CHAPTER 2

LITERATURE REVIEW

In this era of extensive computerization, digitalization is the major revolution in the field of science and Engineering. There is an inexhaustible mine of digital information available on the internet, Cloud database, and Grid database and even in Mainframes. Now researchers are letting their ideas bloom on Big Data. With so much of data, information retrieval has become one of the active research areas. Information is the knowledgeable data about any concept in particular. There are different forms of data available to express the facts about any specific domain. Image is one of the rich sets of disclosure, which provides the intelligent fact about any domain. Image retrieval has been one of the active research domains since 1970’s, as denoted by Rui et al(1999). This domain is the combination of Data management and Computer Vision research areas.

As per the survey of (Huang 1999), Long et al(2003), and Datta et al(2008), image retrieval techniques are classified as Text based image retrieval and Content based image retrieval. Owing to the enormous research activities in this area, the classification is extended in different ways. With the focus being on the aspect of reducing the semantic gap, the literature review in this thesis is presented as shown in Figure 2.1.

Some of the most widely used image retrieval systems are Text based image retrieval, Content based image retrieval, and Ontology based image Retrieval techniques that are reviewed here.
2.1 TEXT BASED IMAGE RETRIEVAL

An image can be expressed in text by annotating it in relation to the object around it. There are certain types of annotation like manual annotation, semi-automatic annotation and automatic annotation. Linguistically the images are all indexed in this annotation concept. Russell et al.(2008) had designed a framework LableMe, where each object on the image can be annotated by selecting it. Indexing of Semi-automatic and automatic image annotation is one of the blooming research areas. In this technique, the images are all annotated using the visual feature of the image. For web images Schroff et al.(2011) designed a framework. Here the image and the non-image are identified and the unwanted images are removed from the search. Then from the text around the image in the web pages are used to index the image. The authors also used probabilistic methods for image annotation. Carneiro et al.(2007) classified the image class using the probability latent semantic (pLSA) concept; the type of feature used to classify was not explained clearly.

In text based image retrieval, the text around the image such as image caption, tag, description, etc., is used to index the image linguistically.
For querying and retrieving the images, the mathematical model such as set theory based standard Boolean or fuzzy set, algebraic model representation and query using the technique such as vector space, latent semantic and neural network are used. Probabilistic mathematical model such as Probability latent semantic, Bayesian network, Latent Dirichlet allocation and so on is used.

2.2 CONTENT BASED IMAGE RETRIEVAL

Content-based image retrieval famously called CBIR, uses the low-level features of the image as a variable to compare and search a given image which was also specified by Smualders et al(2000). As the images are available in different formats and in different pixel arrangements, the features of two images, which look alike, but in different formats are not always the same. To make an efficient search, in recent years, intelligent Machine learning techniques such as supervised or unsupervised learning was employed. Figure 2.2 shows the basic block diagram of CBIR, where the whole technique is divided into two phases. In training phase, the training images of specific domain are used to extract the low-level feature from the image. The different types of low-level feature used to be extracted are all elaborated in the forthcoming survey. These features vector values are indexed in a database. So, in the testing phase, for any given test image (from the given dataset) the same set of low level features are all extracted and compared with the built database using similarity measure such as Euclidean measure, Manhattan measure, Chi-square measure and so on ,which was also specified by Rui et al(1999). If the measure provides an excepted threshold value, then those images are considered as perfect match and the relevant images are shown to the user in hierarchical order. To enhance the performance of CBIR, in recent year’s researchers started to employ machine-learning technique to the areas as shown in Figure 2.2.
In Content Based Image Retrieval (CBIR), the Content is said to be the basic feature of the images. There are nearly uncountable number of features extraction algorithms available as the research of CBIR was started in early 70’s. Therefore, generally from the work of Bernal et al (2009) these features are divided into global features and local features. Mostly researchers worked on global features such as color, shape, and texture based algorithms as listed in Figure 2.3. Some of the extensively used concepts for extracting these global features are elaborated.

### 2.2.1 Color Feature Extraction

Colors in an image provide tremendous amount of information. Using this color information images can be segmented, analyzed, labeled, and indexed. In content-based image retrieval system, color is one of the basic primitive features used. A color is a parameter, which depends upon the frequency of light. Most of the basics of color images are studied from Sridhar (2011) image processing book. In digital image processing the colors are represented as mathematical co-ordinates called Color Models. As specified by Gonzalez (2007) the commonly used color models are RGB,
HSV, HSI, CMY, and YCbCr models with each model having its own characteristic.

Figure 2.3 List of Global low level features

Some of the commonly used color feature extraction concepts are conventional color histogram, Fuzzy color histogram, color correlogram, color moment, MPEG 7 dominant color descriptor, MPEG 7 color layout descriptor, and MPEG 7 scalable color descriptor. The graphical representation of distribution of given data is called Histogram. It represents the number of times the data occur in the given dataset. Thus for the given color space, the conventional color histogram calculates the probability density function of the given continuous random variable, which is the image. These concepts identify the relative likelihood of the color pixels in the image.

To represent and explore the observed uni-variate data, histogram is the oldest technique used. To generate histogram the data are partitioned to a given reference interval into n bins with respect to the number of occurrences. The major drawback of histogram is that it is a discontinuous function. The choice of interval and number of color bin has an effect on estimated density. In recent years, it has been suggested by researchers to replace the binary
partition by fuzzy partitioning. As described by Han & Kai (2002) and Zheng & Min (2013) in Fuzzy color histogram instead of determining the number of occurrences of a pixel in one color space, here it was considered to determine the number of n color bins using fuzzy-set membership function. These concepts eliminate the drawback of color histogram.

