

# A Survey on Content-based Image Retrieval

Mohamed Maher Ben Ismail  
College of Computer and Information Sciences,  
King Saud University, Riyadh, KSA

**Abstract**—The widespread of smart devices along with the exponential growth of virtual societies yield big digital image databases. These databases can be counter-productive if they are not coupled with efficient Content-Based Image Retrieval (CBIR) tools. The last decade has witnessed the introduction of promising CBIR systems and promoted applications in various fields. In this article, a survey on state of the art content based image retrieval including empirical and theoretical work is proposed. This work also includes publications that cover research aspects relevant to CBIR area. Namely, unsupervised and supervised learning and fusion techniques along with low-level image visual descriptors have been reported. Moreover, challenges and applications that emerged to support CBIR research have been discussed in this work.

**Keywords**—Image retrieval; Content-based image retrieval; Supervised learning; Unsupervised learning

## I. INTRODUCTION

The importance of digital image databases depends on how friendly and accurately users can retrieve images of interest. Therefore, advanced search and retrieval tools have been perceived as an urgent need for various image retrieval applications. The earliest search engines have adopted text-based image retrieval approaches. These solutions have shown drastic limitations because digital images to be mined are either not labelled or annotated using inaccurate keywords. In other words, text-based retrieval approaches necessitate manual annotation of the whole image collections. However, this tedious manual task is not feasible for large image databases.

Content-Based Image Retrieval (CBIR) emerged as a promising substitute to surpass the challenges met by text-based image retrieval solutions. In fact, digital images, which are mined using CBIR system, are represented using a set of visual features. As illustrated in Figure 1, typical CBIR system consists of an offline phase which aims at extracting and storing the visual feature vectors from the database images. On the other hand, the online phase allows the user to start the retrieval task by providing his query image. Finally, typical CBIR system returns a set of images visually relevant to the user query. However, its main drawback consists in the assumption that the visual similarity reflects the semantic resemblance. This assumption does not hold because of the semantic gap [1] between the higher level meaning and the low-level visual features.

Despite the promising results achieved by large-scale applications, such as Yahoo! and Google TM, bridging the semantic gap remains a challenging task for CBIR researchers. Also, social network usage, along with the widespread of low cost smart devices, has re-boosted the research related to image

retrieval. This represented a paradigm shift in the research aims of the new generation of CBIR researchers. Image representation, feature extraction and similarity computation also as a critical component of typical CBIR systems. More specifically, in order to design successful CBIR system, researchers investigated various contributions for these components [15, 16, 17]. Comprehensive surveys on CBIR systems have been proposed to report the progress reached by the research community [1, 3, 4, 5, 6, 7]. Other surveys have been elaborated on highly relevant topics to CBIR systems. Namely, researches on high-dimensional data indexing [11], relevance feedback [10], and medical application of CBIR [13, 14] have been surveyed.

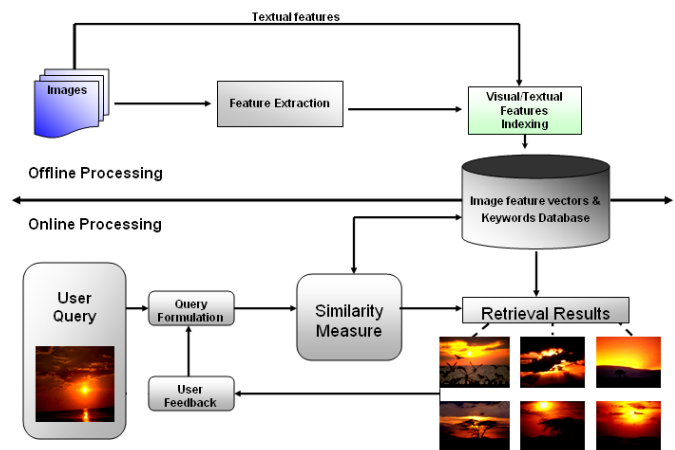


Fig. 1. Overview of typical CBIR system.

The continuous growth of associated research spanning several domains during the last decade and the increase in the number of researchers investigating CBIR are the main motivations of this survey. This article fully surveys, investigates and appraises state of the art research and future facet of CBIR systems. The rest of this article is organised as follows: Section 2 focuses on state of the art methods used to bridge down the ‘semantic gap’. Low-level features proposed to capture high-level query semantic are outlined in Section 3. In Section 4, CBIR recent challenges and applications are addressed. Emerging research issues related to CBIR systems are introduced in Section 5. Finally, Section 6 concludes the survey.

## II. BRIDGING THE SEMANTIC GAP

Researcher contributions to bridge the semantic gap can be categorised into different manner based on the adopted angle of view. In particular, if one takes into consideration the application domain, the state of the art techniques can be

perceived as those focusing on scenery image retrieval [18, 19, 20], web images retrieval [21, 22], artwork image retrieval [23], etc. This article spotlights on the approaches used to develop high level semantic based CBIR. These approaches are grouped into: (i) Approaches based on supervised or unsupervised learning techniques to learn the association between low level descriptors and query semantics, and (ii) Fusion based image retrieval approaches.

#### A. Supervised and Unsupervised Learning

Over the last decade, researches have confirmed the limitations of single similarity measure to yield perceptually meaningful and robust image ranking. Learning based solutions have been proposed as promising an alternative to overcome this weakness. In particular, image categorization/classification has been designed as a pre-processing phase to speed up image retrieval from the large collection [76, 77]. Equivalently, unsupervised learning has been adapted to speedup retrieval process and enhances visualization performance when the images are not labelled or annotated [13, 14]. More specifically, the clustering phase can be represented as early retrieval stage that aims to handle unstructured image collections. On the other hand, classification techniques, along with the distance measurement, form the core of the retrieval process.

Recently, remarkable contributions have been proposed for unsupervised learning and supervised learning techniques and their application in various domains. This work focuses on novel approaches and applications dealing with content based image retrieval and closely related topics. The earlier efforts were focused on similarity measures and feature extraction components. Clustering and fast classification components have been promoted as practical hacks to overcome the scalability problem due to the continuous exponential growth of digital image databases. Clustering can be defined as the process of partitioning patterns into homogeneous categories in an unsupervised manner. It consists in dividing a collection of unlabelled data instances into groups such that instances belonging to different groups are as dissimilar as possible, and instances assigned to the same group are as similar as possible. Clustering aims to improve retrieval and visualization capability of typical image retrieval systems. One should mention that the performance of such retrieval systems is still affected by traditional challenges such as cluster conformity to the ground truth partition and visualization accuracy.

In [35], the authors suggested various taxonomies of clustering methods. Partitional clustering relies on hard or fuzzy objective function optimization. For hard clustering, binary membership value is assigned to each data instance whether it belongs or not to a cluster. Since clusters are rarely completely separated and are usually overlapping in real world applications, the use of crisp logic to describe the data is not appropriate to distinguish between instances laying on the overlapping boundaries. On the other hand, fuzzy logic allows the gradual evaluation of the membership of instances within a group/cluster. The Fuzzy C-Means (FCM) algorithm [36] is a popular fuzzy clustering algorithm. Multiple FCM based contributions have been reported along with different applications [37, 38]. However, these FCM based algorithms fail to discover the ground truth distribution of the data when it

contains asymmetric clusters and may yield non-optimal results. Probabilistic modelling is another alternative to fuzzy clustering. More specifically, mixture modelling based approaches in [80] rely on the assumption that instances in a given cluster are inherited from one of the multiple distributions, and aim at estimating the parameters of these distributions. Recently, in [39], the authors proposed to let data instances, belonging to different clusters, to be issued from various density functions. Such clustering techniques can be roughly categorised into three paradigms: statistical modelling, relational and objective function based paradigm.

