A Literature Survey of Image Descriptors in Content Based Image Retrieval

Abdulrehman Ahmed Mohamed[1], Dr. Cyrus Abanti Makori PhD[2], & John Kamau[3].
Mount Kenya University, School of Computing and Informatics, Department of Information Technology, P. O. Box 342-01000 Thika, Kenya
[1]almutwafy@gmail.com, [2]cmakori@mku.ac.ke, and [3]jkamau@mku.ac.ke

Abstract - As a result of the new communication technologies and the massive use of Internet in the society, the amount of audio-visual information available in digital format is increasing considerably. This has necessitated designing systems that allow describing the content of several types of multimedia information in order to search and classify them. The audio-visual descriptors are used for contents description in Content Based Image Retrieval (CBIR). The CBIR is a technique for retrieving images on the basis of automatically-derived features such as color, texture and shape. It uses digital processing and analysis to automatically generate descriptions directly from the media data. Image descriptors are the descriptions of the visual features of the contents in images, which describe elementary characteristics of images such as the shape, the color, the texture or the motion, among others. A literature survey study is most important for understanding and gaining insight about specific area of a subject. Therefore, in this paper we survey some of the state-of-art technical aspects of image descriptors in CBIR. Even though lots of research works had been published on CBIR, however, in this paper an effort has been made to explore an in-depth chronological growth in this field of image descriptors with respect to performance measure metrics of CBIR systems.

Index Terms – Content Based Image Retrieval (CBIR), Image Descriptors, and Performance Measure Metrics

1 INTRODUCTION

1.1 Background Study

Even though Multimedia databases (MMD) is among the fastest growing emerging technologies in the field of database systems. New technologies pose numerous challenges, and MMD has its share of challenges. Most of MMD challenges are around Content-based Image Retrieval (CBIR) systems. CBIR is a technique for retrieving images on the basis of automatically-derived features such as color, texture and shape. Moreover, multimedia objects contain encoding of raw sensorial data, which compromise the efficient indexing and retrieval. As result of which, Query by Image Content (QBIC) technique using image descriptors for indexing and retrieval of multimedia objects were proposed by various studies to address this problem. However, an effective and precise performance evaluation benchmarking for this technique remains elusive.

1.2 Technological Trends

Since the invent of the Internet, and the availability of image capturing devices such as smart phones, digital cameras, image scanners and geospatial satellite devices, the size of digital image storage is increasing rapidly. Efficient image searching, browsing and retrieval tools are required by end users from various domains, including remote sensing, fashion design, criminology, publishing, medicine, architecture, etc. It is for this reasons that, many general purpose image retrieval systems have been developed. Therefore, for the same reasons we explore the in-depth survey of content based image retrieval technology, descriptors technology and performance measure framework technology in order to gain an insight of this domain field.

1.3 Content Based Image Retrieval (CBIR) Technology

The main object of a Content-Based Image Retrieval (CBIR) system, also known as Query by Image Content (QBIC), is to help users to retrieve relevant images based on their contents. CBIR technologies provide a method to find images in large databases by using unique descriptors from a trained image. The image descriptors include texture, color, intensity and shape of the object inside an image. The urgency of efficient image searching, browsing and retrieval techniques by users from large repositories such as the internet, metrological images and geospatial images is real.

It is reported by [5] that, there are two retrieval frameworks: text-based and content-based. In the text-based approach, the images are manually annotated by text descriptors, which are then used by a database management system to perform image retrieval. There are two disadvantages with this approach. The first is that a human labor at considerable level is required for manual annotation. The second is the inaccuracy in annotation due

to the subjectivity of human perception. To overcome these disadvantages in text-based retrieval system, content- based image retrieval (CBIR) was introduced.

It is asserted by [24], that content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR), is the application of computer vision techniques to the image retrieval problem. It is a technique which uses visual features of image such as color, shape, texture, etc. to search user required image from large image database according to user's requests in the form of a query image. Images are retrieved on the basis of similarity in features where features of the query specification are compared with features from the image database to determine which images match similarly with given features.

According to [11] the CBIR paradigm has three fundamental basses of; visual features extraction, multidimensional indexing, and retrieval system design. The visual features (content) extraction is the basis of CBIR. In broad sense, features may include both text-based features (keyword, annotation) and visual features (color, text, shape, faces). The visual feature can be further classified as; general features, and domain specific features. The former include color, texture, and shape feature. While the latter is application-dependent and may include, for example, human faces and finger prints (pattern recognition). The figure 1 below summaries the image retrieval system architectural

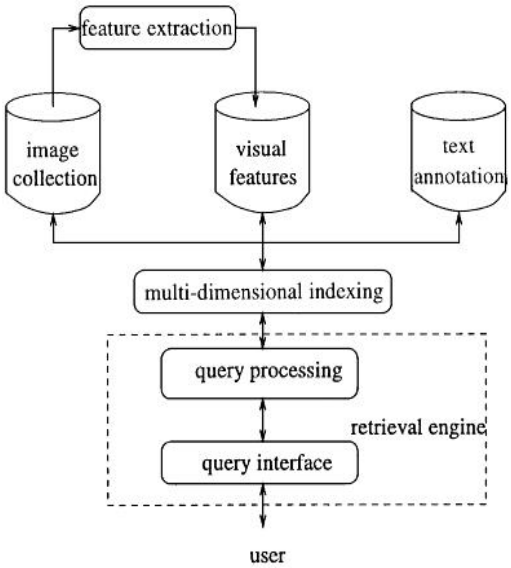

Figure 1 : An image retrieval system architecture, source: [23].

