Journal of Engineering and Sustainable Development

Vol. 23, No.03, May 2019
ISSN 2520-0917
https://doi.org/10.31272/jeasd.23.3.4

Review Article

CONTENT-BASED IMAGE RETRIEVAL: SURVEY

Dr. Hanan Ahmed Al-Jubouri

Lecturer, Computer Engineering Department, Mustansiriyah University, Baghdad, Iraq.

Abstract: Extensive use of digital photographic devices has resulted in large volumes of digital images being acquired and stored in databases. Whether it is for scientific research, medical or social networking, there is a growing demand for effective retrieval of digital images based on their visual content (e.g. colour and texture). Content-Based Image Retrieval systems are developed to meet this demand. However, searching for similar and relevant images from large-scale databases still poses a challenge for Content-Based Image Retrieval systems due to the gap between high-level meaning and low-level visual features. This paper reviews different Content-Based Image Retrieval approaches such as Clustering, Region-of-Interest, Bag-of-Visual-Words, Relevance Feedback, Browsing, and indexing that have been developed to reduce such "Semantic gap" issue. So, the interested researchers can interest to determine which method is benefit to his work.

Keywords: *Bag-of-Visual-Words (BOVW), Browsing, Clustering, Content-Based Image Retrieval (CBIR), Relevance Feedback (RF), Region-of- Interest (ROI), Indexing, and Semantic gap*

استرجاع الصورة بالأعتماد على المحتوى: دراسة بحثية

الخلاصة: الاستعمالات الكثيرة للأجهزة الفوتغرافية الرقمية انتجت كم هائل من الصور الرقمية المكتسبة والمخزونة داخل قواعد البيانات. ان كانت للبحث العلمي ، الطبي، او التواصل الاجتماعي، هنالك طلب متزايد لأسترجاع الصور الرقمية بشكل دقيق بالأعتماد على المحتوى البصري مثل اللون والنسجة. لذلك طورت انظمة استرجاع الصورة بالأعتماد على المحتوى البصري لتلبي هذا الطلب. الا انه البحث عن الصور في قواعد البياتات الكبيرة ما زال يمثل صعوبة امام هكذا انظمة من خلال الفجوة مابين المستوى العالي لمعنى الصورة وهو ما يفهمه البشر وامستوى الواطئ المتمثل بالخصائص الصورية المفهوم لدى او الذي يتعامل معه الجهاز الرقمي. لهذا السبب ورقة البحث هذه تقدم قراءة لمختلف الطرق المبحوثة في مجال استرجاع الصور بالأعتماد على المحتوى الصوري مثل طرق التجميع، المنطقة المنشودة، حقيبة الكلمات الصورية، الاسترجاع لتحديث النتائج، التصفح، والفهرسة. طورت هذه الطرق للتقليص من مشكلة الفجوة المعنوية. وذلك سيمكن الباحثيين من الاستفادة في تحديد الطريقة المناسبة لعملهم.

1. Introduction

Previously, image retrieval system used a supervised learning technique with external labelling and annotations which is infeasible and restricted due to requiring manual annotation of images and unavailability respectively. Meanwhile, CBIR system is an unsupervised learning technique which is more flexible;

Corresponding Author: hananaljubouri@uomustansiriyah.edu.iq images are indexed in the database by extracting features such as colour, texture, and/or shape that reflect the visual content of the images as a vector. Upon request, the system extracts a feature vector from a query image in the same way and compares it with the feature vectors of the images in the database using a similarity measurement. The most similar images are returned to the user as a ranked list as shown in Fig 1. However, CBIR systems face a challenge so-called the *semantic gap* which is a bridge between the high-level meaning of the image and its low-level visual features.

Consequently, researchers from different communities interested CBIR and have been produced different approaches to reduce the challenge and the door is still open for more because image representation as low-level feature is a subjective to the human perceive in addition to the shortcoming of similarity measurements. Therefore, this paper reviews developed algorithms and systems in terms of CBIR methods. Section 2 will explain extracted feature levels. Section 3 will clarify similarity measures. Section 4 will review works from literature related to CBIR approaches. Finally, conclusions will be in Section 5.

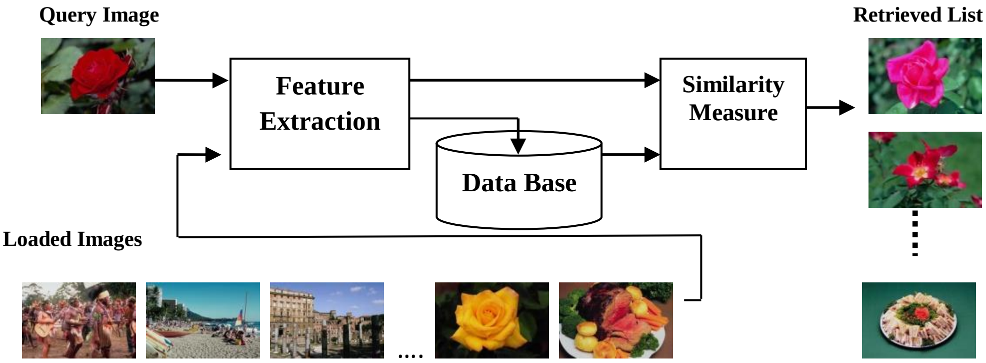

Figure 1: CBIR system.

2. Extraction Features

Feature extraction is the first step of the CBIR process to obtain features that represents an image. Global and local are basically two types of features and can be represented as vectors. For example, the global feature for image Q is a single vector can be represented as $\vec{V}_Q = (v_1, v_2, \dots, v_d)$, where d is the dimension of the vector.

Meanwhile, the local feature is set of vectors can be represented as $S_Q = \{\vec{V}_1, \vec{V}_2, \dots, \vec{V}_n\}$, where $\vec{V}_i = (v_{i1}, v_{i2}, \dots, v_{id})$ $(1 \leq i \leq n)$. More specifically, the image is divided into blocks and then local feature vectors are extracted from each block to represent the image. Generally, features are categories into three levels according to the conceptual meaning as will be clarified in the following [30]:

2.1. Low-Level Features

Colour and texture represent fundamental of low-level features. On one hand, the colour of an image is a function f (x, y), where f is the intensity value of the image I at the location (x, y) in a 3-dimension colour space $\mathbb{R}^3$. Therefore, many colour spaces are exist according to application requirements such as *CMY* (cyan, magenta, yellow), *HSV* (hue, saturation, value), *CIE* (Commission Internationale de l'Éclairage) (either $L^*a^*b^*$ or $L^*u^*v^*$), and *YCbCr* (luminance, chrominance-blue, chrominance-red), etc. All colour spaces based on the primitive *RGB* (red, green, blue) space. On the other hand, the texture is a measure of structure such as smoothness, roughness, or orientation of regions in the image. Structural and statistical are two categories of texture where morphology operator and transformation could be used [29, 30].

2.2. Mid-Level Features

Shape feature is a measurement of geometric attributes of an object in the image or cluster shape which is resulted from grouping low-level features by using a clustering method and also known as segmentation [29, 30].

2.3. High-Level Features

Keywords or phrases represent the semantics of the image and can be used as the high-level features. Many challenges are faced because the image could represent different semantics at the same time or may be subjectively interpreted [29, 30].

3. Similarity Measures

CBIR systems retrieve a list of the most similar images based on a similarity measure between a query image and data base images. This means the similarity measure is calculated to reflect the closeness between two represented images. Therefore, there are variety of the measurements depend on the image representation, space, completeness, outcome of measurement, domain knowledge etc. many factors affect these measures such as noise, semantic, and computational efficiency [1]. A distance function is a dissimilarity measure which is commonly used to calculate the difference between two images. In [2] an overview of matching measurements mostly used was presented. The matching could be two single vector such as City-block and Euclidean (D_{L1} and D_{L2} -norm) as (1) and (2) or two sets of vectors such as Integrated Region Matching (IRM) (3).

$$D_{L_1}(I,J) = \sum_{k=1}^{n} |i_k - j_k| \qquad (1)$$

$$D_{L_2}(I,J) = \left(\sum_{k=1}^{n} (i_k - j_k)^2\right)^{\frac{1}{2}} \qquad (2)$$

where I= (i_1, i_2, …, i_n) and J= (j_1, j_2,…, j_n) are two single vector.

$$D_{IRM}(I^1,J^2) = \sum_{k=1}^{n}\sum_{z=1}^{m} S_{k,z}\, D(f_k^1, f_z^2) \qquad (3)$$

where I= $\{(f_1^1, w_1^1), (f_2^1, w_2^1), \dots, (f_n^1, w_n^1\ )\}$, J= $\{(f_1^2, w_1^2), (f_2^2, w_2^2), \dots, (f_m^2, w_m^2\ )\}$, and weight matrix S(nxm): $\sum_z S_{k,z} = w_k^1, \sum_k S_{k,z} = w_z^2$

4. CBIR Evaluation

Image retrieval and classification techniques are two types of investigation to evaluate developed algorithms or approaches in CBIR. Image retrieval measures the accuracy of *Top m* retrieved images based on similarity values in ascending order which is commonly termed as Average Precision (AP) and mean of average precisions as MAP. There are other measurements such that Precision vs. Recall-graph (PR-graph), rank of the first retrieved image ($Rank_1$), and normalized average rank of relevant images ($\widetilde{Rank}$) [30].

$$AP = \frac{TM}{TN} \qquad (4)$$

where TM is the number of relevant retrieved images and TN is total number of retrieved images.

$$\widetilde{Rank} = \frac{1}{NN_R}\left(\sum_{i=1}^{N_R} R_i - \frac{N_R(N_R-1)}{2}\right) \qquad (5)$$

where N is a database size, N_R the number of relevant images, and R_i is the rank at which the *i*th relevant image is retrieved.

On the other hand, image classification measures the accuracy of classification for an example query image to one image classes based on class labels that are defined previously and associated with images in the database. Hence, the classification is supervised whereas retrieval image is unsupervised technique. Support Vector Machine (SVC) and *k*-Nearest Neighbors (*k*-NN) are two examples of classification methods. Recall measurement can be used in classification the image *I* [30].

$$Recall(I_c) = \frac{N_{CI}}{T_{Im}} * 100 \ (6)$$

where N_{CI} is the number of correct images classification, T_{Im} is total number of images of class I_c.

5. CBIR Approaches

Over the last two decades, many approaches and algorithms in CBIR have been proposed. Clustering, Region of Interest (ROI), Relevance Feedback (RF), Browsing, Bag of Visual Words (BOVW), and indexing are mainly developed among these approaches and regarded as fields of research in CBIR. Feature extraction and similarity measures are principle to all these approaches. This paper reviews different CBIR approaches in terms of features, similarity measures, outstanding issues and trends.

5.1. Clustering Approach

Clustering in terms of CBIR is grouping features into clusters based on similarity function. Many factors affect the effectiveness of process, algorithmic considerations, data and cluster characteristics [3]. Therefore, different clustering algorithms have been developed and categorised accordingly: *Prototype-based*, *model-based*, *density-based*, and *graph-based*. *K*-means, Expectation-Maximization Gaussian Mixture Model (EM/GMM), mean shift, and normalized laplacian spectral algorithms are respectively common examples. Study and details of the four algorithms can be found in [30].

One of the known CBIR system is SIMPLcity which developed by Wang *et al.* [4], where CIE $L^*u^*v^*$ colour space conversion was used as pre-processing a local average colour feature of (4 x 4) block was calculated for each channel. Meanwhile, texture features were three second order moments of wavelet coefficients in HL, LH, and HH high frequency sub-bands of L^* channel. The *K-means* algorithm was used to cluster above local extracted features (i.e. F_1-F_6). In addition, shape feature was calculated using Normalized Intertia (4) as F_7, F_8, and F_9. Experiments of image retrieval were conducted on 200,000 images of the COREL database.

$$NI(Rg_i, \gamma) = \frac{\sum_{x \in Rg_i} \|x - m_i\|^{\gamma}}{P_i^{1+\frac{\gamma}{2}}} \qquad (4)$$

where Rg_i is a region in the image, m_i is the mean of Rg_i, P_i is the number of pixels in the region Rg_i, and $\gamma = 1,2, and$ 3

In [5], *YCbCr* colour space conversion was the pre-processing and then image is divided into 8 x 8 blocks. Discrete Cosine Transform (DCT) was applied on each block for *Y*, *Cb* and *Cr* channels respectively. Local DCT colour and texture feature vector From (12D) was extracted, ($C_Y(0,0)/8$, $C_{Cb}(0,0)/8$, $C_{Cr}(0,0)/8$, $C_Y(0,1)$, $C_Y(1,0)$,$C_Y(1,1)$, $std(B_Y4)$, $std(B_Y5)$, ..., $std(B_Y9)$). The *k*-means clustering algorithm was used to summarize resulted feature vectors by fixing the number of clusters K to ten and selecting only five large clusters to index the image into the database. Experiment was conducted on eight classes from Corel database images. The method was evaluated by classification (*k*-NN) technique using chi-square distance function as dissimilarity measure. The average accuracy of classification was 88%. The method in [6] used the *k*-means algorithm to cluster 7D feature vectors (SQFD): colour from L^*, a^*, and b^* components, coordinates *x* and *y*, contrast *X*, and entropy *ε* value. As a result, an image feature composed of centroids C_i. In addition, each cluster was weighted by $w_i = \frac{|C_i|}{\sum_i |C_i|}$. Also, five global features from MPEG-7 standard were investigated (Scalable Colour (SC), Colour Structure (CS), Colour Layout (CL), Edge Histogram (EH), and Region Shape (RS)). Retrieval experiments conducted on WANG, ALOI, and TWIC databases. Different combinations between above features were tried and effected positively on accuracy of retrieval with respect database images. For example, integrating all MPEG-7 and SQFD features increased the accuracy of retrieval using WANG and TWIC to 59% and 37% respectively. Meanwhile, integrating Sc and SQFD improved the performance using ALOI database to 83%.

Zhang *et al*. [7] presented a feature of 10 dimensional in length which consists of 3 colour R, G, and B, 3 colour deviations of R, G, and B using 5x5 block, 2 gradients of luminance, and 2 locations (x, y). After feature extraction, a finite mixture model was calculated using EM algorithm. Due to the challenge of determine the number of clusters *K* automatically, an adaptive EM algorithm was proposed. The idea is calculating three probabilities, merge, split, and death and making some calculation to select the optimal operation. Proposed a statistical model algorithm was evaluated using Corel and Google image search engine. Table 1 refers to some examples of achieved precisions using Merge (M), Split (S), Merge and Split (MS), and combination MS and Death (MSD). As we can see that combination three operations produces best performance.

Table 1. Precision under different combination of operations for different object classes

Class	S	M	MS	MSD
Sunset	0.82	0.845	0.86	0.88
Sky	0.844	0.826	0.824	0.85
Grass	0.9	0.86	0.89	0.94
Garden	0.744	0.73	0.74	0.75
Beach	0.75	0.755	0.73	0.78
waterway	0.724	0.755	0.745	0.76

Recently, Zeng *at el.* [44] used the EM algorithm to learn GMM in quantizing colours of a histogram termed as spatiogram to represent images. The number of histogram bins was adaptively determined using Bayesian Information Criterion (BIC). The second step was modifying above image representation by replacing color GMM model (i.e. mean vector and covariance matrix) instead of histogram bins. Consequently, similarity measurement was proposed by exploiting Jensen Shannon Divergence (JSD) as third contribution.

Experiments used Corel databases (10, 15, 50, and 100 categories) to evaluate the proposed image representation and similarity matching. Standard spatiogram (i.e. discrete colour histogram bins) and modified spatiogram (i.e. Gaussian colour models) were compared. Results of MAP indicated that the last image representation is more accurate, where 20 query images are randomly selected and 12 images are retrieved (Table 2).

Table 2.Comparision between MAP using standard and modified spatiogram image representation for Corel- (10, 15, 50, and 100) images

Database	Sspatiogram	Mspatiogram
Corel-10	74.50	80.60
Corel-15	63.95	74.14
Corel-50	41.67	51.80
Corel-100	37.64	47.25

5.2. Bag-of-Visual Words (BOVW) Approach

In information retrieval method, words or phrases are used to represent documents as features (bag of vocabulary/words (BOW)). In CBIR, the same idea of BOW is consequently exploited by dividing the image into patches from where visual features are extracted and then quantized by using a clustering algorithm. Resulted clusters correspond to vocabularies and their centroids correspond to the words. Therefore, the method is called Bag of Visual Words (BOVW). In more details, the interested image patches (keypoints) are detected using different detectors such as blob and region, where SIFT (Scale Invariant Feature Transform) and MSER (Maximally Stable Extermal Regions) are robust features examples respectively. Detectors types with their features and equations [8] are surveyed in details.

SIFT descriptor was presented in [9] which invariant to scale and rotation for object recognition. To extract this feature, keypoints locations are firstly selected at maxima and minima of the difference-of-Gaussian function. Keypoints vectors indicate scale, orientation, and location. The best patch to extract these vectors was 4x4 with 8 orientations means the length of SIFT feature is 128D.

Vieux *et al*. [10] implemented BOVW method using SURF descriptor and presented Bag-of-Region (BOR) method to make a comparison. HSV histogram and Local Binary Pattern (LBP) histogram were extracted from image regions to represent colour and texture image visual content respectively. In addition, Vieux presented a incremental clustering algorithm for quantization to compare to *k*-means algorithm. Five different sizes of vocabularies were tested (500, 1000, 2000, 5000, and 10000). Experiment of CBIR was conducted using WANG, SIVAL, and CALTECH101 data bases. Results of image retrieval were measured by MAP and showed that there is no feature outperform than other with all three databases at the same time. In addition, the proposed incremental and *k*-means clustering algorithms were roughly similar. However, the performance of retrieval was the best when the summation between running HSV, LBP, and SURF features as shown in Table 3 for some results.

Table 3. MAP using SURF, HSV, LBP, and SUM features on WANG, SIVAL, and CALTECH101 images

Feature	WANG	SIVAL	CALTECH101
SURF	0.443	0.505	0.177
HSV	0.548	0.443	0.142
LBP	0.533	0.260	0.162
SUM	0.639	0.555	0.197

So far, the image is represented in BOVW method by frequency of visual words according to the visual dictionary. Meanwhile, a method in [11] motivated from visual words representation to visual phrases using n-grams which is a sequence of phrases (i.e. n-visual words). This means the image could be represented by combination of n-visual words. The idea is taking into account the relationship between the visual words instead of regarding them separately. Hence, the spatial information is included in local feature. The n-gram representation is originally used in natural language to process documents, where 1-gram means unigram (one word), 2-gram means bigram (two words) and so on.

Here, the SIFT feature was used and experiments were conducted on Corel, Lung, Medical Image Exams, and Texture databases. City block distance function was calculated to compute the similarity between two normalized histograms (h_Q and h_B) using visual words and two normalized histograms using visual phrases (H_Q and H_B), where Q and B are query and database images. Results of image retrieval (MAPs) showed that 2-gram representation is the best except with the Corel images, where 3-gram representation is outperform both traditional 1-gram and 2-gram as illustrated in Table 4.

Table 4. MAP using (1-3)-gram features on Corel, Lung, Medical Image Exams, and Texture images

Feature	COREL	Lung	Medical Image Exams	Texture
1-gram	0.299	0.365	0.398	0.648
2-gram	0.332	**0.385**	**0.421**	**0.681**
3-gram	**0.340**	0.366	0.412	0.667

In the same motivation, Zeheng and Wang [12] extended the BOVW method to regard the relationship between visual phrases themselves and the proposed approach called *Visual Phraselet*. The SIFT descriptor was extracted from Hussein-affine local regions and different visual words sizes were used (50, 100, 250, 500, 750, and 1000). The method was evaluated on oxford 5K and Paris 6K databases. Then the two databases were extended to Flickr 1M. MAPs of image retrieval showed using soft quantization codebook with the *Visual Phraselet* outperform hard quantization codebook with the *Visual Phraselet* using oxford 5K (Table 5). Meanwhile, the case is opposite using Paris 6K database and extension of the two databases (i.e. +Flickr 1M)

Table 5. MAP using Soft and Hard quantization with *Phraselet* features on oxford 5K, Paris 6K, and extended to Flickr 1M

Feature	oxford 5K	+ Flickr 1M	Paris 6K	+Flickr 1M
Soft+ Phraselet	0.719	0.557	0.528	0.391
Hard+ Phraselet	0.685	0.599	0.564	0.430

Recently, Ren *at el.* [13] extended the traditional BOVW method to Bag-of-Bags of Words (BBOW). First, the SIFT features were extracted and used to build a graph. Then the graph was divided into sub-graphs using Normalize Cuts (NCuts) segmentation algorithm. A histogram was calculated for each sub-graph. Hence, images were represented by K histograms. Following Spatial Pyramid Matching (SPM) and adding some resolutions in partition the graph ($K_r=2^{2r}$), where r is the number of level in pyramid. Therefore, the modification was called Irregular Pyramid Matching (IPM) and the histogram intersection distance function was used to compute the similarity between two images. Experiment of image retrieval applied on Caltech101 database with the 15 categories. Results of MAPs showed that proposed IPM outperform SPM with some classes and versa with others.

More recently, a framework was presented for CBIR [14]. Affine image moment invariants features (AMIs) were extracted from local image regions following the BOVW method, interest keypoints that derived using Speeded Up Robust Features (SURF) blob detector mechanism. The reason behind using SURF is that it returns patch size in addition to the interest keypoints. The AMIs features were fed into *k*-means clustering algorithm at quantization stage in BOVW method. The AMIs (I_1, I_2, I_3, I_6, I_7, I_8, I_9, I_{48}) were calculated for gray-scale images to capture texture information. Meanwhile, the same 8 features were calculated for *R*, *G*, and *B* components and combined to a single value to capture colour information (SetUP1 feature). In addition, the chromaticities of an image were calculated to reduce the effect of the lighting on moment invariants (SetUp2 feature). The third set up feature (SetUp3) is normalization to SetUp2 feature.

To evaluate above three features, experiments were conducted on UCID and UKBench databases using different codebook sizes (32, 128, 512, and 2048). Results of image retrieval were showed that small sizes of codebook (32 and 128) were not sufficient to make moment invariants features discriminate. On the other hand, using 512 and 2048 codebook sizes made a performance of the chromaticities feature significant. In addition, well known BOVW features (SIFT and SURF) were tested for comparison purpose. Results referred to that both features are still rubout.

The performance of the third form of features (setUp3) was higher than other when the number of clusters is 512. However, the performance was not exceeded SIFT and SURF robust features from the literature as illustrated for some results in Table 5.

Table 5. MAP using SURF, SIFT, and SetUP(1-3) features on UCID and UKBench images

Feature	UCID	UKBench
SURF	0.651	0.651
SIFT	0.626	0.604
SetUP3	0.600	0.610
SetUP2	0.537	0.329
SetUP1	0.447	0.279

5.3. Browsing Approach

In CBIR, the most developed approaches present query by example image. Meanwhile, the browsing approach presents a facility to navigate through a large number of database images. The navigation is the core of research with this approach.

Plant and Schaefer [15] report navigation and browsing tools to help users during navigation. Three visualisations are generally categorised, mapping-based, clustering, and graph-based. The aim of the first category is to represent relationships between images and high-dimensional space. The second one aims to cluster similar images into groups based on visual content, metdata, or time to reduce the number of images that required to be visualised on the screen. The third one uses graph, where nodes represent images and edges represent links between similar images. Meanwhile, browsing can be horizontal or vertical. In horizontal browsing, users can panning, zooming, scaling, and magnification while in vertical browsing, the users can navigate through different hierarchal levels.

Hilliges *at el.* [16] presented a browsing system that combines CBIR and zoomable interfaces. First, a user can view groups of images and then can choose the interest group and can decide which images to keep and which images to delete. To cluster images into groups, automatic clustering algorithm was used (X-means) based on colour histogram of chromatic channels of colour spaces. YUV, HSV, and HSL colour spaces were investigated and the achievement of YUV was the best. The interested group of images will be shown as a circle of six regions, where the best images in the centre region and overexposed, underexposed, blurry, blurry- overexposed, blurry-underexposed in the other regions. This quality of clusters measured based on fourteen Haralick texture and four roughness moments features of luminance channel. The best achievement was eleventh Haralick feature to determine sharp/ blurry. Meanwhile, the roughness feature performed well to determine overexposed/ normal/ underexposed. The system was conducted on personal images of few users who evaluated outcomes of the system positively. Therefore, the system needs to be extended to test larger data of images.

Plant and Schaefer [17] developed Honeycomb image browser that visualises a large database of images on a hexagonal lattice with image thumbnails occupying hexagons. Displaying of images based on colour similarity and navigating was in a hierarchal manner. In other words, the median colour similarity was calculated from HSV colour space images to display images as thumbnails in hexagons. A tree data structure was used to provide the hierarchal navigation, where the root image is displayed in a cell if it is a single and linked to a cluster of images if it is a representative. The Honeycomb browser was evaluated on MPEG-7 benchmark colour database of 5,466 images. The average of retrieval was efficient compared to explorer and ImageSorter browsers about 32.4 and 5.5 lower respectively.

Image Hub Explorer tool was developed in [18] and can provide four functions, visualize, detect and examine, solve, and interpret. The first function displays a large number of images via a multi-dimensional scaling and explores characteristics of global image data. The second function enables users to treats with each class separately such as examining point type distribution (safe, borderline, and rare) points in terms of *k*-nearest neighbours (*k-NN*) for each feature representation. The third function allows the users to examine different local image features (SIFT, SURF, BREIF, and ORB) in addition to different metrics ($simcos_s$, $simhub_s$, mutual proximity, NICDM, and local scaling). The fourth function enables the users to explore the usefulness of each above feature. More details in the original paper about these functions. A searching process allows the users to browse the image he/she interests (i.e. query image). Then the procedure of feature extraction is made for the query image and matching with those features in the database to retrieve the most similar images. A clustering algorithm *K*-means ++ was used to cluster above visual words features. Different codebooks sizes were generated with different databases. The Image Hub Explorer tool was conducted on Leeds Butterfly, 17 flowers, Caltech101, Essex face, and ImageNet databases.

More recently, Travi-Navi a vision-guided navigation system has been developed and presented in [19]. Multi-functions were presented by the Travi-Navi system, Motion engine, Trace packing, Navigation engine, Motion hints for image capture, Image matching and retrieval, Lock-on at entrances, and Tracking follower. The usage of the system can be in a mall. The implementation of Travi-Navi was done on the Android platform (version .2.2) and OpenCV library was adopted to implement image processing and matching via JNI. Here, we interested the Image matching and retrieval function, where the BOVW method was used, ORB visual words features were extracted and the *k*-means clustering algorithm was used in quantization.

5.4. Region-of-Interest (ROI) Approach

A specific area is firstly determined in this approach and then features are extracted to represent an image that will be used in CBIR or object recognition. Developed CBIR system in [20], a method sorts images based on their entropies in a database to increase the efficiency of image searching. When a query image is made then its entropy is calculated and compared with those in the database to reach the closest value and regarded as started searching point. The system allows a user to determine ROI of query Q (n x m) by a mouse. This region will be shift as window on the image I (y x z) of database and the entropy is calculated between two regions Q and I. The resulted value compared to the threshold. If the difference value is less, extracting feature process will be applied on both regions using DCT. This procedure is repeated with the remaining images to retrieve a list the most similar images.

Ren *at el.* [21] used the adaptive mean shift algorithm to extract superpixels from the image in $CIE\ L^*a^*b^*$ colour space and then clustered the superpixels by using Gaussian Mixture Model (GMM) based on colour similarity. Because of the sensitivity of GMM to the noise, the mean shift clustering algorithm was applied first. Authors modified RankPage algorithm to overcome the problem of segmentation (i.e. over-segmentation). The result is saliency map that will be treated later to reach more accurate image representation. Three scenarios of image representation were proposed and tested. The first one is local feature extraction (SIFT), Sparse coding, and then max pooling (ScSPM+mp). The second one is local extraction (SIFT), Weighted Sparse coding, and then weighted max pooling (WScSPM-I and -II +mp), where version I means using the weight on both the reconstruction error and sparisty terms because they correspond to the feature coding. And version II means that the weight is used only on the reconstruction error because the important features may have smaller error. The third one is calculating the saliency, Weighted Sparse coding, and then saliency weighted max pooling (WScSPM-I and –II +smp). In addition, ScSPM+smp was tested.

Experiments for object recognition were conducted on four databases, Flower-17, Flower-102, ImageNet-Al, and Pascal VOC 2006. Results showed the order of increasing accuracy as follows: ScSPM+smp, WScSPM-I+smp, WScSPM-II+smp means regarding saliency and weights increase the effectiveness of image representation [21].

Grycuk *at el*. [22] presented an algorithm for CBIR. First, mean shift clustering algorithm was adapted (i.e. *h* radius parameter) and used to cluster SIFT, SURF, or PCA SIFT features that were extracted from binary images. Inspirit the method of document retrieval (DR); a dictionary was created that contains cluster and image ID which means wordID and document ID respectively in DR method. For each image TF-IDF value that represents the specific cluster in a given image and taking account its occurrence in all images, the formula was adapted from DR method. The algorithm evaluation was made on 1280 images which are divided into 11 categories. Averages of retrieval showed that estimated the parameter *h* to 288 approached to fixed *h*=250, where mean averages were 77.36 and 78.36 respectively.

Sara *at el.* [23] developed a retrieval system based on region using a joint scalable Bayesian segmentation for texture images. First, the image was transformed into Wavelet Discrete Transform (WDT) three sub-bands. The initial segmentation was made on approximation sub-band using region growing algorithm and means of regions were calculated and aggregated to represent feature vector F_{app}. Meanwhile, the Bayesian segmentation was applied on details sub-bands and feature F_j was extracted. A Pseudo Maximum Likelihood (PML) was adapted in terms of a smoothing parameter β, if $\beta=0$ means the spatial information doesn't regarded, otherwise the role of neighbourhoods is regarded. This step is modification of the ICM algorithm.

Experiment conducted on SPOT3 database satellite images (urban, field, water, mountain, and aeroport) categories. Results showed that the parameter β with non-zero is effective in segmentation.

5.5. Relevance Feedback (RF) Approach

Interaction between system and user to refine the result (retrieved images) is principle in this approach. A general procedure is presenting a sample of images (training examples) for the user who gives a feedback to the system by selecting positive and/or negative examples. Consequently, the result will be refined by the system to produce adjusted retrieved image list in the next round. The procedure can be run iteratively until the user is satisfied with the desired images [24]. However, many issues with RF approach such as the size of training examples where the small size cannot works with Support Vector Machine (SVM) classifier or the system needs more information rather than (relevant/ irrelevant) example images and this makes users burden. Zhue and Huang presented a comprehensive survey [25].

Proposed RF mechanism in [26] used a probability $P=Sn/Tn$, where Sn is the number of times selected an image by the user and Tn is total number of times image retrieved so far. Two features were extracted to represent images. The first one is a shape feature that uses the canny edge detection to obtain a Gradient image, where ‘1’ represents the edge and ‘0’ represents non edge. The Gradient image was divided into 3 x 3 neighbours and a code from each (3 x 3) block was extracted. At the end, total codes create feature vector to represent the shape descriptor. The second feature is colour histogram of the image. Retrieved images based on the similarity value that was calculated between the shape features of query and database images using Hamming distance. Meanwhile, sorting images was made using Euclidean distance. The retrieved images will be shown to the user who selects images at the session and refined using an associated probability value with each image as mentioned earlier. The system was evaluated on Corel database of ten thousands images with different categories. Results showed average precisions (88, 85, 82, 72, and 68) % for top 60, 70, 80, 90, 100 retrieved images respectively.

Papadopoulo *at el*. [27] developed a system where the pre-processing step is segmenting images into regions. Each region was represented by a visual feature vector by adapting BOVW approach, OpponentSIFT feature was extracted at a set of keypoints and a histogram of 1000 visual words was calculated. This was done for all images at off-line phase. Meanwhile, a RF mechanism was developed to be at on-line phase. The mechanism consists of two modes, training and evaluation. At the training mode, images are formed into sets and then users are asked to test every set. Temporal and spatial gaze features are generated for regions that have been spot by the users. Then users are asked again to annotate relevant and irrelevant regions manually. The calculated gaze features of annotated regions will be used for training the user relevance assessment predictor.

At the evaluation mode, a ranked list of images will be displayed to the user (e.g. 10-top images) who asses images and the predictor can estimate a degree of the user’s relevance assessment. Subsequently, all regions have been seen by the user will be aggregated in a composite image that is based to reorder images according to associated degrees. Authors used many techniques to reduce the dimensionality of gaze features and the best one was Principle Component Analysis (PCA). Experiments conducted on 15 subjects and the best Mean Average Precision (MAP) was at fifth iteration [27].

Kundu *at el*. [28] presented CBIR system using a relevance feedback mechanism based on graph. The pre-processing step was colour space conversion from *RGB* to *CIE* $L^*a^*b^*$ that better represent human perception. In addition *YCbCr* was tested but its performance was less. After the conversion process, images were converted into Non-Subsampled Contourlet Transform (NSCT) that is fully shift-invariant, multi-scale, and multi-direction expansion with fast applicability. The best achievement is using 4 sub-bands 1, 2, 4, and 4 decompositions (1+2+4+4=11) and was done for luminance channel (L^*) to represent texture information and for chromatics channels a^* and b^* to represent colour information (i.e. 33 sub-bands for each image). Each sub-band summarized by its mean, standard deviation and energy. Consequently, the length of the feature is 99D that represents the image in the database. In terms of complexity reduction, Maximal Information Compression Index (MICI) technique was used to reduce the dimensionality.

Relevance feedback mechanism has been incorporated with system to refine a list of retrieved images based on the graph, where vertices refer to images and edges refer to similarities between images using a Gaussian kernel function. In other words, database images were constructed as a graph based on similarities between above represented image features. The system can refine the retrieved list iteratively by following a random-walk-based re-ranking algorithm after marketing relevant images by the user until results satisfied the query.

The system evaluated using three databases, SIMPLIcity (1000 images, 10 categories), Oliva (2600 images, 8 categories), and Caltach256 (2500 images, 100 categories). A multi-class Least-Squares Support Vector Machine (LS-SVM) classifier was used and for training (20, 40, 60, and 100) % was tested. Results showed that the best performance with SIMPLIcity using 60%, Oliva using 40%, and Caltach256 using 20% training [28].

In terms of relevance feedback, the average precision (AP) was increased every iteration. Ten iterations have been shown for each database. APs of three databases were higher than existed RF methods (ESM and SSL). In addition, above three databases can be ordered according to APs as follows: SIMPLIcity, Oliva, and Caltach256. The reason behind is that the first database is the smallest in size comparing with the two others [28].

Recently, Duan *at el*. [45] introduced a method for relevance feedback, where learning machine based on Gaussian kernel called extreme learning machine (ELM). The method extracted three features, SIFT (BOVW), colour histogram (BOVW), and Local Binary Patterns (LBP) to represent images. Obtained features were fed to group of learning machines (ELM classifiers) to solidite class label votes. Updating of the classifiers was made corresponding to positive and negative samples.

Experiments were applied on Corel database (60 classes) and part of it WANG standard database (10 classes), 100 images in each class. Three scenarios were made ELM, 3-ELM, and 9-ELM classifiers in experiments implementation. Results of average precision showed that 9-ELM approaches 3-ELM meanwhile ELM outperforms both, where AP was about 100% after 4 iterations using the WANG images and about 72% after 5 iterations using the Corel images because the database is larger.

5.6. Indexing Approach

A direction of CBIR research using mentioned approaches in previous sections concerns to effectiveness in performance more than efficiency whereas the indexing techniques direction concerns to efficiency as well as accuracy.

In general, different indexing techniques have been proposed for big data (e.g. documents, images, videos, cloud computing). Recent two surveys [31, 32] have presented a rich review that highlighted different tree data structure which concern with data organization (e.g. B-tree [33], R+-tree[34], and KD+-tree[35]). Searching time for demanded query in a large database can be achieved in sublinear $(o(|X|))$, logarithmic $(O(log|X|))$, or constant $(O(1))$ using the indexing techniques compared to that linear searching time $O(|X|)$, where q query point needs to matching with whole database X points. However, tree indexing technique has some drawbacks such as storage requirement and difficult to deal with high-dimensional data. Hashing methods treated these two issues by storing compact binary codes that represent the original data and using Hamming distance (8) to find nearest neighbors efficiently [31].

In principle of the hashing methods is partitioning the data and generating a hash bits according to a hash function such that $HF = \{hf_1, \dots, hf_K\}$ for a object $s \in \mathbb{R}^D$ to compute a *K*-bit binary code $c = \{c_1, \dots, c_K\}$ for s as

$$c = \{hf_1(s), \dots, hf_K(s)\} \qquad (5)$$

where the k^{th} bit is calculated as $c_k = hf_k(s)$. The role of the hash function is to map the original data points to a binary valued space (i.e. Hamming space):

$$HF: s \rightarrow \{hf_1(s), \dots, hf_K(s)\} \qquad (6)$$

For a set of hash functions, mapping all the objects data $S = \{s_n\}_{n=1}^{N} \in \mathbb{R}^{D \times N}$to binary codes as:

$$C = HF(S) = \{hf_1(S), \dots, hf_K(S)\} \qquad (7)$$

The next step can made Approximate Nearest Neighbor (ANN) search in Hamming space, where the computation cost is reduced. Hamming distance between two binary codes c_i and c_j as follows:

$$Dis(c_i, c_j) = |c_i - c_j| = \sum_{k=1}^{K} |hf_k(c_i) - hf_k(c_j)| \qquad (8)$$

The survey [31] clarifies steps in ANN search which are designing *HFs*, generating *C* and indexing the data objects as well as categories of learning based hashing methods.

A. Linear and Nonlinear based on the form of *HF*.
B. Data-Dependent and Data-Independent based on learning project functions from training data or not.
C. Unsupervised, Supervised, and Semi-Supervised depend on availability of label information
D. Pointwise, Pairwise, Triple-wise and Listwise are subcategories to supervised or semi-supervised hashing techniques.
E. Single-Shot learning and Multiple-Shot Learning relate to required characteristics of the hash codes.
F. Non-Weighted and Weighted weights are combined to *HF* to increase discrimination power among binary codes.

Table 6 illustrates some different examples of learning methods corresponding to their categories belong to.

Table 6. Four different learning methods and their categories

Learning Method	Category
LSH [36]	Data-dependent and Linear
Spectral hashing [37]	Non-linear and unsupervised
Semi-Supervised hashing [38]	Semi-supervised
3Column generation hashing [39]	Supervised

Note: For more examples see Ref. [31].

Meanwhile, the survey in [32] presented a comprehensive study which highlighted that above classic indexing methods can not deal with unknown big data such that on cloud whereas Artificial Intelligence (AI) indexing methods can detect behavior of such data. Hence, the study categorized the indexing methods to Non Artificial Intelligence (NAI), Artificial Intelligence (AI), and Collaborative Artificial Intelligence (CAI). NAI is alternative name to traditional indexing methods that are clarified earlier. AI has ability such as fuzzy to explore a pattern of big data which is not fixed. Deep neural networks are deep learning for hashing that attracted researchers since 2006. Eight deep learning hashing methods have been explained and analyzed in [31]. Finally, CAI integrates the first and second categories to down faced issues in both.

In terms CBIR, Liu and Shao [40] presented unsupervised image hashing by developing a framework named Evolutionary Compact Embedding (ECE) to generate corresponding binary codes. Genetic programming (GP) and Adaboost strategy were used to learn each bit of ECE $[c_1(s), \dots, c_k(s), \dots, c_K(s)]$ code iteratively to embed high into low dimension in Hamming space. This is the first version of ECE. Due to time consuming of GP, the random batch parallel learning technique was used. Given a training set of images $I = \{I_1, \dots, I_n, \dots, I_N\}$ with class labels, the first half *N/2* pairs were randomly grouped with label 0 and the second half *N/2* pairs with label 1. So, the learning process can be made for each pair at the same time. This is the second version of ECE.

The experimental retrieval was conducted on SIFT 1M and GIST 1M databases. Table 7 (a-b) illustrates mean average precision of proposed ECE version1 and version2 along code length in bits for SIFT 1M and GIST 1M images. Although the second version is faster than first one, the first version is more accurate especially when code length is increasing.

Table 7. MAP using ECE version1 and 2 on SIFT 1M and GIST databases

Code length	15	30	50	65	80	95
ECE version1	0.3	0.35	0.4	0.44	0.47	0.5
ECE version2	0.28	0.33	0.35	0.45	0.45	0.47

(a) SIFT 1M database

Code length	15	30	50	65	80	95
ECE version1	0.17	0.21	0.24	0.25	0.26	0.27
ECE version2	0.19	0.20	0.22	0.23	0.24	0.26

(b) GIST 1M database

A method in [42] was developed for CBIR that extracts visual and text features to build image database and retrieve similar images. Hence, the method is called semantic-assisted visual hashing (SAVH) that extracts image visual features (SIFT Bag-Of-Visual-Word). At the same time, the method extracts associated text automatically and represents it as visual graph (Bag-Of-Text-Word) in addition to the assistance of topic hypergraph and latent semantic topics to help on modeling and correlating between images. This makes the approach unsupervised and recognizable comparing with the other similar approaches that use fixed labels or texts. The remaining of procedure is learning above two features by visual hash functions to produce corresponding binary codes using a linear regression model.

In terms of experiment implementation, three databases were used Wiki (2,866 images), MIR Flickr (25,000 images), NUS-WIDE (186,643 images). Visual extracted features were (128D, 1000D, and 500D) BOVW and text extracted features were (10D, 457D, 1000D) BOTW respectively.

Table 8 shows mean average precision of image retrieval for code binary length 64 and 128 bits respectively. It is clear that the best performance with the Flicker images using the SIFT feature vector 1000D in length and text word feature vector 457D in length. This means the visual feature is more discriminate with the high dimension meanwhile the text feature with the moderate dimension. In terms of code binary length, there is no big difference between 64 and 128 bits.

Table 8. MAP using 64- and 128-bit for Wiki, MIR Flickr, and NUS-WIDE

Wiki		MIR Flickr		NUS-WIDE	
64-bit	128-bit	64-bit	128-bit	64-bit	128-bit
0.1914	0.1991	0.668	0.6704	0.5193	0.5281

At the same field of research, unsupervised hashing approach based on bilinear projection to map the original data to binary codes was presented in [43]. Here, the SIFT feature was investigated in different image decompositions (4x4, 4x8, and 8x8) and extracted as histograms in length 8, 4, and 2 bins respectively.

The performance of these SIFT local features compared to that of SIFT global feature (1x128). Table 9 illustrates mean average precision resulted from experiments that were conducted on Caltech256, SUN397, and Flickr databases. The best performance of image retrieval was when images were divided into 4x4 blocks and 8-bin histograms were calculated. The length of generated code was 32-bit and the number of extracted SIFT features was 1000.

Table 9. MAP using different block divisions for Caltech256, SUN397, and Flickr 1M

Database	Caltech256 (30607 images)	SUN397 (108754 images)	Flickr 1M (million images)
Test set	1000 images	80964 images	1K images
Training set	The rest images	27790 images	150000 images
Block division	MAP		
8x16	0.239	0.122	0.250
4x32	0.231	0.120	0.241
2x64	0.229	0.117	0.239
1x128	0.225	0.115	0.235

The apposite to unsupervised is supervised learning technique when a label is available as mentioned earlier. Hence, a proposed method in [46] exploited matrix factorization collectively for local features and labels consistency that were formulated by Laplacian matrices. The BOVW local SIFT feature vector (128D) was calculated to represent an image and topics feature vector (10D) was used to represent text.

A cross- model hashing based on obtained collective matrix factorization was made to learn hash functions. Produced binary codes satisfied a good discriminate for Wiki images, where mean average precisions were 0.2572, 0.2759, 0.2863, and 0.2913 for 16, 32, 64, 128 binary code bits respectively. These results were promise compared to similar approaches from literature. The case was opposite when NUS-WIDE database were used. The mean average precisions were degraded compared to the best method SCM_seq from literature and proposed method Supervised Matrix Factorization Hashing (SMFH) as shown in Table 10.

Table 10. MAP proposed SMFH and existed SCM methods

Method	16-bit	32-bit	64-bit	128-bit
SMFH	0.4553	0.4623	0.4658	0.4680
SCM_seq	0.5219	0.5336	0.5365	0.5189

6. Conclusions

Firstly, this paper demonstrated typical components of CBIR system. Different research areas in CBIR field, clustering, bag of visual words (BOVW), browsing, region of interest (ROI), relevance feedback (RF), and indexing are defined and clarified in terms of principle and framework. All developed approaches participates two common components which are extracted features and similarity measures. Researchers in each method try to reduce the challenge of CBIR so-called "*Semantic gap*" between high level conceptual meaning and low-level features as well as the efficiency of image retrieval especially with the indexing field.

The main idea of the clustering is to group low-level features into mid-level features. This means obtained clusters are more meaningful segments. One of challenge is over- and under-segmentation. The case with the BOVW is generated huge of visual words and built a dictionary after quantization process that made by one of clustering algorithms, where *k*-means and its variants are mostly used via its simplicity. Lost spatial information in extracted features is an issue with this method. Meanwhile, the Browsing method bears different idea which is navigating a large collection of images in a screen panel instead of query by image example. How to visualize image collection in a specific screen size is a challenge. In ROI method, determining the region that the user interests and then starting the steps of CBIR. The RF provides interaction with the user to make useful feedback to refine the resulted list of retrieved images until the requirements are satisfied. Involving the user is one issue in this method.

So far, effectiveness is more concerned with these methods meanwhile indexing approach concerned with accuracy and efficiency. Hence, the main idea is how to store images in a way that guarantee response efficiently using tree data structure or representing images by generating binary codes using different hashing functions and learning different paradigms.

We have observed that above methods are overlapped although each one represents research area in CBIR field. For example, clustering is needed in BOVW and the last is used in the others. In addition, too much works have been done using BOVW that is based on idea of information retrieval (i.e. documents).

A review and some analysis for works from literature are made in this paper to help new researchers from different scientific areas who are interested in CBIR field determining the direction of research area.

7. References

1. Du, H., Al-Jubouri, H. & Sellahewa, H., (2014). *"Effectiveness of image features and similarity measures in cluster-based approaches for content-based image retrieval"*. Baltimore, Maryland, USA , SPIE Sensing Technology+ Applications, pp. 912008-912008.

2. Datta, R., Joshi, D., Li, J. & Wang, J. Z., (2008). *"Image retrieval: Ideas, influences, and trends of the new age"*. ACM Computing Surveys (CSUR), 40(2), p. 5.
3. Witten, L. H., Frank, E. & Hall, M. A., (2011). *"Data Mining Practical Machine Learning Tools and Techniques"*. Burlington, USA: Elsevier.
4. Wang, J. Z., Li, J. & Wiederhold, G., (2001). *"SIMPLIcity: Semantics*-*sensitive integrated matching for picture libraries"*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 23(9), pp. 947-963.
5. Nezamabadi-Pour, H. & Saryazdi, S., (2005). *"Object-based image indexing and retrieval in DCT domain using clustering techniques"*. Proceedings of World Academy of Science Engineering and Technology, pp. 98-101.
6. Lokoč, J., Novák, D., Batko, M. & Skopal, T., (2012). *"Visual image search: feature signatures or/and global descriptors"*. Berlin / Heidelberg, Springer, pp. 177-191.
7. Zhang, B., Luo, H. and Fan, J., (2016), *"Statistical modeling for automatic image indexing and retrieval"*. Neurocomputing, *207*, pp.105-119.
8. Tuytelaars, T. & Mikolajczyk, K., (2008). *"Local invariant feature detectors: a survey"*. Foundations and Trends{\textregistered} in Computer Graphics and Vision, 3(3), pp. 177-280.

