatercity.000</td>
<td>0.9776</td>
<td>12.2941</td>
</tr>
</tbody>
</table>

2.3.3.3 Performance Measure

Precision and recall [ZHO00] are the standard performance metrics used to measure the effectiveness of the CBIR system in retrieving most similar images and they are mentioned in Eqs. (2.3.8) and (2.3.9).
where \(|.|\) returns the size of the set, precision \(P(q)\) represents the ratio of the number of images relevant to the query \(q\) among retrieved images to the number of retrieved images; recall \(R(q)\) represents the ratio of the number of images relevant to the query \(q\) among the retrieved images to the number of images relevant to the query in an image database. Hence, the precision versus recall jointly represents the percent of the images relevant to the query image retrieved. The average \((F)\) precision and recall is computed using the formula expressed in Eq. (2.3.10).

\[
R(q) = \frac{|X(q) \cap Y(q)|}{|X(q)|}
\]

(2.3.9)

where \(|.|\) returns the size of the set, precision \(P(q)\)

2.4 Multiresolution with Statistical Features Based Method

The wavelet transform [SMI94] is a powerful tool for characterizing the image features under time-frequency domain since it decomposes the image into various subbands which have multi-scale frequency and orientations. Recently, Discrete Wavelet Transform (DWT) has become popular in image retrieval applications [ANT92], since it provides multiresolution capability, good energy compaction and adaptability to human visual system. Daub-4 DWT [DAU90] decomposes an image into a pyramid structure of sub-images with various resolutions corresponding to the different scales. Since DWT divide the image into an approximation and a detail subband applying through low-pass \((h)\) and high-pass \((g)\) filters and the output of the filters are decimated by 2 is shown in Fig.2.4.1. The approximation represents the
low resolution of the sub-image and detail coefficients provide information about the changes in the horizontal, vertical and diagonal directions, respectively.

**Fig. 2.4.1** Daubechies-4 Wavelet Transform

The approximation and detail functions of Daubechies-4 wavelet transform are computed using a set of four coefficients.
The decomposition process is presented in the algorithmic form:

\[
\text{Procedure DecomposePixel}( m[0\ldots n])
\]

\[
\text{Begin}
\]

\[
\text{While } n > 4 \text{ do}
\]

\[
\text{Begin}
\]

\[
\text{For } i = 0 \text{ to } n-3 \text{ do}
\]
The approximations are stored in half of the upper array \([a_0...a_3]\) and wavelet coefficients are stored in another half of the lower array \([c_0...c_3]\). Subsequently, half of the upper array represents the approximation coefficient, and it is given as input to the next step in the wavelet decomposition. This iteration process is continued until the result oriented optimum level is constructed, which is discussed in the next section. In Fig. 2.4.3 shows the 2-level wavelet decomposition of an image with size \(n \times n\).

**Fig. 2.4.3** Two levels wavelet decomposition of an image using Daubechies-4

The block diagram of the proposed retrieval method is presented in Fig.2.4.4, the query image submitted to the retrieval system is converted to HSV color images. Then each component image
decomposed into a wavelet coefficients $W_{m,n}^{C}, C \in \{H,S,V\}$ which is shown in Fig. 2.4.5, where $m \in \{LL,HL,LH,HH\}$ denotes the subbands scale frequency and $n \in \{1,\ldots,Z\}$ represents the orientations. In the HSV color space, the H and S components are closely related to chrominance information and V component represents the wider bandwidth which contains most texture information. The proposed system extracts features such as color autocorrelogram from $W_{m,n}^H$, and $W_{m,n}^S$ wavelet components, and texture features at $W_{m,n}^V$ wavelet component. Based on these features, the feature vector is constructed.

**Fig. 2.4.4** Block diagram of the proposed system
The overall process of the proposed system is presented in the algorithmic form:

**algorithm**

Step 1: Read image $I$ and convert it into HSV color space.

Step 2: Decompose HSV color image using Daubechies-4 wavelet transform and construct an optimum level pyramid.

Step 3: Compute color autocorrelogram from $W^H_{m,n}$ and $W^S_{m,n}$ subbands and a set of texture features at $W^V_{m,n}$ subband at pyramid level.

Step 4: Find the similarity between computed feature vectors of query and target images by Minkowski-form distance and consider the sorted distance values for $k$-number of relevant images.