1.4 Image Descriptors Technology

As a result of the new communication technologies and the massive use of Internet in the society, the amount of audio-visual information available in digital format is increasing considerably. This has necessitated designing systems that allow describing the content of several types of multimedia information in order to search and classify them. The audio-visual descriptors are in charge of the contents description.

It is defined by [18] that, in computer vision, visual descriptors or image descriptors are defined as the descriptions of the visual features of the contents in images, videos, or algorithms or applications that produce such descriptions. They describe elementary characteristics such as the shape, the color, the texture or the motion, among others.

It is describe by [28], that visual descriptors are divided in two main groups: General information descriptors, which they contain low level descriptors which give a description about color, shape, regions, textures and motion, and specific domain information descriptors which they give information about objects and events in the scene.

In their book [6] describe the general information descriptors as consisting of a set of descriptors that covers different basic and elementary features like: color, texture, shape, motion, location and others. The color descriptor is the most basic quality of visual content. Five tools are defined to describe color; Dominant Color Descriptor (DCD), Scalable Color Descriptor (SCD), Color Structure Descriptor (CSD), Color Layout Descriptor (CLD), and Group of frame (GoF) or Group-of-pictures (GoP). The Texture descriptors are used to characterize image, textures, or regions. They observe the region homogeneity and the histograms of these region borders. The set of descriptors is formed by: Homogeneous Texture Descriptor (HTD), Texture Browsing Descriptor (TBD), and Edge Histogram Descriptor (EHD). The Shape descriptor contains important semantic information due to human's ability to recognize objects through their shape.

However, this information can only be extracted by means of a segmentation similar to the one that the human visual system implements. These descriptors describe regions, contours and shapes for 2D images and for 3D volumes. The shape descriptors are formed by; Region-based Shape Descriptor (RSD), Contour-based Shape Descriptor (CSD) and 3-D Shape Descriptor (3-D SD). While, the Motion descriptors are defined by four different descriptors which describe motion in video sequence. The descriptor set is formed by; Motion Activity Descriptor (MAD), Camera Motion Descriptor (CMD), Motion Trajectory Descriptor (MTD), and Warping and Parametric Motion Descriptor (WMD and PMD). Finally, the Location descriptor elements location in the image is used to describe elements in the spatial domain. In addition,

elements can also be located in the temporal domain. The location descriptors are formed by; Region Locator Descriptor (RLD) and Spatio Temporal Locator Descriptor (STLD).

1.5 Performance Measure Framework Technology

It is asserted by [15], that evaluation has always been an important aspect of information retrieval. Most studies follow the Cranfield paradigm, using a set of ad-hoc queries for evaluation and calculating effectiveness measures such as precision and recall. While the Cranfield paradigm has often been criticized, it is not without merit and is used in large evaluation initiatives such as TREC and CLEF, which were designed to evaluate search results. Most search engine evaluations today are "TREC-style," as they follow the approach used in these tests. They use, however, a somewhat limited understanding of a user's behavior as their results are determinant upon selection behavior, which is influenced by many factors. However, TREC-style evaluations focus on a "dedicated searcher," i.e., someone who is willing to examine every result given by the search engine and follow the exact order in which the results are presented

The performance evaluation of the CBIR systems based on realistic user criteria is nearly an unexplored area in information retrieval. For CBIR algorithms, there are no standard test collections or evaluation frameworks available like TREC in the text retrieval domain

In his book [4] proposed a framework for evaluating the retrieval effectiveness of search engines. The framework consists of five parts, namely queries selection, results collection, results weighting, results judgment, and data analysis. While different choices regarding the individual stages of the test design are made for different tests, guidance for designing such tests is given. In the section pertaining to query selection, the kinds of queries that should be utilized when evaluating search engines are discussed. In the section on results collection, the information collected in addition to the URLS of the results are detailed. The result weighting section deals with the positions and the higher visibility of certain results, due to an emphasized presentation. The section on results judgment addresses who should make relevance judgments, what scales should be used, and how click-through data can support retrieval effectiveness tests. In the last portion of the framework (data analysis), appropriate measures that go beyond the traditional metrics of recall and precision are discussed. The framework, including all of the elements described below, is depicted in Figure 2

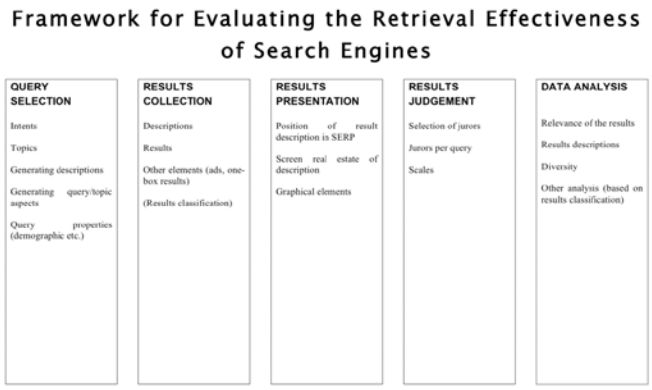

Figure 2 : Framework for the evaluation of effectiveness of search engines, source: [4]

The text-based retrieval techniques are based on manually assigning descriptors such as caption, index term, cataloguing of data among others. While in CBIR employs indexing based on automatic identification and abstraction of indexable visual features within an image using image-processing transformations of low level visual abstraction features such as color, shape and texture, where conventional object recognition techniques cannot recognize these features. A query is typically made by an example image (e.g. photo, drawing, and sketches) and applying partial-match methods to rank retrieved image into calculated similarity order, [19].