9. Lowe, D. G., (2004). *"Distinctive image features from Scale-Invariant Keypoints"*. Int. J. Comput. Vision, 60(2), pp. 91-110. J.
10. Vieux, R., Benois-Pineau, J. & Domenger, J.-P., (2012). "*Content Based Image Retrieval Using Bag-of-regions"*, Berlin, Heidelberg, Springer-Verlag, pp. 507-517.
11. Pedrosa, G.V. and Traina, A.J., (2013). *"From bag-of-visual-words to bag-of-visual-phrases using n-grams"*. In Graphics, Patterns and Images (SIBGRAPI), 2013 26th SIBGRAPI-Conference on (pp. 304-311), IEEE.
12. Zheng, L. and Wang, S., (2013). "*Visual phraselet: Refining spatial constraints for large scale image search"*. IEEE Signal Processing Letters, 20(4), pp.391-394.
13. Ren, Y., Bugeau, A. and Benois-Pineau, J., (2014). *"Bag-of-bags of words irregular graph pyramids vs spatial pyramid matching for image retrieval"*. In Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on (pp. 1-6), IEEE.
14. Karakasis, E.G., Amanatiadis, A., Gasteratos, A. and Chatzichristofis, S.A., (2015). *"Image moment invariants as local features for content based image retrieval using the bag-of-visual-words model"*. Pattern Recognition Letters, *55*, pp.22-27.
15. Plant, W. and Schaefer, G., (2009). *"Navigation and browsing of image databases"*. In Soft Computing and Pattern Recognition, 2009. SOCPAR'09. International Conference of (pp. 750-755).
16. Hilliges, O., Kunath, P., Pryakhin, A., Butz, A. and Kriegel, H.P., (2007). *"Browsing and sorting digital pictures using automatic image classification and quality analysis"*. Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments, pp.882-891.
17. Plant, W. and Schaefer, G., (2010). *"Image retrieval on the honeycomb image browser"*. In Image Processing (ICIP), 2010 17th IEEE International Conference on (pp. 3161-3164), IEEE.
18. Tomašev, N. and Mladenić, D., (2015). "Image hub explorer: Evaluating representations and metrics for content-based image retrieval and object recognition". Multimedia Tools and Applications, 74(24), pp.11653-11682.
19. Zheng, Y., Shen, G., Li, L., Zhao, C., Li, M. and Zhao, F., (2017). *"Travi-navi: Self-deployable indoor navigation system"*. *IEEE/ACM Transactions on Networking*.
20. Yung-Gi, W., (2006). *"Region of interest image indexing system by DCT and entropy"*. GVIP Journal, 6(4), pp. 7-16.
21. Ren, Z., Gao, S., Chia, L.T. and Tsang, I.W.H., (2014). *"Region-based saliency detection and its application in object recognition"*. IEEE Transactions on Circuits and Systems for Video Technology, 24(5), pp.769-779.
22. Grycuk, R., Gabryel, M., Korytkowski, M. and Scherer, R., (2014). *"Content-based image indexing by data clustering and inverse document frequency"*. In International Conference: Beyond Databases, Architectures and Structures (pp. 374-383), Springer.
23. Sakji-Nsibi, S. and Benazza-Benyahia, A., (2016). *"Region-based image retrieval using a joint scalable Bayesian segmentation and feature extraction"*. In Signal Processing Conference(EUSIPCO), 2016 24th European (pp. 1272-1276), IEEE.
24. Pinjarkar, L., Sharma, M. and Mehta, K., (2012). *"Relevance feedback techniques in content based image retrieval: A survey"*. Int. J. Eng, 1(9), pp.19-23.
25. Zhou, X. S. & Huang, T. S., (2003). *"Relevance feedback in image retrieval: A comprehensive review"*. Multimedia Systems, 8(6), pp. 536-544.

26. Yasmin, M., Mohsin, S., Irum, I. and Sharif, M., (2013). *“Content based image retrieval by shape, color and relevance feedback”*. Life Science Journal, 10(4s), pp.593-598.
27. Papadopoulos, G.T., Apostolakis, K.C. and Daras, P., (2014). *“Gaze-based relevance feedback for realizing region-based image retrieval”*. IEEE Transactions on Multimedia, *16*(2), pp.440-454.
28. Kundu, M.K., Chowdhury, M. and Bulò, S.R., (2015). *“A graph-based relevance feedback mechanism in content-based image retrieval”*. Knowledge-Based Systems, *73*, pp.254-264.
29. Feng, D., Siu, W. C. & Zhang, a. H. J., (2003). “*Multimedia informaton retrieval and management technolgical fundamentals and applications”*. Verlag Berlin Heidelberg, Germany: Springer Science & Business Media.
30. Al-Jubouri, H., (2015). “*Multi evidence fusion scheme for content-based image retrieval by clustering localised colour and texture features*”. Doctoral thesis, University of Buckingham.
31. Wang, J., et al, (2015). “*Learning to Hash for Indexing Big Data - A Survey*” IEEE, vol. 104, no. 1, pp. 34-57.
32. Gani, A., Siddiqa, A., Shamshirband, S. et al. (2016) “*A survey on indexing techniques for big data: taxonomy and performance evaluation*”. Knowledge Information System, Springer, vol. 46, pp. 241–284.
33. Graefe, G., (2010). “*A survey of B-tree locking techniques*”. ACM Trans Database Syst., vol. 35, pp.16:1--16:26.
34. Sellis, TK., Roussopoulos, N., Faloutsos, C., (1987). “*The R+-tree: a dynamic index for multi-dimensional objects*”. Paper presented at the proceedings of the 13th international conference on very large data bases, pp. 507-518.
35. Wei, L-Y, Hsu, Y-T, Peng, W-C, and Lee, W-C, (2014). “*Indexing spatial data in cloud data managements*”. Pervasive Mobile Comput, vol. 15, pp. 48-61.
36. Gionis, A., Indyk, P. and Motwani, R., (1999). “Similarity search in high dimensions via hashing”. In Proc. of 25th International Conference on Very Large Data Bases, pp. 518–529.
37. Weiss, Y., Torralba, A., and Fergus, R., (2008). “*Spectral hashing*”. In *Proc. of Advances in Neural Information Processing Systems*, vol. 21, pp. 1753–1760.
38. Wang, J., Kumar, S., and Chang, S.-F., (2012). “Semi-supervised hashing for large scale search”. IEEE Transactions on Pattern Analysis and Machine Intelligence.
39. Li, X., Lin, G., Shen, A., den Hengel, V. , and Dick, A. (2013). “*Learning hash functions using column generation*”. In Proceedings of the 30th International Conference on Machine Learning, pp. 142–150.
40. Liu, L. and Shao, L., (2015). “*Sequential Compact Code Learning for Unsupervised Image Hashing*”. IEEE Transactions on Neural Networks and Learning Systems , vol. 27, pp. 2526 – 2536.
41. Zhu, L., Shen, J., and Xie, L. (2016). “*Unsupervised Visual Hashing with Semantic Assistant for Content-based Image Retrieval”*. IEEE Transactions on Knowledge and Data Engineering, vol. 29, pp. 472-486.

42. Liu, L., Yu, M., and Shao, L., (2015). "*Unsupervised Local Feature Hashing forImage Similarity Search*". IEEE Transactions on Cybernetics, vol. 46, pp. 2548 – 2558.
43. Tang, J., Wang, K., and Shao, L., (2016). "*Supervised Matrix Factorization Hashing for Cross-Modal Retrieval".* IEEE Transactions on Image Processing, vol.25, pp. 3157 – 3166.
44. Zeng, S., Huang, R., Wang, H., and Kang, Z., (2016). "*Image Retrieval Using Spatiograms of Colors Quantized by Gaussian Mixture Models*". Neurocomput, Elsevier Science Publishers B. V., vol. 171, pp. 673-684.
45. Duan L., Dong S., Cui S., Ma W. (2016) "*Extreme Learning Machine with Gaussian Kernel Based Relevance Feedback Scheme for Image Retrieval*". Proceedings in Adaptation, Learning and Optimization, vol 6. Springer, Cham.
46. Tang, J., Wang, K., Shao, L., (2016). " *Supervised Matrix Factorization Hashing for Cross-Modal Retrieval*". IEEE Transactions on Image Processing, Vol. 25, pp. 3157 – 3166.