Step 5: Adopt recall-precision methods to compute the performance of the proposed technique.

Step 6: stop.
2.4.1. Feature Extraction Techniques

In this section, the proposed system considers two feature extraction techniques: one is the color feature extraction method such as color autocorrellogram and the other is the texture feature extraction method such as coarseness, contrast and directionality.

2.4.1.1 Color Feature for CBIR

Color is one of the most widely used features for image retrieval method. Images are characterized by color features and have many advantages such as robustness, effectiveness, implementation simplicity, computational simplicity, low storage requirements. In the subsequent subsections, we describe color histogram, color correlogram and color autocorrellogram. Among these, the color autocorrellogram method is adopted in the proposed system.

Color Histogram

The most commonly used methods to represent color feature of an image is the color Histogram. A color histogram is a type of bar graph, where the height of each bar represents an amount of a particular color of the color space being used in the image [GER92]. The bars in a color histogram are named as bins and they represent the x-axis. The number of bins depends on the number of colors in an image. The number of pixels in each bin denotes y-axis, which shows how many pixels are of a particular color in an image. The color histogram not only easily characterizes the global and regional distribution of colors in an image, but also be invariant to rotation about the view axis.
In color histograms, quantization is a process where number of bins is reduced by taking colors that are similar to each other and placing them in the same bin. The quantization process reduces the space required to store the histogram information and time to compare the histograms. Obviously, quantization reduces the information regarding the content of images; this is the trade-off between space, processing time, and accuracy in results [LEW01]. One of the disadvantages of histogram is that it ignores the spatial organization of colors. The histogram of sample image is presented in Fig. 2.4.6.

![Lena image and its Color Histogram.](image)

**Fig. 2.4.6** Lena image and its Color Histogram.

**Color Correlogram**

Color correlogram describe the global distribution of local spatial correlation of colors, it is easy to compute the size of the feature vector is fairly small, it is stable to tolerate large appearance changes, and it is also scalable to large image databases. Color Correlogram [HUA97] represents the spatial correlation of colors in an image, whereas the color histogram
captures only the color distribution of an image and does not include any spatial information. While the computing and storage costs of correlograms match those of histograms, the presence of spatial information makes the former more stable to tolerate large image appearance changes than the latter. This makes the correlogram very attractive for applications such as content-based image retrieval and cut and scene detection etc.

Let
significant computational benefits over color correlogram [FEN03], it is more suitable for the purpose of image retrieval than the color correlogram.

The color autocorrelogram [HUA97] of image \( I \) provide the probability of finding identical colors at a fixed distance and gives significant computation benefits over the color correlogram [OJA01] which is described as

\[
\alpha^k(l) = \Pr\{ p' \in I \mid \|p - p'\| = k \text{ and } p' \in I(l) \text{ for } p \in I(l) \subset I \} \\
l \in \{0,1,...,L-1\}
\]

(2.4.2)

where \( \Pr[\ . \] denotes the probability satisfying a given condition, \( I \) is the set of all pixels in an image, and \( I(l) \) is the set of pixels of color \( l \) which is quantized with \( L \) levels. Selecting the Manhattan distance method for distances between pixels, the distance between pixels \( p \) and \( p' \) whose coordinates are \((x, y)\) and \((x', y')\), respectively, is expressed as

\[
\|p' - p\| = \max\{|x' - x|, |y' - y|\}
\]

(2.4.3)

In Eq. (2.4.2) shows that given any pixel color \( l \) in the image, \( \alpha^k(l) \) represents the probability that a pixel of distance \( k \) from the given pixel is on color \( l \). As \( k \) is varied in Eq. (2.4.2), spatial correlation in various resolutions between identical colors can be obtained. Fig. 2.4.7 shows the example of two images with the same number of gray levels but different spatial distributions. We can confirm from Fig. 2.4.7 that the color histograms of the two images appear identical, but their color autocorrelagrams differs from histogram, which are shown as Table 2.4.1.
Fig. 2.4.7 Images with the same number of gray levels, but the different spatial distributions.

Table 2.4.1 Computation of color histogram and color autocorrelogram for images shown in Fig. 2.4.7.