Statistical modelling based clustering considers each cluster/category as a restrictively distributed pattern. Thus, the overall dataset is modelled as distribution mixture. The Expectation Maximization algorithm [40] is usually used to estimate the parameters of the mixture components/distributions corresponding to the cluster properties. The main appealing advantage of this mixture modelling approach is the information it provides on the data densities along with the final clustering partition [41]. Note that mixture components are not necessarily modelled as multivariate distribution. For instance, in [42], the authors intended to cluster image regions by characterizing each cluster using a 2-Dimensional HMM. However, if no probability measure is set-up to model a category/cluster, a mixture modelling can be achieved by grouping data instances and representing each cluster in a different similarity preserving space [43]. Typically, this approach represents the dataset for a more accurate classification rather than clustering it. In particular, applications such as remotely sensed image recognition, medical image classification, and automatic image annotation exploit this approach along with specified image collections with labelled training instances [71]. On the other hand, for relational approaches (pairwise distance based approaches) the mathematical representation of the data points is not critical [81]. This makes them widely applicable and appealing for various image based applications such as image retrieval which requires complex formulation of image signatures. However, the computation of the pairwise distances between data instances makes the relational methods timely expensive. In [44], the authors proposed a spectral clustering algorithm [78] to group similar images into homogeneous clusters and use the obtained partition information to enhance the retrieval process. More specifically, given the query image, clusters are learned in an unsupervised manner in order to enhance the retrieval accuracy. Objective function optimization is another traditional unsupervised learning technique. For instance, the popular K-means algorithm [72] minimizes the sum of the intra cluster distances. Notice that a major drawback of K-means is that the number of clusters has to be specified a priori.

A natural alternative to overcome this limitation consists in gradually increasing the number of clusters until the average distance between an instance and its corresponding cluster centre reaches a predefined threshold. The competitive agglomeration algorithm is a more advanced alternative to finding the number of image clusters [45]. From an application point of view, researchers from the multimedia community dedicated more attention for Web image clustering. In fact, the

unsupervised learning (clustering) techniques are valuable when meta-data is collected/extracted in addition to visual descriptors [33, 34, 37]. Unsupervised learning usually serves to recognize new images and assign them to some predefined categories before proceeding with the retrieval phase. Similarly, classification techniques can be grouped into two main categories. The first one contains the generative modelling based approaches. The second category regroups the discriminative modelling approaches such as decision trees and SVM classifiers where the class boundaries and the posterior probabilities are learned. The generative modelling uses Bayes formula along with the densities of data instances within each class to estimate the posterior probabilities. The researchers in [46] adopted Bayesian classification to propose an image retrieval system. Similarly, researchers in [26] used Bayesian classification in their proposed image retrieval approach. Their system aimed to capture high-level concepts of natural scenes using low-level features. Images were then automatically classified into outdoor or indoor images. Similarly, in [134] Bayesian network was adopted for indoor/outdoor image classification. Besides, image classification using SVM as supervised learning technique has been proposed in [47]. Recently, advanced multimedia query processing systems using SVM based MIL framework has been proposed in [48, 49]. MIL framework considers  $l$  training images as labelled bags where the bag  $i$  includes a set of instances represents a region  $i$  extracted from a training image  $i$ , and  $y_i$  indicates a positive or negative example for a given class value. The mapping of these bags to a new feature space, where supervised learning technique can be trained to classify unlabelled instances, is the key component of MIL. An image classification system has been proposed in [50] as a key component of an image retrieval system. Such classification techniques along with new information theory based clustering have boosted the integration of clustering and classification components into typical image retrieval systems. Different supervised learning techniques, such as neural network, were also considered for high-level concept learning. Specifically, in [19], the authors used 11 concepts. Namely, they considered water, fur, cloud, ice, grass, rock, road, sand, tree, skin, and brick. A large training set including low-level region descriptors is then used as input for neural network classifier. This aims to learn the association between high-level semantic (concept labels) and low-level descriptors. The main limitation of this approach is its high computational cost and the relatively large data required for training. Besides these learning techniques, decision trees methods such as ID3, C4.5 and CART are used to predict high-level categories [160]. In particular, the authors in [24] used CART algorithm to derive decision rules that associate image colour features to keywords such as Marine, Sunset, and Nocturne. In [161], a two-class (relevant and irrelevant) categorization model is solved using a C4.5 decision tree. Despite their robustness to noise and handling of missing data, decision trees exhibit a lack of modularity.

### B. Multimodal Fusion and Retrieval

The last decade has witnessed the proposal of various image retrieval approaches [82, 83, 84, 95, 13] which mainly rely on image and text modalities. One should notice that solutions for multimedia and speech retrieval have also been

proposed. This work focuses on image retrieval using text and image modalities only. In particular, it highlights the aggregation of these two modalities to enhance the retrieval accuracy. In other words, it considers this fusion as a typical technique that contributes considerably to the enhancement of the retrieval results. In fact, combining two query modalities can be counter-productive. In such scenario, query fusion aims at learning the optimal model to aggregate the different modalities. Recently, researchers have proposed some fusion techniques and applied them to image retrieval and image annotation systems [51]. In the following, a survey on multi-modal fusion techniques related to image retrieval application is outlined. Traditional fusion approach is intended to learn optimal rules to fuse multiple classifier outputs (decisions). This process requires some ground truth data to validate the obtained rules [89, 90]. Unlike this late fusion approach, another fusion alternative relies on the re-training of individual classifiers in order to optimize the fusion rule. For instance, the authors in [74] formulate the multi-modal fusion as two fold problem. Statistical modelling of the modalities represents the first fold. The second one consists of learning the optimal combination in an unsupervised manner. This fusion learning approach proved to be more effective than naive fusion for image retrieval [52]. Moreover, the fusion learning is performed offline which makes its application computationally inexpensive. This boosted the usage of modality fusion in retrieval related applications. However, over-fitting remains a considerable challenge for fusion learning. Thus, bagging [75] has been used to re-sample the data and prevent/reduce over-fitting. Despite these efforts, including fusion learning as the main component of image retrieval system represents a relatively new research area for pattern recognition and image processing researchers [86, 87, 88]. It is expected that it will boost research for various applications based on modalities and medias such as video, audio and text. In other words, future challenges are to fuse, in an efficient manner, as many information modalities as possible to overcome real world problems.

Local and global are the main approaches for combining diverse learners. Global approach assigns an average confidence degree to each learner based on the training set. On the other hand, local approach dedicates a confidence degree to the subspaces of the training set. This assumes that more accurate classification performance can be achieved using optimal data-based weights. During the training stage, an unsupervised grouping of the input data instances into homogeneous clusters is mandatory for local fusion approach. For supervised learning, unlabelled instances get appointed to regions, and the expert learner corresponding to this regions yield the fusion decision. Dynamic data classification during the testing stage is outlined in [143, 144, 145]. The classifier accuracies are obtained using sample vicinity in the feature space local regions. The most accurate classifier is then used to classify test samples. The Context-Dependent Fusion (CDF) in [145] is a local fusion approach that first groups the training samples into homogeneous context clusters. These clustering and local expert model selection phases are sequentially independent components of CDF. The authors in [146] proposed a generic context-dependent fusion approach which categorizes the feature space and combines the outputs of the

individual expert models simultaneously. Simple linear aggregation is used to predict aggregation weights for the individual classifier models. However, these weights may fail to reflect the integration between the individual learners. The researchers in [147] used clustering and feature selection to determine the most accurate classifier.

More specifically, the unsupervised clustering of the training samples aims to discover the fusion decision regions. Next, the highest-performance classifiers on each local region of the feature space are selected. The principal limitation of this work one classifier only is appointed for each region. In [148], another clustering and selection approach was proposed to partition the training samples into correctly and incorrectly classified samples. In fact, the feature space is partitioned by grouping the training samples. Then, the most accurate classifier in the test sample vicinity is appointed in order to provide the fusion decision. This makes this approach more computationally efficient than the approach in [147]. Recently, in [149, 150], a local fusion approach that partitions the data instances into homogeneous groups using their low-level features was proposed. Notice that the resulting clusters are used to aggregate the individual classifier decisions. In fact, aggregation weights are assigned to each individual classifier within each context. These weights reflect the relative accuracy of the classifiers within the different contexts. In order to address the sensitivity of this approach to noise and outliers, the researchers in [151] proposed a possibilistic approach that adapts the fusion technique to sub-regions of the feature space. The proposed clustering algorithm produces possibilistic memberships reflecting the typicality of data instances in order to reduce noise point impact. Then, expert learners are appointed to the resulting clusters. Notice that the aggregation weights are learned simultaneously for all classifiers. Finally, the aggregation weights corresponding to the closest cluster/context yield individual confidence values. Although this fusion approach proved to be effective for some applications, the proposed objective function remains prone to local minima.

### III. LOW-LEVEL FEATURES

The various promising low-level feature has been proposed to encode image content for CBIR systems. In the following, low-level descriptors and their use to enhance the retrieval accuracy are surveyed.