The correlogram was used to analyze the correlation statistics of the data as described by Huang et al (1997). Color correlogram used to find the spatial correlation of pairs of colors changes with respect to the distance. The small value of distance is sufficient to index the image with respect to the color pair. The drawback of this method is the high dimensionality of the feature space.

The color moment depends upon the distribution of colors in an image. If the colors of an image follow certain probability distribution, then the moments of that distribution can be used as feature of that image. The moments are scaling and rotation invariant. Depending upon the color model the number of color moments used are 9 for RGB color space and 12 moments if the color model is CMYK. The first three color moments such as mean, standard deviation, and skewness are often used which encode the shape and color fact of the image under any lighting condition.

Among the list of Color descriptors, Dominant color is best suitable for local image. For given images maximum of eight dominant colors were identified and labeled with unique numbering. The feature vector for each dominant color can be calculated as

\[ F_{\text{MPEG7}} = \{C_i, P_i, V_i, S_c\} \text{ Where } i=\{1,2,3,\ldots,N\} \]  

(2.1)
In Equation (2.1) $C_i = i^{th}$ Dominant color, $p_i$ = Percentage in 5 bits, $V_i$ = Color variance in 3 bits, $S_c$ = spatial coherency in 5 bits, $N$ = Total number of quantized colors in a region in 3 bits.

Manjunath et al (2001) and Salembier et al (2002) are the masterminds of MPEG 7 Visual feature descriptor. In Color Layout Descriptor, the image is partitioned into $8 \times 8$ blocks and in each block dominant color is determined. For each $8 \times 8$ block Discrete Cosine Transform for Y, $C_r$, and $C_b$ color is determined and quantized for the required bit and using the Zigzag scanning the values are tabulated in matrix form. DCT is widely used for image compression because of its high energy packing capabilities.

In Scalable, color descriptor the image is converted into HSV color space and the histogram of 256-color bin is generated. From the generated histogram, the values are all normalized using Haar transform, which is a square integral function, where a two-histogram value is modified by finding the sum and difference between the value and substituting the difference and sum of the values. Let $v_1$ and $v_2$ be the two-histogram values. First the sum $s= v_1 + v_2$ and the difference $d_{hv} = v_1 - v_2$ are calculated. Then the modified substituted values for $Mv_1 = s - (d/2)$ and $Mv_2 = s + (d/2)$ respectively are found.

### 2.2.2 Texture Feature Extraction

The spatial pixel arrangement of an image is said to be the texture of the image. A texture can be analyzed and calculated. The image textures are classified as Structural, Statistical, Model-based, and Transformation based textures. Some of the commonly used texture features are explained. The structural texture is used to analyze well-patterned image and it is not suitable for local images.
As per Materka & Michal (1998) texture review, the Tamura feature is a statistical texture analysis method. It is designed with respect to human visual perception; it extracted the features such as contrast, coarseness, directionality, regularity, roughness, and line-likeness.

The Wavelet, Gabor, and Contourlets transformation are some of the widely used Transformation based texture analysis. As per Jayaraman (2009) image-processing book, in Wavelet kind of transformation the image is transformed at multi resolution depending upon the desired frequency. This kind of transformation will not provide the directionality and anisotropy of the image. Contourlets are implemented by using desire filter bank that would decouple the multiscale and the directional decompositions. For multiscale and direction decomposition, different filters can be used.

Normally to determine the repeated pattern in the image, it has to be transformed either by using Gabor filter or by Wavelet filter and from the transformed image, the texture patterns are determined; otherwise, the statistical features of the texture are determined. Some of the statistical features are coarseness, directionality, regularity, contrast, line-likeness, roughness and so on. These features work well for Brodatz textures pattern. The WOLD features of periodicity, randomness and directionality have been proved to work well on Brodatz textures patterns also.

Repeated patterns in an image are said to be the Texture. In MPEG 7, there are three different ways to represent a texture. Among them Texture browsing description is used for the browsing of heterogeneous texture patterns. For all the three texture vectors, require 12 bit as shown in equation (2.2). For Homogenous texture pattern from the work of Ro et al (2001) the texture vectors includes the energy, mean, standard deviation etc. kind of texture analysis and used as feature vector.
\[ T_{\text{MPEG7}} = \{ t_1, t_2, t_3, t_4, a_1, \ldots, a_n \} \] (2.2)

In Texture browsing descriptors, these vectors for an image provide the details about texture regularity, directionality, coarseness and its scale.

Edge Histogram descriptor is used to determine the edges of the images. By default, the image is divided into $4 \times 4$ sub images and the edge magnitude of the each sub image is determined by comparing the standard edge provided by the MPEG 7, which is Vertical edge, Horizontal edge, $45^\circ$ edge, $135^\circ$ edge, and non-directional edge.

Let $I_k(i,j)$ represent the sub image block. The filter coefficient of each edge vector is given by Equation (2.3)

\[
    f_{cv}(k) = \text{Vertical edge filter} \\
    f_{ch}(k) = \text{Horizontal edge filter} \\
    f_{cd45}(k) = 45^\circ \text{edge filter} \\
    f_{cd135}(k) = 135^\circ \text{edge filter} \\
    f_{cn}(k) = \text{non directional edge filter}
\]

To calculate the vertical edge magnitude for a sub-block $(i,j)$ is given by

\[
    M_{v}(i, j) = \left| \sum_{k=0}^{64} I_k(i, j) \ast f_{cv}(k) \right| 
\]

(2.3)

2.2.3 Shape Feature Extraction

In general Shape of the image is extracted and classified as contour based, region-based, space domain and transform domain based methods. The
approaches are classified as 1-D function approach, Polygonal approach, spatial interrelation feature, Moments, Scale-Space methods and shape transform domain methods. The most commonly used feature is Region moment, Invariant moment, Zernike moments and Legendre moments, Fourier descriptor and histogram of edge direction.

In images moments are the quantitative measure of the shape of the image with respect to the set of points in the images. The shape of the images can be ascertained with respect to the region or boundary of the image. Among the region based descriptor there exist different kinds of moments.

For a 2-dimensional grayscale image, Let $I(x,y)$ be the image intensities of each pixel. Then the raw- moment (first order moment) $M_{ij}$ is calculated as shown in Equation (2.4)

$$M_{ij} = \sum_x \sum_y x_i y_j I(x,y) \quad (2.4)$$

This is said to be the general Region moment of the image.

In general, moments describe numerical quantities at some distance from a reference pixel point. Regular Cartesian moments have the form of projection over $(x_i y_i)$ which is not orthogonal. Thus, the regular moments contain redundant information. In Zernike-moments, as specified by Khotanzad & Yaw (1990) are a set of complex polynomials which form a complete orthogonal set over the interior of the unit circle.

Zernike moments are the mappings of an image onto a set of complex Zernike polynomials. As they are orthogonal to each other, the Zernike moments can represent the properties of an image with no redundancy or overlap of information between the moments. It solely depends upon the scaling and translation of the object in a region of interest. As the
magnitude of the moment is independent of the rotation angle of the object on image, it can be utilized to describe the shape characteristics of the object. For Zernike moment calculation, only the pixels within the unit circle of the resulting normalized image, I(x,y) are used for subsequent calculations. The Zernike moment of order $n$ with repetition $l$ for a continuous image function $I(x,y)$ that vanishes outside the unit circle is shown in Equation (2.5)

$$Z_{nl} = \frac{n+1}{\pi} \sum_x \sum_y I(x, y)V_{nl}^*(\rho, \theta) \text{ Where } x^2+y^2 \leq 1$$

Where in Equation (2.5) $\rho$ specifies the length of vector from origin pixel to (x,y) pixel and $\theta$ specifies the angle between vector $\rho$ and x pixel in the counterclockwise direction.

The Curvature Scale Space (CSS) approach was used to represent shape for planar curves in the image. The image object boundaries are usually represented as planar curves that do not cross over them. As explained by the authors Abbasi Sadegh et al(1999) the curvature zero crossing of a curve are points where the sign of curvature changes.

The overall literature review is shown in Table 2.1. In Table 2.2, some of the notable research works and prototypes, which continue to provide a foundation for multimedia retrieval systems that are in vogue today, are tabulated.
<table>
<thead>
<tr>
<th>Author</th>
<th>Low-Level feature</th>
<th>Learning Model Used</th>
<th>Dataset Used</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Chai 2011)</td>
<td>SIFT</td>
<td>SVM</td>
<td>102 Oxford flower dataset</td>
<td>76.3%</td>
</tr>
<tr>
<td>Fukuda K et al (2008)</td>
<td>Shape</td>
<td>Fuzzy C-Mean</td>
<td>Not specified</td>
<td>71.1%</td>
</tr>
<tr>
<td>Halaschek-Wiener et al (2005)</td>
<td>MPEG 7</td>
<td>Ontology</td>
<td>Not Specified</td>
<td>50%</td>
</tr>
<tr>
<td>Hao et al (2010)</td>
<td>297-dimensional feature vector</td>
<td>'Random Walk' algorithm</td>
<td>COREL</td>
<td>70%</td>
</tr>
<tr>
<td>Hong et al (2004)</td>
<td>Color histogram</td>
<td>SVM</td>
<td>885 general flowers</td>
<td>Not specified</td>
</tr>
</tbody>
</table>
Table 2.1 (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Low-Level feature</th>
<th>Learning Model Used</th>
<th>Dataset Used</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hsu et al (2011)</td>
<td>Color and Shape</td>
<td>SVM</td>
<td>102 flower dataset</td>
<td>89%</td>
</tr>
<tr>
<td>Kebapci et al (2011)</td>
<td>SIFT</td>
<td>KL Divergence</td>
<td>380 flower Image</td>
<td>88%</td>
</tr>
<tr>
<td>Kherfi et al (2007)</td>
<td>Color, shape, and texture</td>
<td>Probabilistic Model</td>
<td>COREL</td>
<td>60%</td>
</tr>
<tr>
<td>Li et al (2012)</td>
<td>Not specified</td>
<td>Probabilistic Model</td>
<td>FLICKER Pool</td>
<td>58%</td>
</tr>
<tr>
<td>Lindstaedt et al (2009)</td>
<td>MPEG 7</td>
<td>SVM</td>
<td>FLICKER Pool</td>
<td>74%</td>
</tr>
<tr>
<td>Lu et al (2008)</td>
<td>MPEG 7</td>
<td>SVM</td>
<td>COREL</td>
<td>70%</td>
</tr>
<tr>
<td>Nilsback et al (2006)</td>
<td>Color and SIFT</td>
<td>SVM</td>
<td>102 Oxford flower</td>
<td>82.7%</td>
</tr>
<tr>
<td>Nilsback et al (2008)</td>
<td>SIFT</td>
<td>K-mean clustering</td>
<td>102 Oxford flower</td>
<td>72.8%</td>
</tr>
<tr>
<td>Qi et al (2007)</td>
<td>Histogram and modified edge histogram</td>
<td>SVM</td>
<td>COREL</td>
<td>83.25%</td>
</tr>
</tbody>
</table>
Table 2.1 (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Low-Level feature</th>
<th>Learning Model Used</th>
<th>Dataset Used</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saitoh et al (2003)</td>
<td>Color and shape</td>
<td>K-mean clustering</td>
<td>Wild Flower and leaf (20 set of flower)</td>
<td>95%</td>
</tr>
<tr>
<td>Saitoh et al (2004)</td>
<td>Color</td>
<td>SVM</td>
<td>Wild Flower</td>
<td>96.8%</td>
</tr>
<tr>
<td>Schroff et al (2011)</td>
<td>Not specified</td>
<td>SVM</td>
<td>Google Image</td>
<td>47.6%</td>
</tr>
<tr>
<td>Su et al (2010)</td>
<td>Not specifically specified</td>
<td>Fuzzy membership function</td>
<td>Not Specified</td>
<td>76%</td>
</tr>
<tr>
<td>Tianzhu et al (2011)</td>
<td>Shape</td>
<td>Multiple instance learning model and a AdaBoost</td>
<td>KTH</td>
<td>95.33%</td>
</tr>
<tr>
<td>Uijlings et al (2010)</td>
<td>SIFT, SURF, DAISY and Semantic Texton</td>
<td>Random Forest, PCA and Spatial Pyramid</td>
<td>FLICKER Pool</td>
<td>70%</td>
</tr>
<tr>
<td>Zhang et al (2006)</td>
<td>MPEG 7</td>
<td>Not specified</td>
<td>Not Specified</td>
<td>60%</td>
</tr>
<tr>
<td>Author</td>
<td>System</td>
<td>Summarization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Niblack et al (1993), and Flickner et al (1995) | QBIC | - In this prototype system, query would be an image or rough sketch.  
- Used the RGB composition of color, uses coarseness, contrast and directionality features of Texture  
- Shape features use area, circularity, eccentricity, major axis orientation and a set of algebraic moment invariants of shape  
- No relevant feedback system is implemented |
| Smith & Shih (1997) | VisualSeek | - In this prototype system the query for the image retrieval system would be the color of the needed image  
- For precise result the user has to specify the spatial location, minimum size and color composition  
- This work will not provide efficient result for this age internet images. |
| Tom et al (1996) | MARS | - It is one of biggest projects started at University of Illinois supported by NSF/DARP/NASA and so on.  
- They try to provide a standard way of multimedia analysis and retrieval system  
- They too use the basic content of image such as color, texture, shape and layout  
- They mainly concentrate on new approach towards segmentation, shape representation and to support complex queries  
- Still researches are going on in this project. |
Table 2.2 (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>System</th>
<th>Summarization</th>
</tr>
</thead>
</table>
| Smith & Shih (1997) | VisualSeek | • In this prototype system the query for the image retrieval system would be the color of the needed image  
|                  |        | • For precise result the user has to specify the spatial location, minimum size and color composition  
|                  |        | • This work will not provide efficient result for this age internet images. |
| Tom et al (1996) | MARS  | • It is one of biggest projects started at University of Illinois supported by NSF/DARP/NASA and so on.  
|                  |        | • They try to provide a standard way of multimedia analysis and retrieval system  
|                  |        | • They too use the basic content of image such as color, texture, shape and layout  
|                  |        | • They mainly concentrate on new approach towards segmentation, shape representation and to support complex queries  
|                  |        | • They also try to accomplish their project for compressed images  
|                  |        | • Still researches are going on in this project. |
| Chang et al (1997) | VIDEOQ | • This prototype also comes under query by image or sketch.  
|                  |        | • In addition to considering color, shape and texture of the image, it mainly concentrates on spatial temporal constrain  
|                  |        | • The motion vector of the key object searched is considered. |
Table 2.2 (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>System</th>
<th>Summarization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carson et al (2002)</td>
<td>Blobworld</td>
<td>• In this prototype system the main functional areas are the feature extraction and multidimensional image indexing and retrieval system</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The features such as color histogram, color correlogram and wavelets are extracted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Here the author uses the concept of Exception Maximization (EM), an algorithm used to learn a model along with its hidden variable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The extracted features are clustered using the concept of Gaussian clustering by determining the likelihood between them.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The Minimum Description length principle is used to determine the output</td>
</tr>
<tr>
<td>Robles et al (2004)</td>
<td>Using Wavelet</td>
<td>• This retrieval procedure uses the concept of wavelet feature extraction using Haar transform</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• From the transformation the researcher used two types of feature vectors multi-resolution global color histogram and multi-resolution local histogram</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• This system will provide relevant result if used locally</td>
</tr>
</tbody>
</table>
2.3 BRIDGING THE SEMANTIC GAP APPROACH

Initially in CBIR, researchers were keen in deriving different algorithms for extracting the exact feature from the images. All those features had provided high retrieval rate on different test image sets. However, these features have not reduced the semantic gap between the high-level and low-level content of the images. Therefore, instead of studying the state-of-the-art techniques of low-level feature extraction from the image, in this work the techniques used to narrow down the semantic gap were studied. Only through machine learning method, (i.e) the methods used to integrate low-level feature with high-level concept and query concept, will reduce the gaps. Figure 2.4 shows the concept of Semantic gap. The gap between human interpretations of high level, human language based image semantic concept with the image content-based low-level feature is said to be the semantic gap. This gap can be narrowed using the techniques like annotation, semantic visual template, and Ontology.

![Figure 2.4 Procedure to handle Semantic Gap](image)

2.3.1 Annotation Based Approach

The images can be annotated either by using the keyword or by using the local features. Some of the works with respect to annotation are
discussed here. In the work of Jeon et al (2003) the images are manually segmented into blobs. By using the probability reasoning such as cross-relevance models the blobs are categorized and annotated. The specific type of feature vector used has not been explained. In the work of Gao et al (2006) the images are converted to text through the codeword concept. The specific detail of conversion has not been specified. From the converted text using LSI the annotation is performed. Zhang & Ebroul (2006) tried to annotate the image using the MPEG 7 features. In this work, the images are sub-divided into blocks. From the block images, MPEG7 features are all extracted and the images are all annotated manually. These annotations are analyzed using the basic of retrieval concept. Halaschek- Wiener et al (2005) created an ontology based image framework Photostuff. The Photostuff is built with some predefined ontology. So the user can upload any image and annotate it with respect to the available keyword from the already build ontology.

For web image based annotation Peng et al (2007) tried to annotate the web image using Hidden Markov model. From the contest of web document, the keywords are determined from the web pages and using the basic of HMM and knowledge of Wordnet they tried to annotate the image automatically. However, the way the images are related to the identified keyword has not been clearly elaborated in their paper. Image feature based classification and annotation was proposed by Qi & Yutao (2007). They used two different sets of SVM. Initially multiple instance learning (MIL) was used to classify the image block. This classification was enhanced using global feature based SVM, where the global features such as color histogram and modified edge histogram were used to classify the image. Thus, classified image was annotated accordingly. They used COREL dataset. MPEG 7 based feature extraction and annotation was once an active research area. Hentschel et al (2007) used this MPEG 7 based general features and a visual-word dictionary was generated manually using these features. Using this dictionary
as backend, for the given text based query, the relevant image was retrieved. Lu et al(2008) used the MPEG 7 features, for the selection of appropriate feature Genetic algorithm, based Bi-Coded chromosomes were used. For the classification SVM classifiers was employed to annotate the image. For web image, based annotation Lindstaedt et al(2009) used the Folsonomy of the tagged web images and the low level feature with respect to MPEG 7 was used to narrow the semantic gap. For classification, they used SVM classifier. Su et al(2010) tried to reduce the semantic gap using fuzzy membership function. Form the web document both visual, textual data were extracted, and model was created. For image analysis, the algorithm used was not discussed; just the creation of model was specified. From the text, the keywords were all extracted. Both the models were fused with fuzzy set. This concept was implemented for the retrieval process. Again, for social web images, Schroff et al(2011) designed a text based query system. From the retrieved image the relevant image with respect to the given query was classified using SVM classifiers. The authors Li et al(2012) classified the images using the probability reasoning technique by analyzing the tagging of social media images. The summary of annotation based information retrieval is shown in Table 2.3.

Table 2.3 Summary of Annotation based information Retrieval

<table>
<thead>
<tr>
<th>Annotation based approach</th>
<th>Author</th>
<th>Image Feature</th>
<th>Textual feature</th>
<th>Model Used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jeon et al(2003)</td>
<td>Images are segmented into blobs but specific features were not specified</td>
<td>Manual Annotation</td>
<td>Cross-relevance models</td>
</tr>
</tbody>
</table>
Table 2.3 (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Image Feature</th>
<th>Textual feature</th>
<th>Model Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halaschek-Wiener et al (2005)</td>
<td>MPEG 7</td>
<td>Keywords</td>
<td>Ontology</td>
</tr>
<tr>
<td>Peng et al (2007)</td>
<td>Not specified</td>
<td>Keywords from WordNet</td>
<td>Hidden Markov model</td>
</tr>
<tr>
<td>Qi et al (2007)</td>
<td>Histogram and modified edge histogram</td>
<td>Not specified</td>
<td>Multiple instance learning based SVM</td>
</tr>
<tr>
<td>Hentschel et. al. (2007)</td>
<td>MPEG 7</td>
<td>Not specified</td>
<td>Visual Word template</td>
</tr>
<tr>
<td>Lu et. al. (2008)</td>
<td>MPEG 7</td>
<td>Annotation</td>
<td>Bi-Coded chromosome Genetic algorithm and SVM classifier</td>
</tr>
<tr>
<td>Lindstaedt et. al. (2009)</td>
<td>MPEG 7</td>
<td>Web image tag</td>
<td>SVM classifier</td>
</tr>
<tr>
<td>Su et. al. (2010)</td>
<td>Not specified</td>
<td>Keyword</td>
<td>Fuzzy membership function</td>
</tr>
<tr>
<td>Schrooff et. al. (2011)</td>
<td>Not specified</td>
<td>Web image tag</td>
<td>SVM classifier</td>
</tr>
<tr>
<td>Li et. al. (2012)</td>
<td>Not specified</td>
<td>Web image tag</td>
<td>Probability reasoning technique</td>
</tr>
</tbody>
</table>

2.3.2 Visual Template Based Approach

In this approach, a visual codebook kind of template normally called Bag of Visual word (BoW) is created to reduce the semantic gap. The features generally used to create the codebook are derived from Lowe (2004).
SIFT based feature extraction algorithm. As per Tsai (2012) review, with respect to BoW methodology for feature extraction and representation SIFT features were used. For vector quantization of visual feature K-mean clustering algorithm was employed which would generate a nominal rate of 1000 visual words. The learning model used to design the visual template were also paves one of the major contributions of this work. The normally used learning models are SVM,LDA,pLSA and so on. Some of the profound works, which provide the main idea for this thesis, are elaborated.

Uijlings et al(2010) in their paper had evaluated different ways of visual word creation in each step. For feature extraction descriptors such as SIFT, SURF, DAISY and Semantic Texton were used. For visual word creation, they compared the k-mean based visual word classification with Random Forest, PCA and Spatial Pyramid. The classification results were evaluated using chi-square, RBF and Fast Histogram intersection kernel for the SVM. They concluded that SIFT and SURF provided dense visual feature vector and SVM based classification provided accuracy results. They tried to implement their evaluation on Flickr uploaded images. To identify and annotate an action in video Tianzhu et al(2011) used a multiple instance-learning model and an AdaBoost, a discriminative exemplars selection. Initially in the training section actions such, as running, walking, jumping etc., were studied using the SVM based multiple instance-learning model and to annotate the action image they used AdaBoost classifier. They tested their algorithm with KTH dataset and Weizmann dataset. The algorithm to extract the local visual features were not explicitly elaborated in their paper.

Kherfi et al(2007) provides a framework to organize and index the image collection, which can be used for image browsing, summarization, and semantic retrieval. In their work, they proposed a two layered hierarchical structure. Initially the keywords regarding the image collection were provided
in the higher layer. In the lower layer, the local features such as color, shape, and texture were provided. For categorization of the images, the probability based pLSA approach is used as highlighted by the authors Kim & Daijin (2009). The comparative study of this work with other content-based image retrieval system provided satisfactory results. The actual usage of visual feature and their consecutive algorithm and inclusion in pLSA-based classification was not elaborated in their paper. Hao et al (2010) did the actual usage of visual low-level feature and image description. A 297-dimensional feature vector represented grid based color moment; the pixel coordination was determined by local binary pattern (LBP) and normalized edge histogram was extracted from the image. These vectors were combined with the image tag by a hybrid graph approach. The Random Forest based 'Random Walk' algorithm was proposed to traverse the hybrid graph.

**Table 2.4 Summary of visual template based approach**

<table>
<thead>
<tr>
<th>Author</th>
<th>Image feature</th>
<th>Textual feature</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uijlings et. al., (2010)</td>
<td>SIFT, SURF, DAISY and Semantic Texton</td>
<td>Not Specified</td>
<td>Random Forest, PCA and Spatial Pyramid</td>
</tr>
<tr>
<td>Tianzhu et. al.,(2011)</td>
<td>Shape</td>
<td>Action keywords</td>
<td>Multiple instance learning model and a AdaBoost</td>
</tr>
<tr>
<td>Kherfi et. al., (2007)</td>
<td>Color, shape, and texture</td>
<td>Keywords</td>
<td>Probabilistic Model</td>
</tr>
<tr>
<td>Hao et. al., (2010)</td>
<td>297-dimensional feature vector</td>
<td>Image tag</td>
<td>'Random Walk' algorithm</td>
</tr>
</tbody>
</table>
2.3.3 Ontology Based Approach

In ontology based image retrieval system, a domain specific ontology creation is the key contribution. Most of the image retrieval system using ontology had concentrated in creating the ontology with respect to domain knowledge. Initially for Helsinki University museum, Hyvönen et al (2003) created a domain specific ontology. The Museum has heritage images, these images were annotated according to the created ontology terminology. This ontology was used for image retrieval system. As per the annotated keyword, the images were searched through the created ontology. Soo et al (2003) used the concept of Mandarin Chinese Thesaurus. They combined the domain ontology of the Mandarin Chinese and their thesaurus. Then these were combined with the manually annotated historical figures. With all these details, they tried to create an RDF file for query processing.

Schober et al (2004) used the image low-level features like color and texture (the authors have not specified the actual features used in their system) that were extracted from the images. Other than low level feature the textual information such as background membership and spatial relations between the objects in the images were all described in the ontology. The authors called their framework OntoPic. Brooke & Wei (2005) used the annotation concept with respect to web page content. For created hotel, car hire and airline ontology using the inference engine concept the HTML tags were all extracted and populated into the ontology. Osman et al (2007) created an abstract domain ontology for sports domain. In their ontology attributes such as event name, team name, player name and manager name were included as textual attribute and for images; they used the attributes such as size, format, and contrast of the image. In the work of Farah et al(2008) for each satellite image, they had created three types of ontology the scene model ontology Sensor model ontology, and spatial relation model ontology. Then
these ontologies were all merged and used for semantic image retrieval system. For all the created ontology, the image based keyword and certain metric were used. In text based image retrieval system, they had incorporated the idea of ontology to it. The concepts to fill the semantic gap were not elaborated in this technique. For the canine animals, Wang et al (2008) had created the hierarchical domain ontology. For different canine images the basic domain knowledge and image description were used to design the ontology. Then both the textual domain knowledge and image description were merged and used for retrieval system. The author suggested that fusion of textual and image context data would provide better results than the simple text based or content based image retrieval system. Shi et al (2008) had tried to segment the image in region manually. From the segmented image, general features were extracted (which was not specified). Ontology for the collection of images was created. However, the integration of feature with the ontology was not specified. Shareha et al (2009) tried to follow the work of Wang et al (2008). They tried to merge the image ontology with text ontology. Again, the usage of low level feature was not elaborated. They created hierarchical structure for human and animal and they created a text ontology. For image, the annotation with respect to the objects in the images was used. Based with the annotation keyword the image ontology was created. The researcher has attempted to merge these two ontologies, viz., text ontology and image ontology in this work. Koletsis et al (2010) created ontology for Dog images. The MPEG 7 based low-level features were extracted for image retrieval system. The ontology about the text description of the Dog was used for text-based information retrieval. In their dataset, they used the 30 different dog breeds. The procedure to merge the text and low-level ontology was not elaborated in this work. In the work of Yildirim et al (2013) the content of the basketball game video was extracted. The author concentrated only on high level semantic of the videos description. The action, activity of the basket ball game was organized, and a complete ontology for the videos was created.
From the created ontology the exact action and activity of the player or the referee was extracted by indulging the fuzzy concept. Khurana & Chandak (2013) followed the idea of Liu et al (2010) where SIFT feature of the image was extracted from the locomotive image data set. The scores of the images were calculated with respect to the positive and negative image separation and retrieval was done in this study. This concept was implemented in the video by selecting the key frame with locomotive in the image. The image was manually annotated in XML. The authors have not integrated the low-level feature, XML annotation and key frame selection in the ontology.

In all these works, the creation of ontology using textual annotation of the images was used. The combination of textual and visual feature ontology creation was not elaborated in these works. The main problem of this work was to represent pixels of images into Ontology. This problem was one of the main objectives of this work. That is, to understand the image features generically. Creating an ontology generally for images was one of the obstacles in this work, so domain had to be considered before studying the feature of the image. The domain of flower image was considered, and some of the works with respect to flower image were discussed.

Table 2.5 Summary of Ontology based information Retrieval

<table>
<thead>
<tr>
<th>Author</th>
<th>Domain</th>
<th>Approach to create Ontology</th>
<th>Textual Feature</th>
<th>Image Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyvönen et. al., (2003)</td>
<td>Helsinki University Museum</td>
<td>Annotated keyword the images</td>
<td>Not Specified</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Domain</td>
<td>Approach to create Ontology</td>
<td>Textual Feature</td>
<td>Image Features</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------------------------</td>
<td>----------------------------</td>
<td>--------------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>Dai et. al., (2005)</td>
<td>Hotel, Car based domain</td>
<td>Used the HTML tag around the images</td>
<td></td>
<td>Not Specified</td>
</tr>
<tr>
<td>Osman et. al., (2007)</td>
<td>Sports Domain</td>
<td>Names of team, player and manager</td>
<td></td>
<td>Image size, format, and contrast</td>
</tr>
<tr>
<td>Farah et. al., (2008)</td>
<td>Satellite Images</td>
<td>Textual description with respect to scene, sensor and spatial relation</td>
<td></td>
<td>Not Specified</td>
</tr>
<tr>
<td>Su et. al., (2008)</td>
<td>Canine animal</td>
<td>Textual description of canine animal</td>
<td></td>
<td>Descriptive image feature</td>
</tr>
<tr>
<td>Shi et. al., (2008)</td>
<td>Natural Images</td>
<td>Not Specified</td>
<td></td>
<td>Images are sub-divided and general features are used</td>
</tr>
<tr>
<td>Shareha (2009)</td>
<td>Human and animal</td>
<td>Textual keywords</td>
<td></td>
<td>Image annotation</td>
</tr>
<tr>
<td>Koletsis et. al., (2010)</td>
<td>Dog</td>
<td>Not Specified</td>
<td></td>
<td>MPEG 7 features</td>
</tr>
<tr>
<td>Yildirim et. al., (2013)</td>
<td>Basketball</td>
<td>Description of Action and activity of the game</td>
<td></td>
<td>Not Specified</td>
</tr>
<tr>
<td>Khurana et. al., (2013)</td>
<td>Locomotive</td>
<td>XML description</td>
<td></td>
<td>SIFT</td>
</tr>
</tbody>
</table>
2.3.4 Flower Image Retrieval

In image-retrieval research area, flower image retrieval is one of the major sub-domains. Mostly a few authors continually tried to provide an effective way of recognizing given input flower image with respect to their created test images.

In the work of Das et al (1999), the flowers were indexed with respect to the dominant color, which were recognized using the iterating segmentation algorithm. Using the domain knowledge of nearly 300 flower images in the dataset, the images were segmented and they tried to recognize the images. Obviously providing only color as feature to index the multi-dimension image will not provide the satisfactory results. For wild flower recognition, Saitoh & Toyohisa (2003) specified certain term and rule for taking the image of the flower and leaf, which is an inconvenient and laborious technique to handle images. Then 10 features for flowers, such as petal shape by defining the ratio of length to height of petal, number of petals, flower's moment with respect to the surface area, Roundness with respect to flower area, 6 values with respect to the primary and secondary color of the image were extracted. For leaf 11 different features such as aspect ratio with respect to the shape of the flower, moment, roundness and bias center of gravity, stem angle, tip angle, leaf structure, three feature with respect to the color of the leaf were all extracted. These extracted features were compared linearly; the recognition rate of this work is 95% with respect to 20 sets of flower and leaf image from 16 different species.

In the work of Hong et al (2004) the flowers were segmented using region-of-interest concept. Then using the domain knowledge of the flower image the color histogram was extracted. Other than the color histogram of the images, two more shape features were extracted. The features were centroid contour distance curve and angle code histogram. In centroid contour
distance, the distance between the centroid with respect to the edges of the
flower was calculated. For angle code histogram, the angles between two
approximate lines drawn at the contour points were determined. They used
885 flowers from 14 different species. They outlined that their work
outperformed Das et al(1999) which was obvious.

Zou et al(2004) created a framework called Computer Assisted
Visual InterActive Recognition. Initially the flower regions on the training
images were segmented interactively. The ground-truth outlines of training
flower rose images were used as rose-curve model on the test images. The
model was generated using eight different parameters such as petal number,
ratio of rose outer radius to the inner radius, first three moments of hue and
saturation being extracted and used. For the given input rose image’s
silhouette was compared with the trained images ground-truth. This
framework had facilities for re-computing the model parameter and re-ranking
the recognized result. Saitoh et al(2004) work had extended his previous
work. For flower image, segmentation Intelligent Scissor approach was
employed to extract the flower from the given image. The features such as
number of petals, central moment, roundness, ratio of the route length to the
sum of distances between the gravity center and all boundary points, 6-color
feature with respect to HS space were all extracted for recognition of flower
image.

Cho & Chi (2005) proposed a back propagation algorithm with
adaptive crossover and mutation operations for image classification. They
applied this algorithm in flower image retrieval system. The results showed
that the structural representation of images provided a promising recognition
rate. However, they were not specific with the number and type of feature
value to be selected. In most of the systems, the accuracy rate for
reorganization mainly depends on the selection of feature values. Nilsson et
al(2006) used the bag of visual word concept. They had created visual words for the oxford image dataset. To create the visual word, first the color of flower was determined by converting the image to HSV color space. Then from the converted images, the K-mean clustering algorithm was used to determine the centroid of the colors. The centroid of the test and training color was compared using Chi-square measurement. In this approach, their system identified totally 55.3% in their first hypothesis. For shape they used the SIFT feature which actually determined the interesting point feature that provided them the recognition rate of 82.7%. For Texture they used Zisserman's MR8 filter bank which provided the recognition rate of 56.0%

Again Nilsback et al(2008) had tried to improve the previous work by trying to identify some of the exact features of the flower image. They used HSV based k-mean clustering centroid; SIFT for background and foreground image and Histogram of edges. Then they classified using the Multi kernel SVM classifier. Fukuda et al(2008) tried to recognize the flower with respect to the structure such as many-petalous, single-petalous, and sympetalous flower. For identifying the shape, the power spectrum group was determined. For the classification, they used the fuzzy c-mean clustering algorithm whose membership function depends upon the shape of the image. For the experiment, they used 448 images from 112 species taking four images per species. Hsu et al(2011) had created an interactive flower-image recognition system. Initially for this system, the user had to draw an appropriate bounding window to segment the flower object from the image. From the boundary, segmented image nearly 18 different colour and shape features with respect to the flower petal and stamen/pistil were extracted from the image. The author had justified that their system had outperformed the recognition rate of Hong et al(2004) and Satioh et al(2003) system. The intelligent way of identifying the petals and stamen/pistil was not specified
clearly in their report. To reduce the computation time of Nilsback et al (2008) work, (Chai 2011) initially segmented the flower image using the super pixel based framework. Initially the input image was divided into smaller sub-regions called super-pixel. Then using graph cut segmentation concept the foreground flower image was segmented from the background. Then the SIFT feature of the segmented images were all used to classify the images using linear kernel SVM classifier and it would reduce the computation time of the system.

**Table 2.6 Summary of Flower Image Retrieval**

<table>
<thead>
<tr>
<th>Author</th>
<th>Flower type</th>
<th>Features</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Das et. al., (1999)</td>
<td>Not specified</td>
<td>Color</td>
<td>Not specified</td>
</tr>
<tr>
<td>Saitoh et. al., (2003)</td>
<td>Wild Flower and leaf (20 set of flower)</td>
<td>Color and shape related feature of flower and leaf</td>
<td>95%</td>
</tr>
<tr>
<td>Hong et. al., (2004)</td>
<td>885 general flowers</td>
<td>Color histogram, Centroid contour distance curve, and angle code histogram</td>
<td>Not specified</td>
</tr>
<tr>
<td>Zou et. al., (2004)</td>
<td>Rose</td>
<td>The petal number, ratio of rose outer radius to the inner radius, first three moments of hue and saturation</td>
<td>Not specified</td>
</tr>
<tr>
<td>Saitoh et. al., (2004)</td>
<td>Wild Flower</td>
<td>The number of petals, central moment, roundness, ratio of the route length to the sum of distances between the gravity center and all boundary points, 6 color feature with respect to HS space</td>
<td>96.8%</td>
</tr>
<tr>
<td>Author</td>
<td>Flower type</td>
<td>Features</td>
<td>Recognition Rate</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Saitoh et. al., (2004)</td>
<td>Wild Flower</td>
<td>The number of petals, central moment, roundness, ratio of the route length to the sum of distances between the gravity center and all boundary points, 6 color feature with respect to HS space</td>
<td>96.8%</td>
</tr>
<tr>
<td>Nilsback et. al., (2006)</td>
<td>102 Oxford flower dataset</td>
<td>Color, Zisserman's MR8 filter bank and SIFT</td>
<td>82.7%</td>
</tr>
<tr>
<td>Nilsback et. al., (2008)</td>
<td>102 Oxford flower dataset</td>
<td>HSV based k-mean clustering Centroid; SIFT for background and foreground image and Histogram of edges.</td>
<td>72.8%</td>
</tr>
<tr>
<td>Fukuda K et. al., (2008)</td>
<td>Not specified</td>
<td>Shape based power spectrum</td>
<td>71.1%</td>
</tr>
<tr>
<td>(Chai 2011)</td>
<td>102 Oxford flower dataset</td>
<td>SIFT</td>
<td>76.3%</td>
</tr>
<tr>
<td>Hsu et. al., (2011)</td>
<td>102 flower dataset</td>
<td>18 different colour and shape feature with respect to the flower petal and stamen/pistil.</td>
<td>89%</td>
</tr>
</tbody>
</table>
2.4 SUMMARY

In this chapter, the basic local features and prototypes of content-based image retrieval system were discussed. As the main objective of this work is to reduce the semantic gap, the key concepts used to bridge the semantic gaps were reviewed. The recent work held in the area of annotation based image retrieval approach, visual template based image retrieval approach and ontology based information retrieval approach was discussed. The notable work done in flower based image retrieval system was also reviewed.