<table>
<thead>
<tr>
<th>Color schemes</th>
<th>Image (a)</th>
<th>Image (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color Histogram</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Color autocorrelogram</td>
<td>$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \times \frac{1}{4} = \frac{1}{8}$</td>
<td>$\frac{3}{8} + \frac{3}{8} + \frac{3}{8} \times \frac{1}{4} = \frac{3}{8}$</td>
</tr>
</tbody>
</table>

Fig. 2.4.8 shows the procedure for extracting color features from decomposed frequency subbands such as...
After constructing a number of quantization levels using Lloyd algorithm, the color autocorrelogram in Eq.(2.4.2) is employed to extract the color features at optimum level.
measurements for human subjects. The first three attained very successful results in CBIR system and are used in the proposed texture feature extraction. The process of texture feature extraction on
\[ E_{k,h}(x,y) = \left| A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y) \right| \quad (2.4.6) \]

At each point, one then picks the best size which gives the highest output value, where \( k \) maximizes \( E \) in either direction. The coarseness measure is then the average of \( S_{opt}(x, y) = 2^k_{opt} \) over the
and then for every position \((n_0, n_1)\)

\[
\theta = \frac{\pi}{2} + \tan^{-1} \frac{\Delta_T(n_0, n_1)}{\Delta_H(n_0, n_1)}
\]  \hspace{1cm} (2.4.8)

is calculated. These values are then histogramized in a 16 bin histogram \(H\), and the directionality can be calculated as the sum of the second moments around each peak from valley to valley. Based on these texture features, the texture feature vector is constructed \(f_T\)

\[
f_T = [E, F_{con}, \theta]
\]  \hspace{1cm} (2.4.9)

where \(E, F_{con}, and \theta\) are represent coarseness, contrast and directionality respectively.

The visual feature vector
different features of vector dimension and component variances in the similarity computation.

### 2.4.2 Similarity Measure

In image retrieval, features are not only playing an important role, but similarity method also ensures the enhancement of retrieval accuracy. Instead of exact matching, content-based image retrieval calculates similarities between the computed feature vector of query image and each target image in a database. Accordingly, the retrieval result is not a single image, but a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution [RIT08] of features. Minkowski [AND98] distance is a simple method leading to very efficient computation, which in turn makes image ranking scalable (a quality that greatly benefits real-world applications) and is given in equation (2.4.11)
Table 2.4.2 shows the sample distance values obtained from each decomposition level (\(L_0\) to \(L_5\)) between the query and target images using Eq. in (2.4.11). It is clearly seen that the distance values are significantly almost same for all decomposition levels (\(L_0\) to \(L_5\)). Therefore, the proposed CBIR system extracts the features at optimum level, which are considered as sufficient for retrieving image from image database, since the optimum level contains few dominant wavelet coefficients without losing the significant features. It is observed that the experimental results presented in Table 2.4.2 show that the proposed method is invariant for scaling transformation, since the same distance value for various sizes of an image (\(level_0\) to \(level_5\)).
Further, to strengthen the efficiency of the proposed technique, we have presented a graph which comprises the distance values against decomposition levels. The curves show that the distance values are significantly same for all decomposition levels (L_0 to L_5) for the given sample images. For example, the query image ‘leaves’ and target image ‘leaves1’ obtain the distance values almost same for all the decompositions. This shows that the proposed system extracts the features at fifth level instead of the first level. Since the fifth level contains a few coefficients compared to that of the first level.

2.5 Wavelet Based Orthogonal Polynomial Model

Wavelet based orthogonal polynomial technique is build with a set of carefully chosen low-order and high-order polynomial coefficients, which are used to reconstruct the multiresolution subband structure and is used to extract the low-level color and texture features present in the image under analysis.
2.5.1 Multiresolution Structure with Orthogonal Polynomial Model

The multiresolution reordering subband structure is built with a set of low-order and high-order polynomial coefficients are presented. The linear combination of orthogonal polynomial [BER97] model is used to construct scaling and wavelet functions in a multiresolution domain by re-grouping low-order and high-order polynomials coefficients. In general, multiresolution decomposition of an image can be analyzed as linear transformations of original values into a set of basis of the transformed coefficients. The wavelet transform of a function $f(t) \in V_J$ vector subspace in $L^2(R)$ [FRO03] can be defined as

$$f(t) = \sum_{kk} c_{j0}(k) \varphi_{j0,k}(t) + \sum_{kk} \sum_{j=j_0}^{j-1} d_j(k) \psi_{j,k}(t)$$