#### A. Colour Features

The most popular and widely used low level descriptor in image CBIR system is the colour feature. Several colour spaces have been defined for colour feature representation [91]. As reported in [95, 92, 93, 94, 20], the closest colour spaces to human perception include RGB, LUV, HSV, HMMD, YCrCb, and LAB. Also, various colour descriptors/features, such as colour histogram, colour moments, colour-covariance matrix, and colour coherence vector have been proposed for CBIR systems [96, 97, 98]. Similarly, in [99], colour structure, dominant colour, colour layout and scalable colour have been proposed as standard MPEG-7 colour features. Despite these efforts to encode the colour properties of the image, the proposed features have shown limitations to express image high level semantic. In order to alleviate this concern,

researchers proposed averaging colour of all pixels in a region/image as a colour feature [20, 98, 100]. However, this feature is affected by the image segmentation quality. In [100], the authors defined the dominant colour in HSV space as region perceptual colour. The dominant colour considers the largest bin of the colour histogram ( $10 * 4 * 4$  bins) of the region in the HSV space. Then, the dominant colour feature corresponds to the average HSV value of all the pixels in the selected bin. One should notice that if applied to non-homogeneous colour region due to inaccurate segmentation, taking the average colour does not yield representative colour feature. Thus, image pre-processing has been adopted as the main component of CBIR systems in order to remove noise from the images and enhance the segmentation quality [101,102].

#### B. Texture Features

Texture features aim at encoding another important visual property of images. In particular, texture feature represents the best some real world image content such as clouds, skin, trees, fabric, etc. Hence, texture feature contributes efficiently to reducing the gap between image content and their high level semantic for CBIR systems. For instance, spectral features extracted using wavelet transform [103] or Gabor filtering [104] have been widely adopted by CBIR systems. Similarly, statistical features such as wold features [105] and Tamura texture features [106] have been proposed in order to represent image visual content better and improve CBIR accuracy. Later, MPEG-7 adopted some statistical measures proposed in [106], such as directionality, regularity and coarseness, to define standard texture browsing descriptor [94, 98]. However, this statistic measure based features are not robust to scale and orientation variation [107].

Based on researcher contributions to propose accurate CBIR systems, wavelet and Gabor based texture features proved to match the best human vision and achieved the highest performance [98, 104, 108]. However, one should notice that these two texture features are sensitive to the shape of the image region [20, 104]. More specifically, they handle better the rectangular regions than arbitrarily shaped regions. Reshaping these non-rectangular regions by padding or applying some transforms emerged as an intuitive solution to overcome this drawback. Notice that region padding decreases the fidelity of the extracted texture feature to the image content. Another efficient extraction approach using iterative projection onto convex sets (POCS) has been proposed in [109] to extract texture features from non-regular regions. The Edge Histogram Descriptor (EHD) [98] proved to represent natural images efficiently. This edge feature encodes the spatial distribution of images edges. More specifically, it includes local edge histograms extracted from predefined sub-images and grouped into horizontal, vertical, diagonal, anti-diagonal and neutral edges. However, EHD is sensitive to scene and object distortions. Similarly, the researchers in [110] extracted the gradient vector from the sub-band images obtained using wavelet transform.

#### C. Shape Feature

Shape attributes such as consecutive boundary segments, circularity, aspect ratio, moment invariant, Fourier descriptors,

eccentricity and orientation have been widely exploited to represent an image in CBIR systems [20, 97, 111]. In [96], shape descriptors are extracted using area and second-order moments from gross image regions. For object-based image retrieval, MPEG-7 [98] has included three shape descriptors. Namely, a descriptor based on curvature scale space (CSS), a region based feature extracted using Zernik moments, and a 3-D shape descriptor based 3-D meshes of shape surface have been defined as MPEG-7 standard shape features. CSS descriptor is robust to scaling, translation and rotation variations. However, it shows some limitation to represent objects taken from the different point of view due to the resulting distortions. The authors in [112] addressed this limitation and proposed a variation of CSS descriptor that is robust to such affine transform.

#### D. Spatial Location

Spatial location represents another shape feature relevant to CBIR. In fact, if objects/regions exhibit similar texture and colour properties, then their respective spatial locations can serve as a more discriminative feature to represent these regions/objects [113, 114]. Minimum bounding box and the spatial centroid of regions represent the information used as a spatial location in [115]. However, such intrinsic spatial location does not reflect the semantic information in an effective manner compared to a relative spatial relationship. Thus, the authors in [116] used 2D-string, and its derivative structures formulate directional relationships such as 'below/above' and 'left/right', between objects. In [117], topological relationships have been included to enhance the performance of directional relationships. They outlined a spatial context modelling algorithm which relies on 6 pairwise spatial region relationships. Similarly, in [118], a promising approach using a composite region template (CRT) was introduced in order to capture semantic classes and the spatial arrangement of regions.

### IV. CBIR OFFSHOOTS: PROBLEMS AND APPLICATIONS OF THE NEW AGE

In [53], an early age survey on CBIR has been reported. Researcher effort was outlined as novel contributions to information retrieval, computer vision and machine learning applications. Nowadays, CBIR represents relatively mature research field. Moreover, a considerable number of researches shows the emergence of non-typical challenges, yet of high relevance to CBIR systems. In the following, these novel research directions are outlined.

#### A. Automatic Image Annotation

The typical goal of content based image retrieval system is to find relevant images to a given query when meta data is missing or unavailable. However, the uploaded digital images on a daily basis to image databases are rarely coupled with relevant labels or keywords. This triggered researches on automatic image annotation approaches [25, 31, 53, 59, 60, 62, 63]. Figure 2 shows the general architecture of a typical image annotation system. This system uses a set of labelled images for training. First, each image is segmented into regions and local features are extracted and used to describe each region. There are two main segmentation strategies; the first one partitions the image into a set of fixed sized blocks or grid

[138, 139]. The second one partitions the image into a number of homogeneous regions that share common features [140, 141, 142]. Ideally, each region corresponds to a different object in the image. After segmentation, each segmented block or region is represented by a feature vector. After segmenting all training images and extracting visual features from their regions, a machine learning algorithm is used to learn associations or joint probability distributions between these features and the keywords used to annotate the images. The testing part of the system takes, as input, an un-annotated image, segments it into homogeneous regions, extracts and encodes the visual content of each region by feature vectors. Then, it uses the learned associations or joint probability distributions to infer the set of keywords that best describe the visual features. These keywords are then used to annotate the image.

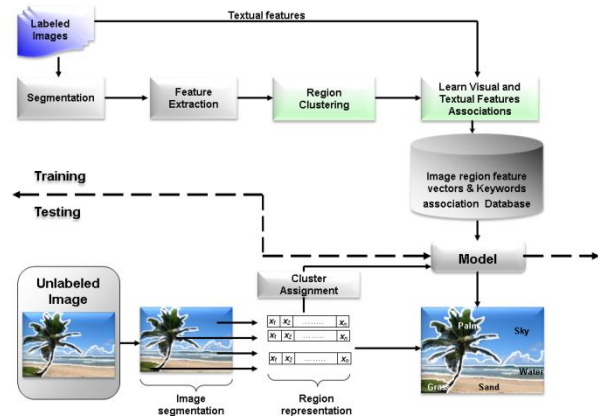


Fig. 2. Overview of a typical automatic image annotation system

Despite the effort made by researchers to propose accurate automatic image annotation approaches, the reported systems show noticeable limitations to label real world images accurately. For instance, the authors in [54] formulated automatic image annotation as a linguistic translation problem with hierarchical text modelling. The approach relies on the assumption that words describing an image represent nodes in a hierarchical concept tree Wordnet [55]. In [56], the researchers extended this approach and used the Wordnet ontology to remove uncorrelated words. In [79], the Latent Dirichlet Allocation (LDA) model was adopted to associate images to textual labels. As one can notice, these approaches encode images as regions, blobs or segments. Thus, images are perceived as bags of words, and joint blob-keyword probabilities are estimated in order to reduce the automatic annotation of images to a likelihood estimation problem. These approaches assume accurate segmentation of the images. Alternatively, Cross Media Relevance Models (CMRM) was proposed in [57, 58] to annotate images automatically. Also, in [59] the authors used the word to word correlations, and proposed coherent language models to enhance image annotation accuracy. The automatic image annotation solutions reported above handle visual features and text modalities separately before modelling their associations. The authors in [60] proposed simultaneous handling of the visual features and the textual keywords. The Probabilistic Latent Semantic analysis (PLSA) is then used to model the resulting uniform vectored data. A variation of this approach, namely the