The performance evaluation of the CBIR systems based on realistic user criteria is nearly an unexplored area in information retrieval. For CBIR algorithms, there are no standard test collections or evaluation frameworks available like TREC in the text retrieval domain

2 THEORETICAL LITERATURE REVIEW

The start-of-art technologies of acquisition, transmission, storage, query and retrieval of multimedia objects allowed the accumulation of large collections of images repository to grow rapidly. With the increase in popularity of the internet, private networks and development of multimedia technologies, users are not satisfied with the traditional information retrieval techniques. So nowadays, the content based image retrieval is becoming a source of exact and fast retrieval. Content Based Image Retrieval (CBIR) is a technique which uses visual features to search user required images from large image database according to user's requests in the form of a query image. Images are retrieved on the basis of similarity in features where features of the query specification are compared with features from the image database to

determine which images match similarly with given features.

Therefore, to gain the insight of these technologies, this paper presents an in-depth desk literature survey study of the three main technologies of: Content-Based Image Retrieval including feature extraction, segmentation, content levels, query level, similarity levels and image taxonomy; Descriptors including taxonomy of descriptors and visual descriptors; and Performance Frameworks including Performance and Correction metrics including recall, fallout, F-Measures, R-Precision, and Mean-Average Precision.

2.1 Content Based Image Retrieval (CBIR)

The term Content-based image retrieval was coined in 1992 by T. Kato to describe experiments into automatic retrieval of images from a database, based on the colors and shapes. Since then, this term has been used to describe the process of retrieving desired images from a large collection on the basis of syntactical image features. The technique used three main technologies of: pattern recognition, signal processing, and computer vision, [6].

In content-based image retrieval (CBIR), the image databases are indexed with descriptors derived from the visual content of the images. Most of the CBIR systems are concerned with approximate queries where the aim is to find images visually similar to a specified target image. In most cases the aim of CBIR systems is to replicate human perception of image similarity as well as possible, [27].

2.1.1 Content Based Image Retrieval Process

The process of CBIR consists of the following six main stages of: image acquisition, image preprocessing, feature extraction, similarity matching, resultant retrieval image and user interface and feedback

2.1.1.1 Image acquisition

It is the process of acquiring a digital image from the image database. The image database consists of the collection of n number of images depends on the user range and choice.

2.1.1.2 Image preprocessing

It is the process of improving the image in ways that increases the chances for success of the other processes. The image is first processed in order to extract the features, which describe its contents. The processing involves filtering, normalization, segmentation, and object identification. Image segmentation is the process of dividing an image into multiple parts. The output of this stage is a set of significant regions and objects, [27].

2.1.1.3 Feature Extraction

It is the process where features such as shape, texture, color, etc. are used to describe the content of the image. The features further can be classified as low-level and high-level features. In this stage visual information is extracts from the image and saves them as features vectors in a features database .For each pixel, the image description is found in the form of feature value (or a set of value called a feature vector) by using the feature extraction .These feature vectors are used to compare the query with the other images and retrieval, [12].

2.1.1.4 Similarity Matching

It is a process that entails the information about each image is stored in its feature vectors for computation process and these feature vectors are matched with the feature vectors of query image (the image to be search in the image database whether the same image is present or not or how many are similar kind images are exist or not) which helps in measuring the similarity. This step involves the matching of the above stated features to yield a result that is visually similar with the use of similarity measure method called as Distance method. There are various distances methods available such as Euclidean distance, City Block Distance, and Canberra Distance, [5].

2.1.1.5 Resultant Retrieved images

It is the process that searches the previously maintained information to find the matched images from database. The output will be the similar images having same or very closest features as that of the query image, [12].

2.1.1.6 User interface and feedback

It is the process which governs the display of the outcomes, their ranking, the type of user interaction with possibility of refining the search through some automatic or manual preferences scheme etc. The Figure 3 below demonstrates the CBIR System and its various components.

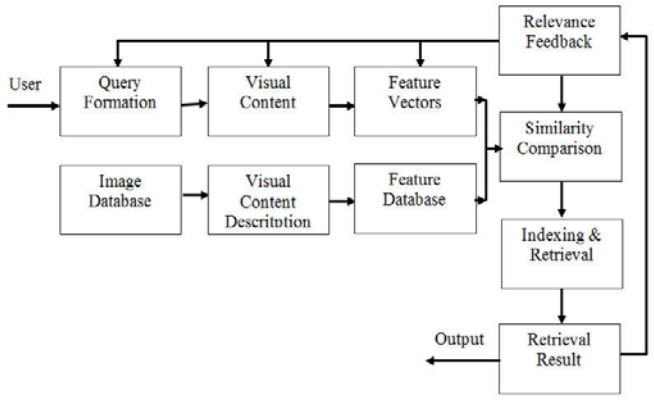

Figure 3: CBIR System and its various components, source: [5].

2.2 Feature Extraction

Feature extraction is the basis of content based image retrieval. Typically two types of visual feature in CBIR: primitive features which include color, texture and shape and domain specific which are application specific and may include, for example human faces and finger prints.

2.2.1 Color

Color represents one of the most widely used visual features in CBIR systems. First a color space is used to represent color images. The RGB space is where the gray level intensity is represented as the sum of red, green and blue gray level intensities. In image retrieval a histogram is employed to represent the distribution of colors in image. The number of bins of histogram determines the color quantization. Therefore the histogram shows the number of pixels whose gray level fails within the range indicated by corresponding bin. The comparison between query image and image in database is accomplished through the use of some metric which determines the distance or similarity between the two histograms. Besides the color histogram several other color features representation like color moments and color sets have been applied [8].

2.2.2 Shapes

In image retrieval, depending on the applications, some require the shape representation to be invariant to translation, rotation and scaling, whiles others do not. In general shape representation can be divided into two categories of; boundary-based which use only the outer boundary of the shape and region-based which uses the entire shape regions. The most successful representative for these two categories are Fourier descriptors and Moment invariants. The main idea of a Fourier descriptor is to use the Fourier transformed boundary as the shape feature. Rui et al. proposed a modified Fourier descriptor which is robust to noise and invariant to geometric transformation [11].

2.2.3 Texture

Texture refers to the visual patterns that have property of homogeneity or arrangement that do not result from the presence of only a single color or intensity. Various texture representations have been investigated in both pattern recognition and computer vision. Haralick proposed the co- occurrence matrix representation of texture feature. This approach explored the gray level spatial dependence of structure. Tamura developed computational approximation to the visual texture properties found to be important in psychology studies. The six visual texture properties were coarseness, contrast, directionality, line likeness, regularity and roughness [5].

2.2.4 Segmentation

According to [24] in computer vision, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Each of the pixels in a region is similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s), [24]. When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like marching cubes.

2.2.5 Content levels

Most researchers accept the assertion that there are multiple levels of content. For example, luminance, and color are regarded as low-level content, and physical objects (such as an automobile or a person) are regarded as high-level content. (Texture and patterns, which blend different types of content, might be regarded as mid-level content.) However, there is no broad agreement about how many levels of content can be perceived by a human, how many types of content there are in each level, or how the

content of a particular image might be classified into types and levels, [3].

2.3 Query levels

It is mentioned by [17] that, there are three levels of queries in CBIR; Level 1: Retrieval by primitive features such as color, texture, shape or the spatial location of image elements. Typical query is query by example, 'find pictures like this'. Level 2: Retrieval of objects of given type identified by derived features, with some degree of logical inference. For example 'find a picture of a flower' and Level 3: Retrieval by abstract attributes, involving a significant amount of high-level reasoning about the purpose of the objects or scenes depicted. This includes retrieval of named events, of pictures with emotional or religious significance, etc. Query example, 'find pictures of a joyful crowd'.

2.4 Similarity Measure

It is asserted by [1] that, the Similarity functions seek to calculate the content difference between two images based on their features. One of the images is given as search parameter and another is stored in the database and had their features previously extracted. There are four major classes of similarity measures: color similarity, texture similarity, shape similarity, and object and relationship similarity.

2.5 Taxonomy of Images

Basically images are broadly classified in four categories as intensity images, indexed images, scaled images, and binary images.

2.5.1 Intensity Images

It represents an image as a matrix where every element has a value corresponding to how bright/dark the pixel at the corresponding position should be colored. There are two ways to represent the number that represents the brightness of the pixel: The double class (or data type). This assigns a floating number ("a number with decimals") between 0 and 1 to each pixel. The other class is called uint8 which assigns an integer between 0 and 255 to represent the brightness of a pixel, [8].

2.5.2 Indexed Images

In an indexed image, the image matrix values do not determine the pixel colors directly. Instead, MATLAB uses the matrix values as indices for looking up colors in the figure's color map. This is a practical way of representing color images. An indexed image stores an image as two matrices. The first matrix has the same size as the image and one number for each pixel. The second matrix is called the color map and its size may be different from the image, [22].

2.5.3 Scaled indexed images

A scaled indexed image uses matrix values. The difference is that the matrix values are linearly scaled to form lookup table indices. To display a matrix as a scaled indexed image, use the MATLAB image display function imagesc, [5].

2.5.4 Binary Images

This image format also stores an image as a matrix but can only color a pixel black or white (and nothing in between). It assigns a 0 for black and a 1 for white, [8].

2.6 Descriptor

Many different techniques for describing local image regions have been developed. The simplest descriptor is a vector of image pixels. The cross-correlation measure can then be used to compute a similarity score between two regions. However, the high dimensionality of such a description increases the computational complexity of recognition. Therefore, this technique is mainly used for finding point-to-point correspondences between two images. The point neighborhood can be sub-sampled to reduce the dimension, [20].

2.6.1 Taxonomy of Descriptors

Basically descriptors are broadly classified into four main categories of: distribution-based descriptor, non-parametric transformations, spatial-frequency techniques, differential techniques, Scale Invariant Feature Transform (SIFT) detector and Visual Descriptors.

2.6.1.1 Distribution-based descriptors

It is a simple descriptor that consists of distribution of the pixel intensities which can be represented by a histogram. A more expressive representation was introduced by Johnson and Hebert in the context of 3D object recognition. Their representation (spin image) is generated using a histogram of the relative position of neighborhood points to the interest point in 3D space, [10].

2.6.1.2 Non-parametric transformations

An approach, interesting for its robustness to illumination changes, was developed by Zabih and Woodfill. It relies on local transforms based on non-parametric statistics, which use the information about ordering and reciprocal relations between the data, rather than the data values themselves. A small region is described by ordered binary relations of the intensities at neighboring points [7].

2.6.1.3 Spatial-frequency techniques

Many techniques describe the frequency content of an image. The Fourier transform decomposes the image content into the basic functions. However, in this representation the spatial relations between points are not explicit and the basic functions are infinite, therefore difficult to adapt to a local approach. The Gabor transform overcomes these problems but a large number of Gabor filters is required to capture small changes in frequency and orientation, that is the description is high dimensional. Gabor filters and wavelets are frequently explored in the context of texture classification, [20].

2.6.1.4 Differential descriptors

A set of image derivatives computed up to a given order approximates a point neighborhood. The properties of local derivatives (local jet) were investigated by Koenderink. While Florack derived differential invariants, which combine components of the local jet to obtain rotation invariance. Freeman and Adelson developed steerable filters, which steer derivatives in a particular direction given the components of the local jet. Steering derivatives in the direction of the gradient makes them invariant to rotation. A stable estimation of the derivatives is obtained by convolution with Gaussian derivatives[10].

2.6.1.5 Scale Invariant Feature Transform (SIFT) detector

It is asserted by [26] that, a SIFT keypoint is a circular image region with an orientation. It is described by a geometric frame of four parameters: the keypoint center coordinates x and y, its scale (the radius of the region), and its orientation (an angle expressed in radians). The SIFT detector uses as keypoints image structures which resemble "blobs". By searching for blobs at multiple scales and positions, the SIFT detector is invariant (or, more accurately, covariant) to translation, rotations, and re scaling of the image. The keypoint orientation is also determined from the local image appearance and is covariant to image rotations. Depending on the symmetry of the keypoint appearance, determining the orientation can be ambiguous. In this case, the SIFT detectors returns a list of up to four possible orientations, constructing up to four frames (differing only by their orientation) for each detected image blob.

2.7 Visual Descriptors

Visual descriptors are divided in two main groups of: General information descriptors: which they contain low level descriptors which give a description about color, shape, regions, and motion, and Specific domain information descriptors: which they give information about objects and events in the scene. A concrete example would be face recognition.

2.7.1 General information descriptors

General information descriptors consist of a set of descriptors that covers different basic and elementary features like: color, texture, shape, motion, location and others. This description is automatically generated by means of signal processing, [7].

2.7.1.1 Color Descriptor

It is the most basic quality of visual content. Five tools are defined to describe color. The three first tools represent the color distribution and the last ones describe the color relation between sequences or group of images: Dominant Color Descriptor (DCD), Scalable Color Descriptor (SCD), Color Structure Descriptor (CSD), Color Layout Descriptor (CLD), and Group of frame (GoF) or Group-of-pictures (GoP), [20].

2.7.1.2 Texture Descriptor

It is also important quality in order to describe an image. The texture descriptors characterize image textures or regions. They observe the region homogeneity and the histograms of these region borders. The set of descriptors is formed by three descriptors of: Homogeneous Texture Descriptor (HTD), Texture Browsing Descriptor (TBD), and Edge Histogram Descriptor (EHD), [10].

2.7.1.3 Shape Descriptor

They contain important semantic information due to human's ability to recognize objects through their shape. However, this information can only be extracted by means of a segmentation similar to the one that the human visual system implements. Nowadays, such a segmentation system is not available yet, however there exists a serial of algorithms which are considered to be a good approximation. These descriptors describe regions, contours and shapes for 2D images and for 3D volumes. The shape descriptors are the defined by three descriptor of: Region-based Shape Descriptor (RSD), Contour-based Shape Descriptor (CSD), and 3-D Shape Descriptor (3-D SD), [18].

2.7.1.4 Motion Descriptors

The motion descriptors are defined by four different descriptors which describe motion in video sequence. Motion is related to the objects motion in the sequence and to the camera motion. This last information is provided by the capture device, whereas the rest is implemented by means of image processing. The descriptor set are

categories as: Motion Activity Descriptor (MAD), Camera Motion Descriptor (CMD), Motion Trajectory Descriptor (MTD), and Warping and Parametric Motion Descriptor (WMD and PMD), [28].

2.7.1.5 Location Descriptor

The elements location in the image is used to describe elements in the spatial domain. In addition, elements can also be located in the temporal domain. The descriptor set are categories as: Region Locator Descriptor (RLD), and Spatio Temporal Locator Descriptor (STLD), [7].

2.7.2 Specific domain information descriptors

These are descriptors, which give information about objects and events in the scene, are not easily extractable, even more when the extraction is to be automatically done. Nevertheless they can be manually processed. Face recognition is a concrete example of an application that tries to automatically obtain this information.

2.7.3 Other techniques

Lowe proposed a descriptor in which a point neighborhood is represented with multiple images. These images are orientation planes representing a number of gradient orientations. Each image contains only the gradients corresponding to one orientation. Each orientation plane is blurred and re-sampled to allow for small shifts in positions of the gradients. This description provides robustness against localization errors and small geometric distortions, [28].

2.8 Performance Framework

It is asserted by [16], that the saliency of an item such as an object, a person, or a pixel, is the state or quality by which it stands out relative to its neighbors. Saliency typically arises from contrasts between items and their neighborhood, such as a red dot surrounded by white dots, a flickering message indicator of an answering machine, or a loud noise in an otherwise quiet environment.

It is suggested by [2] that, instead of depending on the ground truth of eye-tracking data or manually segmented objects, the saliency models ranking can use a Content-based Image Retrieval related criterion as a new evaluation method for the bottom-up attention models. In their proposal the evaluation criterion was based on the ability of a visual attention model to maintain the performance of a CBIR reference method when it is acts as a filter for the key points used by the recognition system. They evaluated the visual attention models basing on Mean Average Precision (MAP) denoted by dMAP = (MAP ORwithout filtering – MAP ORafter filtering)

According to [14] in their article "A novel hybrid method in trademark image retrieval" they, proposed a Fourier-centroid-Histogram Descriptor (FCHD) as a technique to solve trademark image query retrieval, a hybrid region-based and contour-based descriptor retrieval scheme. The method used contour and content information of the images; the binary data of an object from images were collected, and then Fourier-Centroid Descriptor, shape distributions, and their proposed method (FCHD) were computed. To evaluate the performance of the proposed method, the trademark images were chosen from Intellectual Property Office of Taiwan. The database contained 600 images which were divided into circle, triangle, and square three categories for testing the performance of the proposed descriptor, including invariance properties and shape similarity. The metric used to evaluate the similarity between two images of the database was a linear combination of the normalized distances between the individual features. The purpose of the study was to overcome the drawbacks of existing shape representation techniques in order to compare the retrieval efficiency of FCHD with Fourier-Centroid Descriptor (FCD).

It is reported by [21], in their article titled the "Shape distribution" who proposed shape distributions, employing shape functions to describe general shapes. There were five shape functions, such as A3, D1, D2, D3, and D4 shown in Figure 4, which could provide a unique signature of an object by measuring geometric properties of object. Shape distribution functions described the histogram of shapes or surface with metric measurements such as the lengths, angles, areas, and volumes. Importantly, in order to compute a point distance shape distribution function of an object in an image, a representative number of points must be randomly selected. The Euclidean distances were calculated and those measurements were tallied in bins. By the application of shape distribution, the simplification of the description was particularly suitable for natural shapes. However, it was necessary to classify images before retrieving for better effect.

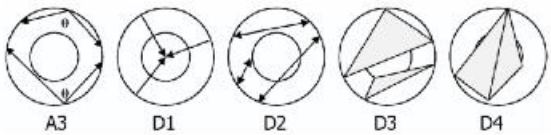

Figure 4: Five simple shape functions based on angles (A3), lengths (D1 and D2), areas (D3), and volumes (D4) – Source: [21].

According to [13], in their article "Content-Based Image Retrieval System: Reviewing and Benchmarking" who proposed the benchmarking of retrieval performance of CBIR systems by comparison and ranking. They presented a measure ranking tool based on the Normalized Average Rank (NAR). The tool allowed comparisons of different CBIR with respect to different queries. The CBIRS were benchmarked by running and evaluating a set of well defined queries on them. They defined a query is an image for the search engines to look for similar images.

The search was done within the images of the query domain defined. The CBIRS were categories according to parameter setting (psets) such as those of texture features, color features, shape features, keywords, interactive relevance feedback, MPEG-7 support, designed for web, classification and image regions. They refer a CBIRS with a specific pset as a system. The psets are named with letters of alphabet starting with a. For example QBIC has 5 psets, a, b, c, d, and e.

The ground truth for the benchmark was the human judge, which was established by random selection of experts with no prior knowledge of CBIR techniques. Each expert was presented with 14 queries to compare the query image with all images from the query domain. The experts were asked to judge similarity from 0 (no similarity) up to 4 (practical equal images). The ground truth consisted of query image and image results, which were mapped to similarity value. The value is the mean of all similarity values from different survey results.

The proposed tool for comparison and ranking is then applied to the result. The aim of the Benchmark is to compare CBIRs to each other based on their accuracy, i.e. their ability to produce results in order to accurately match ideal results defined by a ground truth.

They assume a CBIRs does a good job if it finds images which the user himself would choose. No other queues were used, like e.g. relevance feedback. They used query by example only, no query by keyword, sketch or others. The authors of the study established a relation with developers of CBIRs and scholarly community.

System	reference	texture	color	shape	keywords	interactive relevance feedback	MPEG-7 support	designed for web	classification	image regions
Behold	[YHR06]	*	*		*					*
Caliph&Emir	[LBK03]	*	*		*		*			
Cortina	[QMTM04]	*	*		*	*	*	*	*	
FIRE	[DKN04]	*	*		*	*				*
ImageFinder	[att07]	?	?	?					?	*
LTU ImageSeeker	[LTU07]	*	*	*	?	?			*	*
MUVIS	[GK04]	*	*							
Oracle Intermedia	[oim07]	*	*	*	*[2]					
PictureFinder	[HMH05]	*	*	*	*	*				*
PicSOM	[KLLO00]	*	*	*		*	*			*
QBIC	[FSN+01]	*	*	*	*	*				
Quicklook	[CGS01]	*	*	*	*	*				
RETIN	[CGPF07, FCPF01]	*	*			*				
SIMBA	[Sig02]	*[3]	*							
SIMPLIcity	[LWW00, WLW01]	*	*	*					*	*
SMURF	[VV99]		*	*						
VIPER/GIFT	[Mül02]	*	*			*				*

Figure 5: Live CBIR Systems Source [13].

The Figure 5 gives an overview of all live CBIR systems. For each CBIRS (rows), its features are listed (columns). A star ('*') in a column marks the support for a feature and a question mark ('?') if the support is unknown. If a system doesn't support the feature, the table entry is left empty.

2.9 Performance and correctness measures

Many different measures for evaluating the performance of information retrieval systems have been proposed. The measures require a collection of documents and a query. All common measures described assume a ground truth notion of relevancy: every document is known to be either relevant or non-relevant to a particular query. In practice queries may be ill-posed and there may be different shades of relevancy [29].

2.9.1 Precision

Precision is the fraction of the documents retrieved that are relevant to the user's information need.

$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

In binary classification, precision is analogous to positive predictive value. Precision takes all retrieved documents into account. It can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n or P@n. Note that the meaning and usage of "precision" in the field of Information Retrieval differs from the definition of accuracy and precision within other branches of science and technology.

2.9.2 Recall

Recall is the fraction of the documents that are relevant to the query that are successfully retrieved.

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

In binary classification, recall is often called sensitivity. So it can be looked at as the probability that a relevant document is retrieved by the query. It is trivial to achieve recall of 100% by returning all documents in response to any query. Therefore recall alone is not enough but one needs to measure the number of non-relevant documents also, for example by computing the precision.

2.9.3 Fall-out

The proportion of non-relevant documents that are retrieved, out of all non-relevant documents available:

$$\text{fall-out} = \frac{|\{\text{non-relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{non-relevant documents}\}|}$$

In binary classification, fall-out is closely related to specificity and is equal to $(1 - \text{specificity})$. It can be looked at as the probability that a non-relevant document is retrieved by the query. It is trivial to achieve fall-out of 0% by returning zero documents in response to any query.

2.9.4 F-measure

The weighted harmonic mean of precision and recall, the traditional F-measure or balanced F-score is:

$$F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{(\text{precision} + \text{recall})}.$$

This is also known as the F_1 measure, because recall and precision are evenly weighted. The general formula for non-negative real β is:

$$F_\beta = \frac{(1 + \beta^2) \cdot (\text{precision} \cdot \text{recall})}{(\beta^2 \cdot \text{precision} + \text{recall})}.$$

Two other commonly used F measures are the F_2 measure, which weights recall twice as much as precision, and the $F_{0.5}$ measure, which weights precision twice as much as recall. The F-measure was derived by (Rijsbergen, 1979) so that F_β "measures the effectiveness of retrieval with respect to a user who attaches β times as much importance to recall as precision". It is based on Rigsbergen's effectiveness measure.

$$E = 1 - \frac{1}{\frac{\alpha}{P} + \frac{1-\alpha}{R}}$$

Their relationship is

$$F_\beta = 1 - E \text{ Where } \alpha = \frac{1}{1 + \beta^2}$$

2.9.5 R-Precision

Precision at R-th position in the ranking of results for a query that has R relevant documents. This measure is highly correlated to Average Precision. Also, Precision is equal to Recall at the R-th position.

2.9.6 Mean average precision

Mean average precision for a set of queries is the mean of the average precision scores for each query.

$$\text{MAP} = \frac{\sum_{q=1}^{Q} \text{AveP(q)}}{Q}$$

Where Q is the number of queries

2.9.7 Ground-truth

Ground truth is a term used in remote sensing; it refers to information collected on location. Ground truth allows image data to be related to real features and materials on the ground. The collection of ground-truth data enables calibration of remote-sensing data, and aids in the interpretation and analysis of what is being sensed. Examples include sending technicians to gather data in the field that either complements or disputes airborne remote sensing data collected by aerial photography, satellite side scan radar, or infrared images.

The team of ground truth scientists will be collecting detailed calibrations, measurements, observations, and samples of predetermined sites. From this data, scientists are able to identify land use or cover of the location and compare it to what is shown on the image. They then verify and update existing data and maps [9].

3 RECOMMENDATIONS AND CONCLUSION

It is evidence from the literature survey that various methodologies using difference descriptors were used such as Mean Average Precision (MAP) for performance of a CBIR method which acts as a filter for the key points used in the system recognition; Fourier-centroid-Histogram Descriptor (FCHD), a technique to solve trademark image query retrieval; Shape Distribution Functions, describing the histogram of shapes with metric measurements such as the lengths, angles, areas, and volumes and; Benchmarking of retrieval performance by comparison and ranking of CBIR systems. However, an effective and precise performance evaluation benchmarking techniques remains

elusive. This creates a gap, where the study proposes a performance framework measure using descriptors such as color, textural and shape in Query by Image Content to close this gap for further research works.

REFERENCE

[1] Akbarpour, S. (2013). A Review on Content Based Image Retrieval in Medical Diagnosis. *Technical and Physical Problems of Engineering*, *5*(15), 148–153.

[2] Awad, D., Mancas, M., Richie, N., Courboulay, V., & Revel, A. (2015). A CBIR-BASED EVALUATION FRAMEWORK FOR VISUAL ATTENTION MODELS.

[3] Black Jr., J. A., Fahmy, G., & Panchanathan, S. (2002). A method for evaluating the performance of CBIR systems. *Arizona State University*.

[4] Christophe, J. (2012). *Next Generation Search Engine: Advanced Models for Information Retrieval*. Hershey, PA: IGI Global. Retrieved from http://www.igi-global.com/book/next-generation-search-engines/59723

[5] Danish, M., Rawat, R., & Sharma, R. (2013). A Survey: Content Based Image Retrieval Based On Color, Texture, Shape & Neuro Fuzzy. *Int. Journal Of Engineering Research And Application*, *3*(5), 839–844.

[6] Fierro-Radilla, A., Perez-Daniel, K., Nakano-Miyatakea, M., Perez-Meana, H., & Benois-Pineau, J. (2014). An Effective Visual Descriptor Based on Color and Shape Features for Image Retrieval. In A. Gelbukh, F. C. Espinoza, & S. N. Galicia-Haro (Eds.), *Human-Inspired Computing and Its Applications: 13th Mexican International Conference on Artificial Intelligence, MICAI 2014, Tuxtla Gutiérrez, Mexico, November 16-22, 2014. Proceedings, Part I* (pp. 336–348). Cham: Springer International Publishing. Retrieved from http://dx.doi.org/10.1007/978-3-319-13647-9_31

[7] Gibbon, D. C., Huang, Q., Liu, Z., Rosenberg, A. E., & Shahraray, B. (2012, March). System and method for automated multimedia content indexing and retrieval.

[8] Gonzalez, R. C., & Woods, R. . (2010). *Digital Image Processing* (3rd Ed). India: Prentice-Hall.

[9] Ground truth. (2013, October 22). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Ground_truth&oldid=578297670

[10] Hu, R., & Collommosse, J. (2013). A Performance Evaluation of Gradient Field HOG Descriptor for Sketch Based Image Retrieval.

[11] Jaswal, G., & Kaul, A. (2009). Content Based Image Retrieval – A Literature Review. In *Communication and Control*. India: National Institute of Technology.

[12] Kekre, H. . (2011). Survey of CBIR Techniques and Semantics. *International Journal of Engineering Science and Technology (IJEST*, *3*(5).

[13] Kosch, H., & Maier, P. (n.d.). Content-Based Image Retrieval Systems - Reviewing and Benchmarking.

[14] Lee, M.-T., P. Wu, H.-H., & Yu, Y.-H. (2010). A novel hybrid method in trademark image retrieval.pdf. *Journal of Statistics and Management Systems*, *13*(5), 1029–1044. http://doi.org/10.1080/09720510.2010.10701518

[15] Lewandowski, D. (2012). A Framework for Evaluating the Retriev al Effectivenes s of Se a r c h Eng ines. *IGI Global*.

[16] Li, J., & Levine, M. (2012). Visual Saliency Based on Scale-Space Analysis in the Frequency Domain. *IEE Trans Pattern*, *35*(4).

[17] Liu, Y. (2009). Asurvey of content-based image retrieval with high-level semantics. *Elsevier Science Direcct*.

[18] Manjunath, B. ., Salembier, P., & Sikora, T. (2002). Visual Descriptors. *Wiley and Sons*.

[19] Markkula, M., & Tico, M. (2001). A Test Collection for the Evaluation of CBIR Algorithm: A User and Task-Based Approach. *Information Retrieval*.

[20] Mikolajezk, K., & Schmid, C. (2010). A performance evaluation of local descriptors.

[21] Osada, R., Funkhouser, T., Chazelle, B., & Dobkin, D. (2002). Shape Distributions. *ACM Transactons on Graphics*, *21*(4), 807–832.

[22] Palm, W. (2015). Courseware based on MATLAB and Simulink. Retrieved March 6, 2016, from http://www.mathworks.com/academia/courseware/?s_tid=acmain_ed-pop-cw_gw_bod

[23] Rui, Y., Huang, T. S., & Chang, S.-F. (1999). Image Retrieval: Current Techniques, Promising, Directions, and Open Issues. *Journal of Visual Communicatio and Image Representation*, *10*, 39–62

[24] Stockman, G. C., & Shapiro, L. G. (2001). Compter Vision (2001st ed., pp. 279–325). New Jersey: Prentice-Hall.

[25] Subitha, S., & Suhatha, S. (2013). Survey paper on various methods in content based inforfmation retrieval. *IMPACT: International Journal of Research in Engineering and Technology*, *1*(3), 109–120.

[26] Vedaidi, A. (2013). Scale Invariant Feature Transform (SIFT). *VLFea*.

[27] Viitaniemi, V. (2002). *Image Segmentation in Content-Based Image Retrieval* (Master Thesis). Helsini University OF Technology, Department of Electrical and Communications Engineering, Finland.

[28] Wu, J., & M, J. (2010). CENTRIST: A Visual Descriptor for Scene Categorization. *IEEE Transactions on Pattern Analysis and Ma*, *33*(8), 1489 – 1501.

[29] Zhu, M. (2004). Recall, Precision and Average Precision.