(2.5.1)

where $c_{j0} = \langle f(t), \varphi_{j0,k}(t) \rangle$ and $d_j(k) = \langle f(t), \psi_{j,k}(t) \rangle$ are scaling coefficients and wavelet coefficients respectively, $\varphi(t)$ and $\psi(t)$ are scaling and wavelet functions. If the signal $f(t)$ is in polynomial form, the coefficients $c$’s and $d$’s are linear combination of different order moments [BUT94]. If the moment of wavelet function is zero up to certain order $n-1$, for any polynomial with order lower than $q$, all its wavelet coefficients will be zero. The approximation properties of wavelet basis are the wavelet of length at least $2^{N-1}$ and polynomial degree $N-1$ has vanished inner product [RIT08] with the wavelet at all scales. The polynomials up to degree $N-1$ are reproduced by linear combination of scaling functions. Therefore, the signal falling into a scaling function or scaling space has the power to represent a
polynomial of degree up to $N-1$ [WAN08]. In this connection, the proposed system incorporates the linear combination of orthogonal polynomial to construct the scaling functions and wavelets by re-grouping of low and high order polynomials coefficients [FRO03] via

$$
\varphi_{ji}(x) = \sum_{k=0}^{2^j} a_{jik} P_k(x)
$$

$$
\psi_{ji}(x) = \sum_{k=2^j+1}^{2^{j+1}} b_{jik} P_k(x)
$$

(2.5.2)

where $\varphi_{ji}(x)$ and $\psi_{ji}(x)$ are scaling and wavelet functions respectively, $a_{jik}$ and $b_{jik}$ are coefficients of scaling and wavelet functions. Both scaling function and wavelet are time localized [WAN08] and the wavelet coefficients represent higher frequency. Wavelet and scaling functions spanning an orthogonal basis of a multi-resolution analysis [YOU08, KRI12] and need to fulfill the following orthogonal conditions:

$$
\langle \varphi_{ji}, \varphi_{ji} \rangle_w = \delta_j,
$$

$$
\langle \psi_{ji}, \psi_{ml} \rangle_w = \delta_j \delta_{jm}
$$

$$
\langle \varphi_{ji}, \psi_{ml} \rangle_w = 0, (m \geq j)
$$

(2.5.3)

where $w$ represents the weight, and $\langle \varphi_{ji}, \psi_{ml} \rangle_w = 0$ represents orthogonal between scaling and wavelet function [BER97], we obtain the following equations
\[ \langle \varphi_{jl}, \varphi_{jl} \rangle = \sum_{k=0}^{2^j} \sum_{n=0}^{2^j} a_{ijk} a_{mln} < P_k, P_n > w \]
\[ \langle \psi_{jl}, \psi_{ml} \rangle = \sum_{k=2^j+1}^{2^{j+1}} \sum_{n=2^m+1}^{2^{m+1}} b_{ijk} b_{mln} < P_k, P_n > w \]
\[ \langle \varphi_{jl}, \psi_{ml} \rangle = \sum_{k=0}^{2^j} \sum_{n=2^{j+1}}^{2^{m+1}} a_{ijk} b_{mln} < P_k, P_n > w \]

(2.5.4)

It is obvious that the orthogonality of the polynomial \( < P_k, P_n > w = \delta_{kn} \) the choice of the coefficients \( a, b \) alone determines the orthogonality properties of the basis. Therefore, we get

\[ \langle \psi_{jl}, \phi_{jl} \rangle = \sum_{k=0}^{2^j} a_{ijk}, \]
\[ \langle \psi_{jl}, \phi_{jl} \rangle = \delta_{jm} \sum_{k=2^j+1}^{2^N} b_{ijk}, \]
\[ \langle \phi_{jl}, \psi_{ml} \rangle = 0(m \geq j) \]

(2.5.5)

Here we have used the Legendre polynomials with associated weight \( w(x) = 1 \) as basis function [FRO03]. The orthogonal polynomial coefficients \( c_j \) and \( d_j \) obtained from the Eq.(2.5.1) are reordered to provide an image subband in a multiresolution decomposition like structure. Transforms the coefficients are structured into \((3 \log_2 N + 1)\) multiresolution subbands where \( N \) power of 2 is.

### 2.5.2 Wavelet Packet

Wavelet packet transform (WPT) is a generalization of the dyadic wavelet transform (DWT) that offers a rich set of decomposition structures. The WPT was first introduced by Coifman et al. [COI92, WIC91] for dealing with the non-stationarities of the data better than DWT does. The WPT is
associated with a best basis selection algorithm. The best basis selection algorithm decides a decomposition structure among the library of possible bases, by measuring a data dependent cost function. The generic wavelet transform decomposes the image into approximation and detail components; in the next level, it considers only the approximation components, which is decomposed again and again into approximation and detail component. But WPT decomposes the image into approximation and detail components; further similarly both approximation and detail components are again decomposed into approximation and detail, this process similar to the complete binary tree. As a result, in the case of generic transform, the detail component are left, there are many possibilities to lost the information; but at the same time in the case of WPT, we considers the both approximation and detail components, there is no possibilities for lose the information. Thus, the WPT yields better retrieval results compare to that of generic wavelet transforms. Here in this research, the WPT is considered to retrieve the images.

Wavelet packet transform is a square integrable modulated waveform well localized in both time-frequency and it has three parameters, namely: scale (resolution level) parameter $j$, position-localization parameter $k$ (translation level) and oscillation parameter $p$, which can be defined as

$$W_{j,p,k}(x) = 2^{-j/2} W_p(2^{-j} x - k)$$ (2.5.6)

To generate a sequence of wavelet packets by inducting the following sequence of function:
\[ W_n(x), n = 0, 1, 2, \ldots \] by

\[
W_{2n}(x) = \sqrt{2} \sum_{k=0}^{2^{N-1}} h(k)W_n(2x - k)
\]

\[
W_{2n+1}(x) = \sqrt{2} \sum_{k=0}^{2^{N-1}} h(k)W_n(2x - k)
\]

(2.5.7)

Where \( W_0(x) = \phi(x) \) the scaling is function and \( W_1(x) = \psi(x) \) is the wavelet function. Wavelet packet decomposition produces tree with a total of \( N \log_N \) coefficients as shown in Fig. 2.5.1.

![Wavelet packet decomposition](image)

**Fig. 2.5.1** Wavelet packet decomposition

The Wavelet packet decomposition of binary tree at depth level 2 using the daubechies-4 (Daub-4) wavelet and Shannon entropy \([COI92]\) of a diabetic retinopathy image is presented in Fig. 2.5.2, which is used to extracts the features.


Fig. 2.5.2 Wavelet packet transform of a diabetic retinopathy image decomposition tree level at 2.

2.5.3 Feature Extraction

The proposed wavelet orthogonal polynomial technique converts the given input query image to the HSV color space. The input query image is decomposed to 2-level multiresolution structure such as $LL_1, LLL_2, HH_1, HHHH_2$. The color feature extracted on $LL_1$ and $LLL_2$ subbands using color autocorrelogram technique and the texture features are extracted from the subband such as $HH_1, HHHH_2, HHLL_2$, using the GLCM (Gray-level cooccurrence matrix) based Heralic texture feature information available in the subbands. The overall feature extraction process is depicted in Fig. 2.5.3.

The WPT transformed orthogonal polynomial coefficients of query image is submitted to proposed retrieval system, and is it converted into HSV.
color space with three H, S and V components. Each subcomponent image $I_c, C = \{H, S, V\}$ is then decomposed and transferred to a wavelet coefficients $W^{C}_{m,n}, c = \{H, S, V\}$ using WPT, where $m = \{LL, HL, LH, HH\}$ denotes the subbands and $n = \{1, \ldots, z\}$ represents the orientations [YOU08]. The H and S components are closely related to chrominance information and V component represents the wider bandwidth which contains most texture information. Extract the features such as color autocorrelogram from $W^{H}_{m,n}$ and $W^{S}_{m,n}$ components and set of texture features at $W^{V}_{m,n}$ component for each orientation $n$, and then combine these different modalities of features to construct the feature vector. The similarity method is applied on feature vector of both query image and target images, and return $k$ relevant images from image dataset according to the distance values indexed in ascending order.