nonlinear latent semantic analysis was proposed in [61] to annotate images automatically. Another approach consists in formulating automatic image annotation as a classification task where unlabelled images are assigned to a set of predefined concepts such as landscape, city and sunset [62]. The researchers in [63] solved the automatic image annotation problem using a saliency measure based on WordNet and a structure composition modelling. Automatic Linguistic Indexing of Pictures (ALIP) system, introduced in [64], adopts a 2-Dimensional multi-resolution Hidden Markov Models (HMM) to recognize the intra-scale and inter-scale spatial correlations of the visual properties characterizing given semantic classes. For this approach, single classes are first modelled independently. Then, based on the learned class/model, the likelihoods of the query image is calculated and the statistically salient keywords of the most likely classes are chosen for annotation. Similarly, Automatic Linguistic Indexing of Pictures - Real time (ALIPR) system was proposed in [65] as a novel variation of ALIP. ALIPR allows real-time estimation of statistical likelihoods due to its simpler modelling approach. As a pioneer real time automatic image annotation system, ALIPR triggered remarkable interest for real world applications [66]. The authors in [67] outlined concept/class learning using Gaussian mixture models and user feedback when image databases dynamically change over time. In [68], a soft annotation approach based on Bayes point machines to generate confidence function for the predefined semantic keywords. Also, a soft fusion of SVM classifiers was proposed in [64, 69] to overcome automatic annotation challenges. The authors in [49] used Multiple instance learning to automatically categorize images and associate image regions to semantically relevant keywords [70]. The amount and diversity of learning techniques and approaches used to annotate images show how challenging this problem is automatically. Moreover, the image segmentation techniques, which represent a critical component of the proposed system, exhibit considerable limitations to extract the objects and regions in the images accurately. Thus, associating image regions to semantic concepts get more acute. Recently, researchers aimed at bridging the retrieval annotation gap [63] by using keyword queries by default, regardless of label availability with the images.

### B. Multiple Query-Based CBIR

For multiple query-based CBIR, a set of query images is provided by the user to represent his interest. The low-level features are extracted from each one of these query images. Like for typical CBIR system, the visual descriptor extraction is done offline. The key component of multiple query-based CBIR systems consists in the pair-wise distance computation between the query image set and the images in the database. More specifically, rather than computing the distance between the low-level feature vectors corresponding to the unique query image and image from the database, multiple query-based CBIR requires the distance/similarity estimation between a the low-level features representing the query set and a feature vector from the original database [152, 153, 154, 155]. The Multiple query set is intended to be more representative of the user retrieval interest. The authors in [152] presented a CBIR system based on multiple query set. The proposed approach relies on the multi-histogram intersection to measure the distance between the query image set and images in the

database using texture and low-level colour features. The query image set includes images which represent the texture information, and others that reflect the colour information. The similarity of the query image set to images from the database is formulated as a weighted sum of the individual similarities obtained using texture and colour features separately. The authors in [163] introduced a CBIR system based on multiple query images. They formulate the user query using a set of relevant images, and another set of irrelevant images to the user interest. Namely, they used multiple positive sets and multiple negative sets to express the user's semantic. More specifically, the similarity of a query set to images from the dataset is obtained using the similarity of the dataset images with the means of the positive and negative query image sets. In [155], structure, colour and texture descriptors are used to calculate partial distances between images from the query set and database images. Then, relevance weights are associated with these partial distances along with weighted summation to yield individual distances. Finally, the overall distance between an image from the database and the query image set is introduced as the minimum individual distance between each query image and the given database image. One should notice that such approaches suffer from over-fitting. More specifically, the weights associated with the visual descriptors are affected by the dataset content. In other words, weight tuning/learning is required for each image collection. Thus, the relevance weights represent the visual properties of the database images rather than the semantic the user is interested in.

In [156], the authors proposed an approach for optimal query image learning using Mahalanobis distance. Given query images set  $I_Q = \{I_Q^i (i = 1, \dots, M)\}$  and its goodness scores set  $v_i (i = 1, \dots, M)$ , the distance between the query image  $I_Q^i$  and image  $I_D^j$  from the database is formulated as:

$$D(I_Q^i, I_D^j) = (F_Q^i - F_D^j)^T A (F_Q^i - F_D^j) \quad (1)$$

where,  $F_Q^i$  and  $F_D^j$  represent the optimal feature vector of the query image  $I_Q^i$  and image  $I_D^j$  from the database. On the other hand, matrix  $A$  defines the Mahalanobis distance. The learning of the optimal feature vector  $F_Q^i$  and the Mahalanobis matrix  $A$  is achieved through the minimization of the following objective function [166]:

$$\min_{i, F_Q^i} \sum_{j=1}^N v_i (F_Q^i - F_D^j)^T A (F_Q^i - F_D^j) \quad (2)$$

subject to

$$\det(A) = 1 \quad (3)$$

The minimization of this objective function using the Lagrange multiplier [157] yields:

$$F_Q^i = \frac{\sum_{j=1}^N v_j F_D^j}{\sum_{j=1}^N v_j} \quad (4)$$

and

$$A = \det(C)^{\frac{1}{N}} C^{-1} \quad (5)$$

where,  $C$  is the covariance matrix of the feature vectors  $F_D^j$ . The user expresses his interest using query images and their

corresponding goodness scores. One should mention that a large number of query images should be provided to learn accurate Mahalanobis matrix representing the user high level semantic. Moreover, the computation of the Mahalanobis matrix exhibits high time complexity with highly dimensional features.

In [164], the researchers used the Euclidean distance, and assumed that the relationship between an image from the database and a query image set is an AND logical operation to ensure that the retrieved images are similar to all query images. This yields:

$$D(I_Q^1, \dots, I_Q^M, I_D^j) = \max_i (ED(I_Q^i, I_D^j)) \quad (6)$$

where,  $ED(I_Q^i, I_D^j)$  is the Euclidean distance between a database image  $I_D^j$  and the query image  $I_Q^i$ . As it can be seen, this approach does not assign feature weight and consider all features equally relevant. The authors in [158] introduced another multiple query based image retrieval approach using several visual descriptors. The system relies on logic OR distances between the distances from a given query image  $I_Q^i$  to database image  $I_D^j$  using the different features. Besides, it uses a logic AND operator between the distances from a given database image and of the query images. This approach is formulated using the equation below:

$$D(I_Q^1, \dots, I_Q^M, I_D^j) = \max_i (\min_s D_s(I_Q^i, I_D^j)) \quad (7)$$

where,  $D_s(I_Q^i, I_D^j)$  represents the distance between the database image  $I_D^j$  and the query image  $I_Q^i$  obtained using all features. One should notice that rather than assigning feature weights, this approach [158] considers one single feature only, and discards the others. On the other hand, the authors in [159] proposed to linearly combine distances to express the user interest based on the provided query image set and  $s$  set of goodness scores. The proposed approach is formulated as:

$$D(I_Q^1, \dots, I_Q^M, I_D^j) = \sum_{i=1}^M v_i D(I_Q^i, I_D^j)^t \quad (8)$$

where,  $v_i$  expresses the goodness score of the query image  $I_Q^i$  while  $D(I_Q^i, I_D^j)$  represents the distance between the database image  $I_D^j$  and  $I_Q^i$ . Besides,  $t$  is a positive constant larger or equal to 1. The goodness scores  $v_{i=1, \dots, M}$  are provided by the user to express his interest.

As it can be seen, some of the existing multiple query based CBIR approaches do not conduct features relevance weighting. Instead, they consider one of the provided query images as the most representative one, and ignore the other query images [164, 168]. Other approaches [166, 169] require user scoring of the query images to include it in the pair-wise similarity among images. One should notice an important limitation of some state-of-the-art multiple queries based CBIR approaches [162, 163, 165] which are the considerable number of query images required to learn appropriate relevance weights. Furthermore, these relevance weighting relies on cross-validation using particular dataset, and requires a learning process per dataset. This makes the obtained relevance weights reflect the visual

characteristics of the training set rather than the semantic user interest.

### C. Benchmarking

The state-of-the-art proved that no standard benchmark image collection and/or performance measures had been universally used by researchers to evaluate the proposed CBIR systems.

#### 1) Performance Evaluation

Usually the retrieval performance of CBIR systems is assessed using precision and recall. The precision represents the proportion of retrieved images that is relevant to the query. It assesses the capability of the system to find all relevant images. On the other hand, the recall represents the proportion of relevant images that are retrieved by the system. It assesses its capability to find relevant images only. The precision is computed as follows:

$$Precision = \frac{\# \text{ of retrieved relevant images}}{\text{total\#of retrieved images}} \quad (9)$$

Similarly, the recall is calculated as:

$$Recall = \frac{\# \text{ of retrieved relevant images}}{\text{total\#of relevant images}} \quad (10)$$

Researchers aim to achieve high Precision and Recall values. Therefore, rather than assessing the retrieval performance using Precision or Recall individually, the curve Recall Vs Precision has been widely used to evaluate retrieval systems [4]. However, unlike text-based retrieval systems, the retrieval performance for CBIR systems is not accurately reflected by such curve [119]. Thus, the rank (Ra) measure [120, 121] defined as the average rank of the retrieved images, emerged as a promising alternative to overcome this limitation. The smaller the obtained rank value is, the better the achieved performance is. Another performance measure that has been adopted to assess the retrieval performance is the Average Normalised Modified Retrieval Rank (ANMRR) [122]. It includes the order of the retrieved images. The ANMRR values are within the [0, 1] range. If the ANMRR value is close to zero, then the retrieval is highly accurate.

#### 2) Image Databases

Corel image dataset [123] has been most widely used to empirically evaluate the performance of CBIR systems outlined in the surveyed papers. Many researchers believe that Corel image dataset, with its heterogeneous content and the available manual ground truth label represent an appropriate mean to assess CBIR system [130]. However, others perceive Corel image database unsuitable due to the quality of the associated ground truth labels which are often too high-level to be relevant for the retrieval assessment [131, 132]. Thus pre-processing the meta-data associated with Corel images may be a natural alternative to enhance its quality and exploit its high intra-class variance. Thus, in [133], the authors introduced a novel reference data set to evaluate CBIR systems. The proposed data set was collected using real human evaluations of retrieval results. The authors considered 20k evaluations of query result pairs for query by example approach, and 5k pairs for text-based query approach. The resulting data set is assumed to be independent of any specific image retrieval

algorithm. The authors claimed that this data set is sufficient to assess any CBIR related algorithms objectively. Alternatively, researchers used either different digital image collections such as Kodak consumer images [124], LA resource pictures [125] or their own collected images sets. One should mention that specific datasets have also used for particular applications of CBIR models. For instance, Brodatz textures [126] have been adopted to validate applications that rely on perceptual texture descriptor [127, 128, 129]. Also, the Internet represents another alternative data source for Web image retrieval applications [21, 25].

## V. RESEARCH ISSUES

### A. Query Formulation

Query formulation is a key component to reducing the semantic gap between images low-level content and user high level interest. The researchers in [134] introduced OQUEL query language as novel retrieval language. The simple or complex combination of keywords is supported by OQUEL. In [145], a natural query language is proposed to query digital image collections. The language vocabulary consists of elementary semantic indicators such as “tree”, “sea”, etc..., and a syntax that reflects natural patterns perceived by human such as “outdoor scenes” and “people” [135]. The authors in [136] used image regions to express the semantic content the user is looking for by retrieving images of interest in collections including objects and metadata. More specifically, the semantic content is encoded using texture features based on wavelet transform, and the multi-scale colour coherent descriptor. Despite these efforts, the researchers in [134] considered query language as ill-understood and require more focus.

### B. Image Benchmark and Performance Measures

Subsets of Corel image collection, along with precision and recall, are usually used to assess the performance of CBIR systems. However, the researchers in [137] proved that using Corel image subset and these performance measures yield subjective results. In particular, they claim that the obtained results depend on the submitted queries. In their experiments, the authors submitted various query images and relied on different ground truth data. Moreover, they proved that a CBIR system could yield different retrieval results using the same image collection and performance measures. Thus, they concluded that such performance evaluation couldnot be objective without specifying and reporting the test images used to query the system. One can conclude that standard image collection with specified query images, and appropriate performance measures are urgently required for objective CBIR performance evaluation.

## VI. CONCLUSIONS

During the past decade, Content-Based Image Retrieval (CBIR) related research has reserved more attention for digital image processing, visual descriptor extraction, and learning techniques. Advanced researches proved that visual descriptors are unable to capture higher level semantic the user is interested in. In other words, they made CBIR systems fail to bridge the gap human semantics and image low-level content. This work surveyed recent research contribution aiming to

reduce the “semantic gap”. It also outlined the state-of-the-art low-level features adopted to bridge the “semantic gap”. Despite the considerable quantity and quality of work proposed in this area, no standard approach has been defined for image retrieval based on high level semantics. CBIR systems using unsupervised, supervised learning or fusion techniques were proposed to reduce the gap between low-level visual descriptors and the richness of high-level semantic. Moreover, it has been noticed that objective evaluation and comparison of CBIR systems cannot be achieved without standard image dataset availability and unified performance measures. In conclusion, mature content based image retrieval system able to capture high level semantics stands mainly in need of intelligent learning techniques, and appropriate visual descriptor extraction.

## ACKNOWLEDGMENT

This work was supported by the Research Centre of the College of Computer and Information Sciences, King Saud University. The author is grateful for this support.

## REFERENCES

- [1] A. W. Smeulders, M. Worring, Santini, S., Gupta, A., and R. Jain, Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (12) (2000), 1349-1380.
- [2] P. Aigrain, H. Zhang, and D. Petrovic, Content-based representation and retrieval of visual media: A review of the state-of-the-art. *Multimed. Tools Appl.* 3(3) (1996), 179-202.
- [3] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra, Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Trans. Circ. Syst. Video Technol.* 8(5)(1998), 644-655.
- [4] Y. Rui, T. Huang, Optimizing learning in image retrieval. In *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*, 2000.
- [5] Y. Rui, T. Huang, and S. MEHROTRA, Content-based image retrieval with relevance feedback inMars. In *Proc of the IEEE International Conference on Image Processing*, 1997.
- [6] Y. Rui, T. Huang, and S. F. Chang, Image retrieval: Current techniques, promising directions and open issues. *J. Visual Commun. Image Represent.* 10(1)(1999), 3962.
- [7] C. Sonek and M. Worring, Multimodal video indexing: A review of the state-of-the-art. *Multimed. Tools Appl.* 25(1) (2005), 535.
- [8] M. Rehman, M. Iqbal, M. Sharif and M. Raza. Content Based Image Retrieval:Survey, *World Applied Sciences Journal* 19 (3) (2012): 404-412.
- [9] N. Singhai, S K. Shandilya, A Survey On: Content Based Image Retrieval Systems, *International Journal of Computer Applications* 4(2) (2010).
- [10] D. Zhou, J. Weston, A. Gretton, O. Bousquet and B. Scholkopf, Ranking on data manifolds. In *Proc of the Conference on Advances in Neural Information Processing Systems (NIPS)*, 2003.
- [11] C. Bohm, S. Berchtold, and D. A. Keim, Searching in high-dimensional space index structures for improving the performance of multimedia databases. *ACM Comput. Surv.* 33,(3) (2001).
- [12] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, A review of content-based image retrieval systems in medical applications Clinical benefits and future directions. *Int. J. Medical Inf.* 73(1) (2003), 1-23.
- [13] H. Alraqibah, M. M. Ben Ismail and O. Bchir, “Empirical Comparison of Visual Descriptors for Content based X-ray Image Retrieval”, *International Conference on Image and Signal Processing (ICISP'14)*, Cherbourg, June 2014.
- [14] H. Alraqibah, O. Bchir and M. M. Ben Ismail, “X-Ray Image Retrieval System Based on Visual Feature Discrimination”, *Proceeding SPIE*, vol. 9159, *International Conference on Digital Image Processing (ICDIP)*, Athens, April 2014.



- [15] G. Carneiro and N. Vasconcelos, Minimum Bayes error features for visual recognition by sequential feature selection and extraction. In Proc of the Canadian Conference on Computer and Robot Vision, 2005.
- [16] Y. Chen and J. Z. Wang, A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval, IEEE Trans. Pattern Analysis and Machine Intelligence, 24(9) (2002),2521-267.
- [17] T. Gevers and A. W. M. Smeulders, PicToSeek: Combining Color and Shape Invariant Features for Image Retrieval, IEEE TRANSACTIONS ON IMAGE PROCESSING, 9(1) (2000).
- [18] R. Shi, H. Feng, T.-S. Chua, C.-H. Lee, An adaptive image content representation and segmentation approach to automatic image annotation, International Conference on Image and Video Retrieval (CIVR), 2004, pp. 545-554.
- [19] C.P. Town, D. Sinclair, Content-based image retrieval using semantic visual categories, Society for Manufacturing Engineers, Technical Report MV01-211, 2001.
- [20] V. Mezaris, I. Kompatsiaris, M.G. Strintzis, An ontology approach to object-based image retrieval, Proc of the ICIP, vol. II, 2003, pp. 511-514.
- [21] D. Cai, X. He, Z. Li, W.-Y. Ma, J.-R. Wen, Hierarchical clustering of WWW image search results using visual, textual and link information, Proc of the ACM International Conference on Multimedia, 2004.
- [22] D. Cai, X. He, W.-Y. Ma, J.-R. Wen, H. Zhang, Organizing WWW images based on the analysis of page layout and web link structure, Proc of the International Conference on Multimedia and Expo (ICME), Taipei, 2004.
- [23] P.L. Stanchev, D. Green Jr., B. Dimitrov, High level color similarity retrieval, Int. J. Inf. Theories Appl. 10(3) (2003), 363-369.
- [24] I.K. Sethi, I.L. Coman, Mining association rules between low-level image features and high-level concepts, Proc of the SPIE Data Mining and Knowledge Discovery, vol. III, 2001, pp. 279-290.
- [25] H. Feng, T.-S. Chua, A bootstrapping approach to annotating large image collection, Workshop on Multimedia Information Retrieval in ACM Multimedia, 2003, pp. 556-2.
- [26] A. Vailaya, M.A.T. Figueiredo, A.K. Jain, H.J. Zhang, Image classification for content-based indexing, IEEE Trans. Image Process. 10 (1) (2001) 117-130.
- [27] J.R. Smith, C.-S. Li, Decoding image semantics using composite region templates, IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL-98), 1998, pp. 913.
- [28] Y. Zhuang, X. Liu, Y. Pan, Apply semantic template to support content-based image retrieval, Proc of the SPIE, Storage and Retrieval for Media Databases, vol. 3972, 1999, pp. 442-449.
- [29] W. Chang, J. Wang, Metadata for multi-level content-based retrieval, Third IEEE Meta-Data Conference, 1999.
- [30] D. Cai, X. He, Z. Li, W.-Y. Ma, J.-R. Wen, Hierarchical clustering of WWW image search results using visual, textual and link information, Proc of the ACM International Conference on Multimedia, 2004.
- [31] H. Feng, R. Shi, T.-S. Chua, A bootstrapping framework for annotating and retrieving WWW images, Proc of the ACM International Conference on Multimedia, 2004.
- [32] F. Jing, M. Li, L. Zhang, H.-J. Zhang, B. Zhang, Learning in region based image retrieval, Proceedings of the International Conference on Image and Video Retrieval (CIVR2003), 2003, pp. 206-215.
- [33] WANG, X.-J., MA, W.-Y., HE, Q.-C., AND LI, X. 2004. Grouping Web image search result. In Proc of the ACM International Conference on Multimedia.
- [34] GAO, B., LIU, T.-Y., QIN, T., ZHENG, X., CHENG, Q.-S., AND MA, W.-Y. Web image clustering by consistent utilization of visual features and surrounding texts. In Proc of the ACM International Conference on Multimedia, 2005.
- [35] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Comput. Surv. 31 (3) (1999) 264-323.
- [36] C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
- [37] W. Pedrycz, V. Loia, S. Senatore, Fuzzy clustering with viewpoints, IEEE T. Fuzzy Syst. (TFS) 18 (2) (2010) 274-284.
- [38] W. Pedrycz, G. Vukovich, Fuzzy clustering with supervision, Pattern Recognit. 37 (7) (2004) 1339-1349.
- [39] X. Zhang, W.K. Cheung, C.H. Li, Learning latent variable models from distributed and abstracted data, Inf. Sci. Online (2013).
- [40] MCLACHLAN, G. AND PEEL, D. 2000. Finite Mixture Models. Wiley-Interscience.
- [41] DO, M. N. AND VETTERLIM., Wavelet-Based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. IEEE Trans. Image Process. 11(2), (2002) 146-158.
- [42] LI, J. AND WANG, J. Z. Studying digital imagery of ancient paintings by mixtures of stochastic models. IEEE Trans. Image Process. 13(3) (2004), 340-353.
- [43] LI, J. AND WANG, J. Z., Real-time computerized annotation of pictures. IEEE Trans. Pattern Anal. Mach. Intell. 30(6)(2008), 985-1002.
- [44] CHEN, C.-C., WACTLAR, H., WANG, J. Z., AND KIERNAN, K., Digital imagery for significant cultural and historical materials: An emerging research field bridging people, culture, and technologies. Int. J. Digital Libr. 5(4)(2005), 275-286.
- [45] SAUX, B. L. AND BOUJEMAA, N. Unsupervised robust clustering for image database categorization. In Proc of the International IEEE Conference on Pattern Recognition (ICPR), 2002.
- [46] VAILAYA, A., FIGUEIREDO, M. A. T., JAIN, A. K., AND ZHANG, H.-J., Image classification for content-based indexing. IEEE Trans. Image Process 10(1)(2001), 117-130.
- [47] GOH, K.-S., CHANG, E. Y., AND CHENG, K.-T. SVM binary classifier ensembles for image classification. In Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM), 2001.
- [48] PANDA, N. AND CHANG, E. Y. Efficient top-k hyperplane query processing for multimedia information retrieval. In Proc of the ACM International Conference on Multimedia, 2006.
- [49] CHEN, Y. AND WANG, J. Z., Image categorization by learning and reasoning with regions. J. Mach. Learn. Res. (5)(2004), 913-939.
- [50] DATTA, R., GE, W., LI, J., AND WANG, J. Z., Toward bridging the annotation-retrieval gap in image search. IEEE Multimed. 14(3)(2007), 24-35.
- [51] HAUPTMANN, A. G. AND CHRISTEL, M.G., Successful approaches in the TREC video retrieval evaluations. In Proc of the ACM International Conference on Multimedia, 2004.
- [52] JOSHI, D., NAPHADE, M., AND NATSEV, A., A greedy performance driven algorithm for decision fusion learning. In Proc of the IEEE International Conference on Image Processing (ICIP), 2007.
- [53] SMEULDERS, A. W., WORRING, M., SANTINI, S., GUPTA, A., AND JAIN, R., Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. 22(12)(2000), 1349-1380.
- [54] DUYGULU, P., BARNARD, K., DE FREITAS, N., AND FORSYTH, D., Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In Proceedings of the European Conference on Computer Vision (ECCV), 2002.
- [55] MILLER, G., Wordnet: A lexical database for English. Comm. ACM 38(11)(1995), 39-41.
- [56] JIN, Y., KHAN, L., WANG, L., AND AWAD, M., Image annotations by combining multiple evidence and Wordnet. In Proc of the ACM International Conference on Multimedia, 2005.
- [57] JEON, J., LAVRENKO, V., AND MANMATHA, R. Automatic image annotation and retrieval using cross-media relevance models. In Proc of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003.
- [58] M. M. Ben Ismail, and O. Bchir, Automatic Image Annotation based on Semi-supervised Clustering and Membership based Cross Media Relevance Model. International Journal of Pattern Recognition and Artificial Intelligence 26(6)(2012).
- [59] JIN, R., CHAI, J. Y., AND SI, L., Effective automatic image annotation via a coherent language model and active learning. In Proc of the ACM International Conference on Multimedia, 2004.

- [60] MONAY, F. AND GATICA-PEREZ, D., On image auto-annotation with latent space models. In Proceedings of the ACM International Conference on Multimedia, 2003.
- [61] W. LIU, and X. TANG, Learning an image-word embedding for image auto-annotation on the nonlinear latent space. In Proc of the ACM International Conference on Multimedia, 2005.
- [62] VAILAYA, A., FIGUEIREDO, M. A. T., JAIN, A. K., AND ZHANG, H.-J. Image classification for content-based indexing. IEEE Trans. Image Process. 10(1)(2001), 117-130.
- [63] DATTA, R., JOSHI, D., LI, J., and WANG, J. Z., Tagging over time: Real-world image annotation by lightweight meta-learning. In Proc of the ACM International Conference on Multimedia, 2007.
- [64] LI, B., GOH, K.-S., and CHANG, E. Y., Confidence-based dynamic ensemble for image annotation and semantics discovery. In Proc of the ACM International Conference on Multimedia, 2003.
- [65] LI, J. AND WANG, J. Z., Real-Time computerized annotation of pictures. In Proc of the ACM International Conference on Multimedia, 2006.
- [66] ALIPR. Alipr homepage. <http://www.alipr.com>, 2006.
- [67] DONG, A. AND BHANU, B. Active concept learning for image retrieval in dynamic databases. In Proc of the IEEE International Conference on Computer Vision (ICCV), 2003.
- [68] CHANG, E. Y., GOH, K., SYCHAY, G., AND WU, G. CBSA: Content-based soft annotation for multimodal image retrieval using Bayes point machines. IEEE Trans. Circ. Systems Video Technol 13(1), 26-38, 2003.
- [69] NATSEV, A., NAPHADE, M. R., AND TESIC, J., Learning the semantics of multimedia queries and concepts from a small number of examples. In Proc of the ACM International Conference on Multimedia, 2005.
- [70] YANG, C., DONG, M., and FOTOUHI, F. Region based image annotation through multiple- instance learning. In Proc of the ACM International Conference on Multimedia, 2005.
- [71] M. M. Ben Ismail, Image Annotation and Retrieval Based on Multi-Modal Feature Clustering and Similarity Propagation. PhD thesis packaged, produced and published by BiblioLabs under license by ProQuest UMI (available online for sale). 2011.
- [72] MacQueen, J. B.: "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, (1), 281-297, 1967.
- [73] WANG, J., LI, J., AND WIEDERHOLD, G., SIMPLiCity: Semantics-sensitive integrated matching for picture libraries. IEEE Trans. Pattern Anal. Mach. Intell. 23(9), 947-963, 2001.
- [74] WU, H., LU, H., AND MA, S, Willhunter: Interactive image retrieval with multilevel relevance measurement. In Proc of the IEEE International Conference on Pattern Recognition (ICPR), 2004.
- [75] Breiman, L., Bagging predictors Machine Learning, 24(2), 123-140, 1996.
- [76] Shereena V.B. and Julie M. David, CONTENT BASED IMAGE RETRIEVAL: CLASSIFICATION USING NEURAL NETWORKS, The International Journal of Multimedia & Its Applications (IJMA) 6(5)(2014), 1-14.
- [77] De Oliveira, J. E., Araujo A. A., and Deserno T. M., Content-based image retrieval applied to BI-RADS tissue classification in screening mammography, World Journal of Radiology 3(1)(2011), 24-31.
- [78] SHI, J. AND MALIK, J., Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22(8)(2001), 888-905.
- [79] BLEI, D. M. and JORDAN, M. I., Modeling annotated data. In Proc of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003.
- [80] Ben Ismail M. M. and Frigui H., Unsupervised Clustering and Feature Weighting based on Generalized Dirichlet Mixture Modeling". Information Sciences, 274(2014), 35-54.
- [81] Ben Ismail M. M. and Bchir O., Content based Video Categorization using Relational Clustering with Local Scale Parameter, International Journal of Computer Science and Information Technology 8(1)(2016), 27-46.
- [82] Guo J. M.; Prasetyo H.; Chen J. H., Content-Based Image Retrieval Using Error Division Block Truncation Coding Features, IEEE Transactions on Circuits and Systems for Video Technology, 25 (3)(2015), 466-481.
- [83] Thepade S.p D.; Shinde Y. D., Robust CBIR using sectorisation of hybrid wavelet transforms with Cosine-Walsh, Cosine-Kekre, Cosine-Hartley combinations, in International Conf on Pervasive Computing (ICPC), 2015.
- [84] Gupta N.; Das S.; Chakraborti S., Extracting information from a query image, for content based image retrieval," in Eighth International Conference on Advances in Pattern Recognition (ICAPR), 2015.
- [85] Hassekar P.P.; Sawant R.R., "Experimental analysis of perceptual based texture features for image retrieval," in International Conference on Communication, Information & Computing Technology (ICCICT), 2015.
- [86] Walia E., Pal A., Fusion framework for effective color image retrieval, J. Vis. Commun. Image 25, 1335-1348, 2014.
- [87] Dong J., Li X., Liao S., Xu J., Xu D., Du X., Image Retrieval by Cross-Media Relevance Fusion, Proc. of ACM MM, 2015.
- [88] Yang F., Matei B., Davis L. S., Re-ranking by Multi-feature Fusion with Division for Image Retrieval, Conf on Applications of Computer Vision (WACV), 2015.
- [89] L. I. Kuncheva. Combining Pattern Classifiers. Wiley-Interscience, New York, 2004.
- [90] P. Duygulu, K. Barnard, N. de Freitas and D. Forsyth, Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary, in Proc. Seventh European Conf. Computer Vision, 97-112, 2002.
- [91] K.N. Plataniotis, A.N. Venetsanopoulos, Color Image Processing and Applications, Springer, Berlin, 2000.
- [92] P.L. Stanchev, D. Green Jr., B. Dimitrov, High level color similarity retrieval, Int. J. Inf. Theories Appl. 10(3) (2003) 363-369.
- [93] Y. Liu, D.S. Zhang, G. Lu, W.-Y. Ma, Region-based image retrieval with perceptual colors, Proceedings of the Pacific-Rim Multimedia Conference (PCM), 931-938, 2004.
- [94] R. Shi, H. Feng, T.-S. Chua, C.-H. Lee, An adaptive image content representation and segmentation approach to automatic image annotation, International Conference on Image and Video Retrieval (CIVR), 545-554, 2004.
- [95] B.S. Manjunath, et al., Color and texture descriptors, IEEE Trans. CSVT 11 (6) (2001) 703-715.
- [96] F. Jing, M. Li, L. Zhang, H.-J. Zhang, B. Zhang, Learning in regionbased image retrieval, Proceedings of the International Conference on Image and Video Retrieval (CIVR2003), 206-215, 2003.
- [97] C.P. Town, D. Sinclair, Content-based image retrieval using semantic visual categories, Society for Manufacturing Engineers, Technical Report MV01-211, 2001.
- [98] J.Z. Wang, J. Li, D. Chan, G. Wiederhold, Semantics-sensitive retrieval for digital picture libraries, Digital Library Magazine, 5(11), 1999.
- [99] B.S. Manjunath, et al., Introduction to MPEG-7, Wiley, New York, 2002.
- [100] W. Wang, Y. Song, A. Zhang, Semantics retrieval by content and context of image regions, Proceedings of the 15th International Conference on Vision Interface (VI'2002), 17-24, 2002.
- [101] K.N. Plataniotis, et al., Adaptive fuzzy systems for multichannel signal processing, Proc. IEEE 87(9)(1999) 1601-1622.
- [102] R. Lukac, et al., Vector Itering for color imaging, IEEE Signal Process. Mag. (2005) 74-86.
- [103] J.Z. Wang, J. Li, G. Wiederhold, SIMPLiCity: semantics-sensitive integrated matching for picture libraries, IEEE Trans. Pattern Anal. Mach. Intell. 23 (9) (2001), 947-963.
- [104] W.Y. Ma, B. Manjunath, Netra: a toolbox for navigating large image databases, Proceedings of the IEEE International Conference on Image Processing, 568-571, 1997.
- [105] F. Liu, R.W. Picard, Periodicity, directionality, and randomness: wold features for image modeling and retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 18(7)(1996) 722-733.

- [106] H. Tamura, S. Mori, T. Yamawaki, Texture features corresponding to visual perception, *IEEE Trans. Syst. Man Cybern.* 8 (6) (1978) 460-473.
- [107] W.K. Leow, S.Y. Lai, Scale and orientation-invariant texture matching for image retrieval, in: M.K. Pietikainen (Ed.), *Texture Analysis in Machine Vision*, World Scientific, 2000.
- [108] J.Z. Wang, J. Li, G. Wiederhold, SIMPLiCity: semantics-sensitive integrated matching for picture libraries, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (9) (2001) 947-963.
- [109] Y. Liu, X. Zhou, W.Y. Ma, Extraction of texture features from arbitrary-shaped regions for image retrieval, *International Conference on Multimedia and Expo (ICME04)*, Taipei, 2004, 1891-1894.
- [110] P.W. Huang, S.K. Dai, Image retrieval by texture similarity, *Pattern Recognition* 36 (2003) 665-679.
- [111] R. Mehrotra, J.E. Gary, Similar-shape retrieval in shape data management, *IEEE Comput.* 28 (9) (1995) 57-62.
- [112] F. Mokhtarian, S. Abbasi, Shape similarity retrieval under a-ne transforms, *Pattern Recognition* 35 (2002) 31-41.
- [113] Y. Song, W. Wang, A. Zhang, Automatic annotation and retrieval of images, *J. World Wide Web* 6 (2) (2003) 209-231.
- [114] A. Mojsilovic, B. Rogowitz, ISee: perceptual features for image library navigation, *Proceedings of the SPIE, Human Vision and Electronic Imaging*, vol. 4662, (2002), 266-277.
- [115] W.Y. Ma, B. Manjunath, Netra: a toolbox for navigating large image databases, *Proc of the IEEE International Conference on Image Processing*, 1997, pp. 568-571.
- [116] S.K. Chang, Q.Y. Shi, C.W. Yan, Iconic indexing by 2D string, *IEEE Trans. Pattern Anal. Mach. Intell.* 9(3)(1987) 413-428.
- [117] W. Ren, M. Singh, C. Singh, Image retrieval using spatial context, *Ninth International Workshop on Systems, Signals and Image Processing*, Manchester, 2002.
- [118] J.R. Smith, C.-S. Li, Decoding image semantics using composite region templates, *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL-98)*, pp. 913, 1998.
- [119] J. Huang, S. Kuamr, M. Mitra, W.-J. Zhu, R. Zabih, Image indexing using color correlogram, *Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 762-765, 1997.
- [120] R. Zhao, W.I. Grosky, From feature to semantics: some preliminary results, *Proceedings of the International Conference on Multimedia and Expo, 2000*, pp. 679-682.
- [121] R. Zhao, W.I. Grosky, Negotiating the semantic gap: from feature maps to semantic.
- [122] C. Kavitha, A. Krishnan, and K. Sakthivel, Similarity based retrieval of image database: using dynamic clustering, in *Intelligent Sensing and Information Processing*, 2005. *Proc of International Conference on*, 147-151, 2005.
- [123] University of California at Irvine, 20 May 2013. [Online]. Available: <http://archive.ics.uci.edu/ml/datasets/Corel+Image+Features>.
- [124] J. Luo, A. Savakis, Indoor vs outdoor classification of consumer photographs using low-level and semantic features, *International Conference on Image Processing (ICIP)*, vol II, 2001, 745-748.
- [125] Z. Yang, C.-C. Jay Kuo, Learning image similarities and categories from content analysis and relevance feedback, *Proc of the ACM Multimedia Workshops*, 2000, 175-178.
- [126] P. Brodatz, *Textures, A Photographic Album for Artists & Designers*, Dover, New York, NY, 1966.
- [127] W.K. Leow, S.Y. Lai, Scale and orientation-invariant texture matching for image retrieval, in: M.K. Pietikainen (Ed.), *Texture Analysis in Machine Vision*, World Scientific, Singapore, 2000.
- [128] F. Liu, R.W. Picard, Periodicity, directionality, and randomness: wold features for image modeling and retrieval, *IEEE Trans. Pattern Anal. Mach. Intell.* 18(7) (1996) 722-733.
- [129] C.-Y. Chiu, H.-C. Lin, S.-N. Yang, Texture retrieval with linguistic descriptors, *IEEE Pacific Rim Conference on Multimedia*, 2001, 308-315.
- [130] Y. Rui, T.S. Huang, Optimizing learning in image retrieval, *Proc of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2000, 1236-1243.
- [131] X.Y. Jin, CBIR: difficulty, challenge, and opportunity, Microsoft PPT, 2002.
- [132] H. Mueller, S. Marchand-Maillet, T. Pun, The truth about Coreevaluation in image retrieval, *Proceedings of the International Conference on Image and Video Retrieval (ICIVR)*, 2002, pp. 38-49.
- [133] N.V. Shirahatti, K. Barnard, Evaluating image retrieval, *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, vol. 1, 2005, pp. 955-961.
- [134] C.P. Town, D. Sinclair, Language-based querying of image collections on the basis of an extensible ontology, *Int. J. Image Vision Comput.* 22 (3) (2004) 251-267.
- [135] A. Mojsilovic, B. Rogowitz, ISee: perceptual features for image library navigation, *Proceedings of the SPIE, Human Vision and Electronic Imaging*, 4662 (2002), pp. 266277.
- [136] P.H. Lewis, et al., An integrated content and metadata based retrieval system for art, *IEEE Trans. Image Process.* 13 (3) (2004), 302-313.
- [137] H. Mueller, S. Marchand-Maillet, T. Pun, The truth about Coreevaluation in image retrieval, *Proc of the International Conference on Image and Video Retrieval (ICIVR)*, 2002, pp. 38-49.
- [138] Y. Mori, H. Takahashi, & R. Oka. Image-to-word transformation based on dividing and vector quantizing images with words. In *MISRM'99 First International Workshop on Multimedia Intelligent Storage and Retrieval Management*, 1999.
- [139] Jiwoon Jeon and R. Manmatha. Using Maximum Entropy for Automatic Image Annotation. *International conference on image and video retrieval, IRLANDE*, 2004.
- [140] ImageShack, Imageshack, <http://www.imageshack.us/>.
- [141] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, & R. Jain, Content-based image retrieval at the end of the early years, *IEEE Trans. Patt. Analysis Mach. Intell.*, 22(12)(2000).
- [142] Ormager, S.: Image retrieval: Theoretical and empirical user studies on accessing information in images. In: *Proceedings of the 60th American Society of Information Retrieval Annual Meeting*. 34(1997) 202-211.
- [143] A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene, and A. Gelzinis, Soft combination of neural classifiers: A comparative study, *Pattern Recognit. Lett.* 20(1999), 429-444.
- [144] H. Frigui, L. Zhang, P. D. Gader, and D. Ho, Context-dependent fusion for landmine detection with ground penetrating radar, in *Proc of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets, FL, USA, 2007*.
- [145] H. Frigui, P. Gader, and A. C. Ben Abdallah, A generic framework for context-dependent fusion with application to landmine detection, in *SPIE Defense and Security Symposium, Orlando, 2008*.
- [146] K. Woods, J. Kegelmeyer, W. P., and K. Bowyer, Combination of multiple classifiers using local accuracy estimates, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (4) (1997), 405-410.
- [147] L. Kuncheva. Clustering-and-selection model for classifier combination. In *Knowledge-Based Intelligent Engineering Systems and Allied Technologies*, 185-188, 2000.
- [148] R. Liu and B. Yuan, Multiple classifiers combination by clustering and selection, *Information Fusion*, 163168. 2001.
- [149] H. Frigui, L. Zhang, P. D. Gader, and D. Ho., Context-dependent fusion for landmine detection with ground penetrating radar. In *Proc of the SPIE Conference on Detection and Remediation Technologies for Mines and Minelike Targets, FL, USA, 2007*.
- [150] A. C. Ben Abdallah, H. Frigui, P. D. Gader, Adaptive Local Fusion With Fuzzy Integrals. *IEEE T. Fuzzy Systems*, 20(5)(2012), 849-864.
- [151] M. M. Ben Ismail and O. Bchir, Semisupervised Local Fusion Approach for Mine Detection in SONAR Data. *Int. J. Intell. Syst.* 30(11)(2015), 1161-1183.
- [152] J. Tang, S. Acton, An Image Retrieval Algorithm Using multiple query Images., in *Seventh International Symposium on Signal Processing and its Applications*, Paris, 2003.
- [153] M. Nakazato and T. S. Huang, Extending Image Retrieval with Group-Oriented Interface, in *IEEE ICME, Lausanne, Switzerland, 2002*.

- [154] S. Joseph and K. Balakrishnan, Multi Query Image Retrieval System with Application to Mammogram Images, International Journal of Advanced Research in Computer Science, 3(3)(2012), 469-474.
- [155] Q. Iqbal and J. K. Aggarwa, Feature Integration, Multi-image Queries and Relevance Feedback in Image Retrieval, in International Conference on Visual Information Systems, Florida, 2003.
- [156] Y. Ishikawa, R. Subramanya C. Faloutsos, MindReader: Querying databases through multiple examples," in VLDB Conference, New York, 1998.
- [157] M. Hazewinkel, Lagrange multipliers, Encyclopedia of Mathematics, Springer, 1998.
- [158] S. Nepal, M.V. Ramakrishna, MultiFeature Query by Multiple Examples in Image Databases, Advances in data management, McGrawHill Publishing Company Ltd, 2000.
- [159] L. Zhu and A. Zhang, "Supporting multi-example image queries in image databases," in IEEE International Conference on on Multimedia and Expo, New York, 2000.
- [160] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 2001.
- [161] S.D. MacArthur, C.E. Brodley, C.-R. Shyu, Relevance feedback decision trees in content-based image retrieval, Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'00), 68-72, 2000.