Artificial Intelligence Review (2019) 52:1189–1266
https://doi.org/10.1007/s10462-018-9673-8

A survey of image data indexing techniques

Saurabh Sharma[1] · **Vishal Gupta**[1] · **Mamta Juneja**[1]

Published online: 15 December 2018

Abstract

The Index is a data structure which stores data in a suitably abstracted and compressed form to facilitate rapid processing by an application. Multidimensional databases may have a lot of redundant data also. The indexed data, therefore need to be aggregated to decrease the size of the index which further eliminates unnecessary comparisons. Feature-based indexing is found to be quite useful to speed up retrieval, and much has been proposed in this regard in the current era. Hence, there is growing research efforts for developing new indexing techniques for data analysis. In this article, we propose a comprehensive survey of indexing techniques with application and evaluation framework. First, we present a review of articles by categorizing into a hash and non-hash based indexing techniques. A total of 45 techniques has been examined. We discuss advantages and disadvantages of each method that are listed in a tabular form. Then we study evaluation results of hash based indexing techniques on different image datasets followed by evaluation campaigns in multimedia retrieval. In this paper, in all 36 datasets and three evaluation campaigns have been reviewed. The primary aim of this study is to apprise the reader of the significance of different techniques, the dataset used and their respective pros and cons.

Keywords Image retrieval · Hashing · Metric · Indexing · Nearest-neighbor search

1 Introduction

With the explosive growth of multimedia data technologies, it becomes challenging to fulfill diverse user needs related to textual, visual and audio data retrieval. The advances in the integration of computer vision, machine learning, database systems, and information retrieval have enabled the development of advanced information retrieval systems (Gani et al. 2016). As multidimensional databases are gigantic, it has become important to develop data accessing and querying techniques that could facilitate fast similarity search. The

✉ Vishal Gupta
vishal_gupta100@yahoo.co.in

Saurabh Sharma
saurabh.subi@gmail.com

Mamta Juneja
mamtajuneja@pu.ac.in

[1] University Institute of Engineering and Technology, Panjab University, