CHAPTER 2

LITERATURE SURVEY

In this chapter, an overview on the discussion of how the CBIR methods have been evolved in the research community throughout the years is presented, which provides some of the references for the contributions. Related to an image retrieval important questions have been discussed in the survey such as intent of user, the nature understanding and the scope of the image data as for the search results the different types of queries and visualization. Datta et al. explores CBIR as “any of the technologies which helps in organizing the digital picture archives by the content of visual”. The semantic gap in CBIR is the fundamental concept and difficulty, which normally states as the distance between the semantic similarity of high-level concepts, and the visual similarity of low-level features. The most important of the semantic similarity is depended on the situation. For example if an user wants the images of the sports car to be retrieved, then it is preferred as a measurement of the semantic similarity, therefore the sports car phrase can also contain a definition of rather loose of any of the car which is considered as to build for the speed. Alternatively users are interested in a unique sports car built in 1964-65 which is called as Shelby Cobra Daytona Coupe, then an increase will be there in the importance of the exact visual similarity. Therefore, the semantic similarity for the preference is depended on the application in mind and the query. For the illustration of the growth of CBIR research, Datta et al. have introduced an exercise which is very interesting. The search for publications containing the phrase "Image Retrieval" within each year from 1995 to 2005 is performed using Google Scholar and the digital libraries of ACM, IEEE and Springer. A roughly exponential growth has been found in the interest of image retrieval. For an example, to find the trends in counts of publication they queried Google Scholar with the phrase "image retrieval" together with other CBIR-related phrases. The area of research which is strongly increased in the counts of publications seems to be Categorization or Classification. The images of limited variability is included typically within a narrow domain, such as airplanes, faces, etc., A domain which is broad includes the high variability images, for example a large collection of images which includes the mixed contents which is downloaded from the

internet. In today's life, the separation into narrow and broad domains is a widely used distinct and well-recognized stream.

2.1 Low Level Feature

In representation of image retrieval the shape [9] is a qualified study conducted between the scales, rotation and translation invariant descriptors. Therefore the shape is most widely used feature of image which is exploited in the systems of content-based image retrieval, for each descriptor the number of coefficients are required for indexing and their retrieval performance evaluation. The shape is represented by curvature scale space, image moment descriptors, Fourier and Angular Radial Transform (ART). By using the standard methodology and the two most appropriate available databases; four shape descriptors are evaluated against each other. The moment descriptor shows to be the better shape representation quality where in the lowest descriptor is presented by the ART.

A combination of color and shape [10] descriptor detects the mushroom species. The Centroid-Contour Distance is used for shape extraction and color moments. To get the Receiver Operating Characteristics (ROC) curve the 25% of training data is used as testing data for the accuracy. The quality of accuracy is shown under the curve area that shows the result lie between 0 to 1. On 20 species of mushroom, the accuracy is 0.65751 with 5 data in each species and for 32 species is 0.69649. For the 32 species with 5 data species, 7 data species and 10 data each species the accuracy is 0.69649, 0.65407 and 0.60008 respectively. The performance evaluation is only 60% to 69%.

To implement the CBIR many algorithms and techniques are proposed by many researchers. Based on wavelet transform [12] the performance of different CBIR system by the implementation of combined feature of shape, texture and color as prominent feature is discussed. In the image retrieval, the choice of technique for the feature extraction is to determine the CBIR system performance. On the basis of wavelet decomposition using Local Binary Patterns (LBP), using threshold on wavelet decomposition and wavelet decomposition using morphology operators the performance of the three CBIR system is evaluated on Wang's benchmark image database. When compare to wavelet decomposition using threshold and morphological operators, the retrieval efficiency is better in wavelet decomposition with LBP.

For any computer vision application object identification is very much essential. To detect the object of interest in a scene, object identification algorithm is used. If object of interests are exist then the algorithm can identify the position of objects. An effective object detection approach must identify the object of interests from an arbitrary scene without affecting by scaling or object rotation, alterations in camera viewpoints. Object detection [13] is an art used for the identification of known objects and object identification which comes under specific category like car, building, elephant etc., Feature extraction and feature matching after extraction is a common technique used to detect objects which work by identifying the objects using shape feature and with suitable segmentation approach. Object shape can be reflected as object descriptor. The object boundary identified with the help of Discrete Fourier Transform coefficients. Then suitable segmentation approach used on target image. The detected contour of Objects are compared with the current contour objects. The decision of object identification depends on the comparison.

For the content-based image retrieval the most important discriminative elements are color, texture and shape. In the problems of the shape based image retrieval the mainly used descriptors are the Fourier descriptors [14]. It represents a method of retrieving Fourier descriptors from the simplest shape signature – complex coordinates. Normalization with the sum of magnitudes of all harmonics is used instead of the commonly used scale normalization with the magnitude of the first harmonic. The results is improved shape scale normalization. All the results of an experiments which indicates that the proposed method outperforms many other state of-the-art the methods which are based on Fourier descriptors, in the terms of both of computational time and retrieval performance.

According to the image content the CBIR system search the images with texture, color, shape and other prominent features within the image. Throughout these features, the image appearance and the structure of visual can be provided by the shape. Zernike moment et. al., introduces a system for binary image retrieval [15] , angular radial histogram together with vertical and horizontal histogram of the image. The region-based shape approach is mainly focused and the image features are extracted. The query image and database images are processed for similarity-matching. As a result, according

to the rank value the extracted images are shown to the user. Experiments are conducted on MPEG-7 CE Shape-1 Part-B dataset.

For the accounting of human visual perception and the stochastic nature of textures a new metrics for texture similarity [37] is created. According to the human judgment an essentially identical, metrics are completely dependent on the statistics of local image and the substantial point-by-point deviations between textures are been allowed. The ideas of structural similarity is extended by the proposed metrics and in texture analysis-synthesis are been guided by research's. By using a steerable filter decomposition they are been implemented and a concise set of sub band statistics, sliding windows or computed globally is incorporated. A systematic test is performed to investigate performance of metric in the context of "known-item search," the retrieval of textures that are "identical" to the query texture. It results in eliminating the usage for cumbersome subjective tests, therefore on a large database it enables the comparisons with performance of humans.

Based on the bag-of-words (BoW) model the step taken in the retrieval of image is visual matching which is a very crucial step performed. In the method of baseline, if the SIFT descriptors are quantized to the same visual word only then the two key points are considered as a matching pair. Although there are two limitations in the SIFT visual word. Initially, during the quantization the more of its discriminative power is lost. Secondly, the feature of local texture is only described by SIFT. The power of the BoW model is impaired to the discriminated due to both drawbacks which also leads to false positive matches. To handle this type of problems, multiple binary features [38] were introduced at the indexing level. To model correlation between features, a scheme is introduced called as multi-IDF, through which different features of binary are coupled into the inverted file. Based on the binary features the matching verification methods, such as Hamming embedding is incorporated in the framework effectively. As an extension, the fusion of binary color feature is explored into the image retrieval. The precision of visual matching is greatly enhanced by the joint integration of the SIFT visual word and binary feature which reduces the impact of false positive matches.

To improve the performance of image indexing and retrieval algorithm a framework [43] for the partitioning of an image into non-overlapping tiles of different

sizes results in a high level of retrieval efficiency. The two different regions (tiles) of the images are divided and each of tiles transform the coefficients of the energy and standard deviation of Hartley, which serves as the local descriptors of texture, as the sub-features is been extracted. Next, to record the shape features the invariant moments of edge image have been used. A robust feature set for image retrieval is provided by the shape features and combination of sub-features of texture. For matching of textural features the most similar highest priority (MSHP) principle is used and for shape features matching the Canberra distance is required. The image which has less MSHP is the retrieved image and the query image is from the Canberra distance.

During the past few years the CBIR has attracted substantial attention with the help of the potential practical applications. To improve the performance of CBIR systems, a powerful tool as a variety of relevance feedback (RF) schemes have been developed to bridge the semantic gap between the high-level semantic concepts and the low-level visual features. One of the most popular and important techniques in CBIR among all the various RF approaches is support-vector-machine (SVM)- based RF. In spite of being success, there are two main drawbacks while directly using SVM as an RF scheme. Initially, the positive and negative feedbacks are treated equally, which is not appropriate since the distinct properties is been given to the two groups of training feedbacks. Next, the unlabeled samples is not taken into the account by most of the techniques called as SVM-based RF, however for constructing a good classifier they are very much helpful. To overcome these two drawbacks and to explore the solutions, a semi supervised image retrieval [45] with distinct properties of feedbacks can be integrated with SVM-based RF schemes can be utilized. The positive feedbacks are differentiated from the negative ones which is based on local analysis by the biased maximum margin analysis(BMMA) by introducing a Laplacian regularize to the BMMA the SemiBMMA can integrate information of unlabeled samples effectively. This problem is formulated into a general subspace learning task in complete formal way and then an automatic approach of determining the dimensionality of the embedded subspace for RF is proposed.

A great potential is shown by the content-based histology image retrieval [46] systems in support for making a decision in teaching, biological research and clinical

activities. An important key role is played by the feature combination in the content-based image retrieval. It aims to help in the enhancement of the descriptive power of visual features which is corresponded to the semantically meaningful queries. For interpreting varying tissue composition and architecture into histological concepts it is particularly valuable in histology image analysis where the intelligent mechanisms are required. A method of multi-objective learning is the core of this approach, which results in understanding an optimal visual-semantic matching function which is jointly considering as the query images for the different preferences of the group.

The most important research areas which include the applications in multimedia databases, digital libraries and the internet is the CBIR. The main important features which is used in image retrieval are shape, color, texture and spatial relations. For the identification of an object the shape features is the most powerful clues. The retrieval methods were used for the improvement of retrieval accuracy, region-based shape and dissimilarities [47] of contour. In the retrieval performance, a considerable improvement is been caused only when it is assumed that the fusion of two categories of shape description. To coincide semantic and Euclidean distances, the main goal in this study is to propose a new feature vector. To get through this, a distance-driven non-linear feature extraction method has learned the desired topological manifold.

A content-based image retrieval system for skin lesion [53] images as a diagnostic aid. From the skin lesions the rate of correct retrieval of images measures the effectiveness. To retrieve digital images and the name of the disease category from an image data repository the architecture which is proposed is being used, by the contents in the image, such as color, shape and texture which is extracted from the image. For the diagnostic purpose the algorithm uses regression tree, feature vector and classification which is used to retrieve comprehensive reference sources.

For the image retrieval throughout the various approaches, local key point descriptors [54] and the scale space representation have been displayed as a promising approach. The characteristics of the scale space operation and also an extended method of scale space operation is provided which significantly increase the accuracy in the image matching in the context of image retrieval.

In content based microscopic image retrieval [57] system the design, work and development for the microscopic images utilizing a reference database which includes the images of more than one disease are described. To classify and retrieve microscopic images a multitier approach is used which involves their specific subtypes, which are most difficult to classify and discriminate. Among the retrieved images, to protect the semantic consistency both slide-level image retrieval and the multi-image query is been enabled by this system. For multiple-image query and retrieval a term called as new weighting term is defined which is inspired from information retrieval theory.

A mechanism of relevance feedback [59] is incorporated by the CBIR systems in which users identify images as relevant or irrelevant and then retrieval is adapted. This type of relevance decisions are often assumed to be category-based. Even though the database is unfamiliar with it and irrespective of context, the users are forced to decide upon category membership of an image. In the form of relative similarity judgments the alternative is to obtain user feedbacks. Depending on the interface the user gets the ability to provide meaningful feedback which displays, facilitates the feedback and retrieved images. Based on pre-defined, fixed category labels, deeming retrieved results by simulating usage the algorithms for CBIR are mostly characterized empirically as the relevant and if with the query they share a category label.

In computer vision field the detection of human is an active field of research. As the main characters in comic book stories is not trivial which is extending this to human-like drawings. The analysis of the Comics is very modern field of research at the intersection of objects, people recognition, texts and graphics [16]. An essential step towards a fully understanding of an automatic comic book is the detection of the main characters of the comic is an essential step. By using color palette and content-based drawing retrieval provides a color-based approach for the character of comic's retrieval

An efficient content-based image search a novel approximate indexing scheme called as Multi-Sort Indexing [17] introduces scheme analyzes high dimensional image descriptor vectors, by employing the cardinalities value of their dimensions. Due to the existence of methods of several extraction the value of cardinalities significantly vary as it was expected. To reorder the dimensions of the descriptors a multiple sort algorithm as per the cardinalities values is used such that it can increase the probability of two similar

images to lie within a close constant range. Following the probabilistic analyses and deterministic, the expected bounds of the constant range is defined in detail. The introduced scheme is suitable fully for indexing real-time of images with an algorithm of efficient query-processing the retrieving and searching of relevant images. The research is carried out with five real datasets, against the hashing methods the superiority of the proposed approach, and for approximate similarity search is also suitable.

A new hashing method called semi supervised kernel hyperplane learning (SKHL) [18] for semantic image retrieval by modeling each of the hashing function as a nonlinear kernel hyperplane is developed by the unlabeled dataset. Furthermore, to learn the hashing functions and optimal kernel hyperplanes a Fisher-like criterion is used with side information using only weakly labeled training samples. Besides different types of features to be integrated, for leading towards better hashing functions the multiple kernel learning (MKL) is incorporated into the SKHL (called SKHL-MKL). The effectiveness of SKHL and SKHL-MKL is demonstrated by the comprehensive experiments on NUS-WIDE and CIFAR-100.

In many of the computer vision applications the use of high-dimensional features has become a normal practice. Upon the number of data points the highest dimension of these features is a limiting factor, which might be effectively processed and stored. An approach to learn a compact binary encoding [19] exploits both class-label information and pairwise proximity. The development of encodings are allowed for exploiting this extra information which however is compact, in the terms of retrieval performance or final classification the original high-dimensional features is been outperformed. The methods in general, both parametric and nonparametric learning methods are applicable in it. This generality states that for a wide variety of computer vision tasks the embedded features are suitable such as content-based image retrieval and image classification.

An important tool is been developed by Relevance feedback approaches for an interactive search, allowing the user to convey their requirements. A semi-supervised learning [20] is used for the relevance feedback so that it can be used in image retrieval tasks. The aim of the semi-supervised algorithm is using simultaneously both unsupervised and supervised learning approaches. By using the information which is retrieved from the user feedback the supervised step is performed, but the intrinsic dataset

structure is exploited by the unsupervised step, which presents in the terms of ranked lists of images. For different image retrieval tasks many of the experiments performed involving texture descriptors, color, different datasets and shape. The multimodal retrieval tasks, considering visual and textual descriptors is also evaluated. The effectiveness of the proposed approach is demonstrated by the experimental results.

The features are extracted from the Error-Diffusion Block Truncation Coding (EDBTC) [21] to index color images. Two color quantizes and a bitmap image is been produced by the EDBTC to generate the image feature descriptor it is further processed using Vector Quantization (VQ). To measure the similarity between a target image and the query image with in the database the two main features introduced called as Bit Pattern Histogram Feature (BHF) and Color Histogram Feature (CHF). Respectively, VQ-indexed bitmap image and VQ-indexed color quantize computes the CHF and BHF. To measure the similarity between two images the distance computed from CHF and BHF is used. The experimental results outperform with the method of indexing. Thus EDBTC not only examined with good capability but for the content-based image retrieval (CBIR) system it also provides an effective way to index the images.

Modern machine learning techniques and leveraging computational image processing methods is used by the automatic analysis of histopathological images. For the disease detection, decision support in this area and diagnosis both content-based image retrieval (CBIR) and computer-aided diagnosis (CAD) systems have successfully been developed. In current days, the amount of annotated medical data is increasing, to offer a promise of bridging the semantic gap between diagnostic information and images, a large-scale and data-driven methods have emerged. To cope intelligently with massive histopathological images [22] focuses on scalable image-retrieval techniques to be developed. A supervised kernel hashing technique uses a small amount of supervised information which is leveraged in learning such that 10,000-dimensional image feature vector can be compressed into tens of binary bits only with a preserved informative signatures. This signature then inserted into hash table as binary codes and indexed which enables real-time retrieval of images in a large database. Crucially, to bridge the semantic gap between high-level diagnostic information and low-level image features the

supervised information is employed. In terms of retrieval tests 800 queries are executed in only 0:01 seconds and image classification accuracy is 88:1%.

A graph-based query specific fusion [23] approach based on local and holistic features conducting a link analysis on a fused graph where the multiple graphs are merged and re-ranked. By assessing the consistency of the top candidates' nearest neighborhoods the retrieval quality of an individual method are measured on the fly. Therefore, for different query images by using holistic features or local it is capable of adaptively integrating the strengths of the retrieval methods. This methods does not require any kind of supervision, it has very less parameters, and the implementation is very easy. On the four public datasets through extensive experiments have been conducted, i.e., the Holidays, Corel-5K, UK bench and the large-scale San Francisco Landmarks datasets. A very competitive performance is gained including on several data sets the state-of-the-art results, e.g., the N-S score 3.83 for UK bench.

In image retrieval the color is one of the most important low-level features used as the image features by the most of the CBIR systems. However, image retrieval using only color features will not fetch satisfactory results as similar colors do not have similar content. As the result of this problem, Discrete Wavelet Transform (DWT) and Color Edge Detection [24] is used for CBIR. A feature vectors which combines both color and edge features is generated, Manhattan distance is used for similarity measurement to detect the final image rank. Experimental results shows a better perform when the query images are altered.

One of the interesting methods in image processing is segmentation. This method segment the objects in an image. A close reconstructions (SEGON) and dual multi-Scale Gray level morphological [25] used so that a background (BG) gray-level variation mesh can be built, which might results into identifying BG and object regions. On image BG gray levels from a macroscopic perspective was developed and by using standard procedures it was implemented, thus robustly dealing with large-scale database images. To improve object segmentation accuracy the image segmentation capability of existing methods were exploited by the BG mesh. To evaluate the probability of coherent segmentation labeling, the segmentation accuracy, i.e., the normalized probability random index (PRI) between the hand-labeled and a computer-segmented image one

computed for comparisons. To evaluate the object segmentation capability in dealing with large-scale database images the CBIR system was carried out. With and without SEGON the retrieval precision–recall (PR) and rank performances are compared. With the shape feature for the multi-instance, to select salient common feature elements an AdaBoost was used. For the color features, between two scalable HSV descriptors the intersection of histogram was calculated, for multi-instance retrieval the mean feature vector was used. When both positive and negative queries are provided the distance measure for color feature can be adapted. In order to perform multi-instance with multi feature query the normalized correlation coefficient of features among query samples computed to integrate the similarity ranks of different features. The object segmentation method outperforms others by 21% in the PRI which was resulted by the experiments.

The local tetra patterns(LTrPs) [42] computes gray-level difference between the referenced pixel and its surrounding neighbors which is encoded by the local ternary pattern (LTP) and the standard local binary pattern (LBP). The referenced pixel and its neighbors directions calculated using the first-order derivatives in horizontal and vertical directions. Furthermore, to compute n^{th} order LTrP using $(n-1)^{th}$ order vertical and horizontal derivatives a generic strategy was proposed such that to analyze the effectiveness of their proposed algorithm and an efficient CBIR by combining it with the Gabor transform.

An unsupervised visual hashing [44] approach called semantic-assisted visual hashing (SAVH) intention was to extract the rich semantics latently effectively which is embedded in auxiliary texts of images such that the effectiveness of visual hashing can be boosted without any explicit semantic labels. To gain the target, by simultaneously preserving visual similarities of images a unified unsupervised framework is developed such that a hash codes can be learnt, modeling high-order relationships of inter-images on integrating the semantic assistance from auxiliary texts and between images and shared topics characterizing the correlations.

An image retrieval technique known as lifting wavelet-based color histogram [48] is a combination of color histogram (CH) and the Haar wavelet transformation, where CH is rotation and translation invariant. The accuracy of the retrieval system

increases as the Haar wavelet transformation extract the local characteristics and texture features of an image.

Region-based image retrieval [51] is achieved by gaze-based Relevance Feedback (RF) approach. An iterative estimation of the real-world objects are comprised as a fundamental idea where the image retrieval system exploits the user interest for refining the results. The primary novelties are realizing user's relevance assessment prediction at region-level where a new set of gaze features are introduced and designing time-efficient and effective object-based RF framework. Apart from the temporal attributes, the spatial characteristics of the gaze signals are also represented by the proposed features, which is not been studied in the literature so far. In the other words, to overcome the main limitation of region-based RF approaches is the main aim of the developed object-based RF mechanism, i.e., the frequently inaccurate estimation of the regions of interest in the retrieved images.

To improve the performance of content-based image retrieval the relevance feedback (RF) schemes have been mainly designed. RF require the user to label a large number of samples to be labeled is not appropriate. By resorting to the auxiliary information the labeling efforts of the user is reduced as the aim of collaborative image retrieval [52]. As the most informative ones for the user to label with the help of the optimal hyper plane of SVM an ambiguous samples can be selected by the Support vector machine (SVM) active learning and therefore labeling efforts is alleviate of the conventional RF. Although, the optimal hyper plane of SVM is normally as unstable and inaccurate with small-sized training data. Other than the conventional manifold regularization framework, the most informative samples for the user to label in image retrieval have been selected by the new method effectively.

A classification-driven biomedical image retrieval [58] framework is based on similarity fusion and image filtering wherein a supervised learning techniques is employed. In this type of framework, at the first instance multi class SVM classifier is used to reduce the search space to filter out irrelevant images of database and query images are exploited which results for the similarity matching. According to their modalities the Images are classified at a global level which is based on the different low-level, concept, and key point-based features. It is difficult to find a unique feature for all

types of queries. Therefore, by relying on the image classification and feedback information from users a query-specific adaptive linear combination of similarity matching approach is intended. An individual pre computed weights of different features are adjusted online which is Based on the prediction of a query image category. In some of the cases the prediction of the classifier may be incorrect and the users retrieved images may have a different semantic interpretations. Therefore, by considering both precision and rank order information each individual feature weights are finally determined which is represented by considering top retrieved relevant images as the users judged.

2.2 CBIR Techniques

Due to a tremendous improvement of information technology and the simultaneously number of available multimedia data is increasing, based on visual content the task of retrieving information has become a subject of scientific interest. To retrieve images in a semantic way the bag-of-visual-words (BOVW) [26] model is adopted by the recent approaches. Content-based image retrieval tasks a remarkable performance shown by BOVW, over the global and local feature (LF) representations; the better retrieval effectiveness is exhibited. It strongly depends on the performance of the BOVW approach, although the ideal codebook size prediction, a difficult and database-dependent task. There are threefold contribution. Initially, to calculate the most appropriate size of a codebook for a given database the technique uses a self-organized and self-growing neural gas network. In the next technique called as soft-weighting each of the LF is classified into only one visual word (VW). Lastly, the method which automatically detects the number of VWs is derived by combining the information.

The content based image retrieval issue is to narrow down the gap between high-level semantic concepts and low-level image features. In the relevance feedback scheme for the enhancement of performance a biased discriminant analysis with feature line embedding [28] is introduced. At local neighborhoods the margin between irrelevant and relevant samples is tried to be maximized. The subspace is reduced and the possibility of the relevant images are increases.

In recent years, due to the explosive amount of digital images and crowd sourcing tags the Tag-based image retrieval (TBIR) has drawn much attention. Motivated by the

method of subspace clustering [29] where in a subspace-clustering model the tag completion problem is formulated, which is assumed as from the subspaces the images are been sampled, and by using the state-of-the-art Low Rank Representation (LRR) method the tags are completed. Additionally, to refine further the tags a matrix completion algorithm is introduced. Thus introduced algorithm outperforms state-of-the-art approaches when handling missing and noisy tags as shown by the results on multiple benchmark datasets for image annotation.

A generalized formulation of the multiple instance-learning problem is considered. Under this formulation, both positive and negative bags are soft, in the sense that negative bags can also contain positive instances. This reflects a problem setting commonly found in practical applications, where labeling noise appears on both positive and negative training samples. A novel bag-level representation [30] using instances that are most likely to be positive and its ability to separate soft bags, depending on their relative composition in terms of positive and negative instances, is studied. The study inspires a new large-margin algorithm for soft-bag classification based on a latent support vector machine that efficiently explores the combinatorial space of bag compositions. Empirical evaluation on three datasets shown to confirm the main findings of the theoretical analysis and the effectiveness of the soft-bag classifier.

One of the powerful technique in lieu of the popular vector quantization is called successful image-based object recognition technique [31]. The local features that are quite similar can quantized into quite distinct visual words is one of the major drawbacks of the sparse space-based methods. This issue is addressed for the object recognition with a novel approach is called sparse spatial coding, wherein a spatial constraint coding stage and a sparse coding dictionary learning are combined efficiently. An experiment evaluation was conducted on the Caltech 101, Caltech 256, Corel 5000, and Corel 10000 data sets, which is mainly designed for the evaluation object recognition and on scene recognition reported on COsy Localization Dataset (COLD) and on the MIT-67 resulted in high performance.

CBIR is a developing trend in Digital Image Processing for searching and retrieving the query image from wide range of databases. Conventional CBIR scheme has the following limitations: 1. It is slow 2. Difficult to label negative examples; 3.

Accuracy is poor in a single step; 4. Users may introduce some noisy examples into the query. This in turn explores solutions to a new issue that image retrieval using unclean positive examples. CBIR classification using SVM [32] uses two-step strategy in which first step is feature extraction using low level features (color, shape and texture) while SVM classifier is used in the second step to handle the noisy positive examples. Thus, an efficient image retrieval algorithm is based on color-correlogram for color feature extraction, wavelet transformation for extracting shape features and Gabor wavelet for texture feature extraction. Further, multiple features and different distance metrics combined to obtain image similarity using SVM classifier. Results found are encouraging in terms of color, shape and texture image classification. After the features selected, an SVM classifier [62] is used to distinguish between relevant and irrelevant images accordingly.

CBIR uses the visual contents of a picture like global features-color feature, shape feature, texture feature, and local features-spatial domain present to signify and index the image. CBIR method combines global and local features. Haar Discrete Wavelet Transform [33] divides an image into horizontal, vertical, diagonal region and Gray Level Co-occurrence Matrix (GLCM) for feature extraction. SVM is used for classification. The experimental results show improved results.

RGIRS (Remote Geo-system Image Retrieval System) is a system of retrieving similar image using image features like color feature, texture feature and shape feature. Content based image retrieval system extracts features relevant to query image using feature extraction method. Many RGIRS systems are used to retrieve accurate similar image but the problem is no method provides accurate results. Thus Statistical Rule Model [34] retrieve the image from the database using threshold based Euclidean distance calculation. The input query image features extracted and sent to the SVM Model. The SVM will compare input image features with trained features and test the features with distributed density function. Once the image class identified, the features are rearranged and calculate the distance using Euclidean distance calculation. The threshold value is fixed for retrieval. If the calculated distance is less than the threshold, value the image will retrieved or if the distance exceeds threshold value then the image will not be retrieved.

Content Based Image Retrieval (CBIR), which combines color, texture and shape features [35] of the image is based on three noticeable algorithms: color distribution entropy (CDE), color level co-occurrence (CLCM) and invariant moments. CDE takes the correlation of the color spatial distribution in an image into consideration. CLCM matrix is the texture feature of the image, which is grounded on co-occurrence matrix to seize the alteration of the texture. Hu invariant moments frequently used owing to its invariance under translation, changes in scale, and rotation. The retrieval result is better when use the texture or color or shape feature used alone. The similarity measure matrix is based upon Euclidean distance.

An interactive image browsing and retrieval [36] approach is based on relative similarity feedback obtained from 2D image layouts. It incorporates online maximal-margin learning to adapt the image similarity metric used to perform retrieval. A user starts a session by browsing a collection of images displayed in a 2D layout. User may choose a query image perceived to be similar to the envisioned target image. A set of images similar to the query are then returned. The user can then provide relational feedback and/or update the query image to obtain a new set of images. Algorithms for CBIR are often characterized empirically by simulating usage based on pre-defined, fixed category labels, deeming retrieved results as relevant if they share a category label with the query. In contrast, the purpose of interactive image browsing and retrieval is to enable browsing and retrieval without predefined categories. The performance is evaluated on a target-based setting by quantifying the efficiency with which target images are retrieved given initial queries.

Recent years have witnessed a number of studies on distance metric learning to improve visual similarity search in content-based image retrieval (CBIR). Despite their successes, most existing methods on distance metric learning are limited in two aspects. First, they usually assume the target proximity function follows the family of Mahalanobis distances, which limits their capacity of measuring similarity of complex patterns in real applications. Second, they often cannot effectively handle the similarity measure of multimodal data that may originate from multiple resources. To overcome the limitations, an online multiple kernel similarity (OMKS) learning [39] algorithm learns a

flexible nonlinear proximity function with multiple kernels to improve visual similarity search in CBIR.

How to represent an image is an essential problem of the image retrieval task. To build a powerful image representation another method named cross-regions-pooling [40] (CRP) combines two key ingredients. Initially region are detected by objectness detection technique; later deep attributes (DA) that is the outputs of the softmax layer of off-the-shelf convolutional neural network are pre-trained on a large-scale dataset. The ultimate representation of an image is the aggregation (e.g. max-pooling) of DA extracted from all the regions.

An effective feature for color image retrieval based on block truncation coding (BTC) and vector quantization [41] (VQ). Each input colour image is decomposed into Y, Cb and Cr components. BTC is performed on the 4×4 Y blocks, obtaining a mean pair sequence and a bitplane sequence, and then they are quantized with the contrast pattern codebook and visual pattern codebook to obtain the contrast and visual pattern co-occurrence matrix. VQ is performed on the 4×4 Cb blocks and Cr blocks with the Cb codebook and Cr codebook, respectively, to obtain the color pattern co-occurrence matrix.

Content representation for images with well-defined inter-class boundaries in the feature space remains to be a difficult task. Simple distance-based retrieval approach operate on the feature space for content-based image retrieval (CBIR) are claimed to be inefficient by many researchers. Different CBIR approaches have been proposed to surmount the drawbacks of SDR scheme. The class membership-based retrieval [49] efforts to reduce the overall search time of CBIR.

Local binary pattern (LBP) are widely adopted for efficient image feature description and simplicity. To describe the color images, it is required to combine the LBPs from each channel of the image. The traditional way of binary combination is to simply concatenate the LBPs from each channel, but it increases the dimensionality of the pattern. In order to cope with this problem a method called multichannel decoded local binary patterns [50] is adopted. Here the adder and decoder based two schemas for the combination of the LBPs from more than one channel is introduced. Image retrieval

experiments are performed to observe the effectiveness of the approaches which shows better performance when compared with the existing multichannel techniques.

With many potential industrial applications, content-based image retrieval (CBIR) has recently gained more attention for image management and web searching. As an important tool to capture users' preferences and thus to improve the performance of CBIR systems, a variety of relevance feedback (RF) schemes have been developed in recent years. One key issue in RF is: which features (or feature dimensions) can benefit this human-computer iteration procedure? The theoretical and practical comparisons between principal and complement components of image features are made in CBIR RF. Most of the previous RF approaches treat the positive and negative feedbacks equivalently although this assumption is not appropriate since the two groups of training feedbacks have very different properties. That is, all positive feedbacks share a homogeneous concept while negative feedbacks do not. To some extent the solutions to this important problem can be solved by an orthogonal complement component analysis [55].

Based on the semantic categorization and region-based similarity measure a synthetic aperture radar [56] image retrieval method is inspired by the existing content-based image retrieval (CBIR) techniques and is oriented toward the Earth observation (EO). First, due to the large sizes of SAR images, new method semantically classifies the land covers in the patch level rather than the pixel level by the classic semi supervised learning (SSL), which could reduce the workload of selecting the representative image patch and decrease the searching space in the similarity calculation component. Furthermore, to overcome the inevitable classification error, synthetic aperture radar (SAR) method provides an error recovery scheme, preventing the errors produced in categorization to contaminate the retrieval results. Third, the similarity between two patches is calculated by the improved integrated region matching (IIRM) measure based on the region-based similarity measure, which fails to meet the expectation in SAR images.

Content-based image retrieval (CBIR) is a search technique based on the similarity of visual features and has demonstrated potential benefits for medical diagnosis, education, and research. However, clinical adoption of CBIR is partially hindered by the difference between the computed image similarity and the user's search

intent, the semantic gap, with the end result that relevant images with outlier features may not be retrieved. Furthermore, most CBIR algorithms do not provide intuitive explanations as to why the retrieved images were considered similar to the query (e.g., which subsets of features were similar), hence, it is difficult for users to verify if relevant images, with a small subset of outlier features, were missed. Users, therefore, resort to examining irrelevant images and there are limited opportunities to discover these "missed" images. A guided visual exploration of the search space through a tool, called visual analytics for medical image retrieval [60] facilitates interactive exploration of the complete database using the query image as the point of reference.

2.3 Research Gap

Literature presented proves that the content based image retrieval using shape and HOG descriptor is a complex and limited work carried out till date. Content-based image retrieval is a critical process. Use of shape descriptor based on templates is preferred compare to existing conventional techniques. HOG (histogram of gradients) approach is beneficial when compared to existing approaches. The proposed scheme employs canny edge algorithm to identify strong key points on the edges, where the generalized distance transform scheme has used for key point set formation.

To the best of our knowledge there exists very few architecture till date that incorporates shape and HOG descriptor to reduce the gap between high-level and low-level features. Shape descriptor provides an object shape identification based on key points using templates. HOG features provides low-level feature to identify actual object. Therefore, both the shape and HOG features together can provide a complete identification and structure of the object.

In the literature survey, several existing techniques are described for shape and HOG descriptor [9] , [11] , [13-15] , [20,21] , [36] , [51,52] for content-based image retrieval. However, these techniques are much concerned about either shape identification or shape structure improvement but in the proposed model integration of shape and HOG descriptor enables both image identification and structure.