**Fig. 2.5.3** Block diagram of proposed system
The proposed system considers the features for an effective combination of color and texture. The color and texture are the abstract information which is embedded in an image. The color autocorrelogram constructed from low-order orthogonal polynomial coefficients based on scaling function. In addition, the low-order orthogonal polynomial coefficients perfectly reordered into approximation function of wavelet [BER97]. The texture features extracted based on wavelet detail function, related to high-order polynomial coefficients.

### 2.5.3.1 Color Autocorrelogram

In order to extract the color features, color autocorrelogram method is used in various approximation subband, which is discussed in section 2.4.1.1. Fig.2.5.4 shows the procedure for extracting color features from decomposed frequency sub-bands such as $W^H_{m,n}$, and $W^S_{m,n}$. The $Q^H_{m,n}$ and $Q^S_{m,n}$ denote color quantization of frequency sub-bands. Each frequency subband is quantized using Lloyd algorithm [GER92].

![Fig. 2.5.4 Process of color feature extraction.](image)

After constructing a number of quantization levels for given image $I$, color autocorrelogram is employed to extract the color features from
Fig. 2.5.5 Co-occurrence matrix generation for \( Ng=4 \) levels and four different offsets: \( P_H(0^\circ), P_{RD}(90^\circ), P_V(45^\circ) \) and \( P_{LD}(135^\circ) \).

It is observed that two neighboring pixels \((2,0)\) of the input image \( I \) is revealed in \( P_H \) concurrence matrix as 3, since there are 2 occurrences of the pixel intensity value 2 and intensity value 0 adjacent to each other in the input image \( I \). The neighboring pixels \((0,2)\) with occur again 2 times in \( P_H \), which makes these matrices symmetric. In the same procedure other matrices \( P_V \), \( P_{RD} \), and \( P_{LD} \) are computed. The four matrices are used separately for image retrieval, then the final decision is formed by fusing the four decisions. As these matrices are symmetric, it is more convenient to use the upper or lower diagonal matrix coefficients in forming the vectors. So, instead of having a vector length of \( Ng \times Ng \), the vector size is reduced to \((Ng \times Ng + Ng)/2\) which helps to speed up the process without
affecting the recognition performance. The process of the texture feature extraction process is given in Fig. 2.5.6.

![Diagram of texture feature extraction process](image)

**Fig. 2.5.6** Process of texture features extraction.

Haralick [HAR73], et al. proposed a method to consider the relative frequencies for which two neighboring pixels are separated by a distance on the image. The following set of Haralick statistical texture features are computed based on GLCM using
correlates a pixel to its neighborhood over $HHH1$ subband, and $f_3$ measures
the closeness of distribution of elements in the $HHH2$ subband. The
diagonal distribution oriented features $f_4$ computed from $HHLL2$. Based on
these features, texture feature vector is constructed in the Eq. 2.5.14

$$f_T = [f_1, f_2, f_3, f_4, f_5, f_6, f_7]$$

(2.5.14)

where $f_T$ represents the texture feature vector.

### 2.5.3.3 Feature Vector Formation

The color feature vector $f_c$ and the texture feature vector $f_t$ expressed
in Eqs. (2.5.8) and (2.5.14) are combined, which form the feature vector as
shown in Eq. (2.5.15), since the combined features yield better retrieval
performance when compared to that of the standalone.

$$f = \left[ \frac{f_c}{\mu_c \sigma_c}, \frac{f_t}{\mu_t \sigma_t} \right]$$

(2.5.15)

where $f$ denotes the combined feature vector space, $f_c$ and $f_t$ are the color
and texture feature vector of dimension $\mu_c$ and $\mu_t$, and $\sigma_c$ and $\sigma_t$ are their
standard deviation vector obtained from respective database. The combined
color and texture features component is normalized by its dimension and
standard deviation, since it reduces the effect of the different features
vector dimension and component variance. The feature extraction process
of the proposed work is presented in the form of flowchart in Fig. 2.5.7.
2.5.4 Similarity Measure

In order to validate the extracted features and for better retrieval, the similarity measure, Manhattan distance [SER08] is adopted, and is expressed in eq. (2.5.16)
where $q$ and $t$ represent the query and target images in the database; $f_i(q)$ and $f_i(t)$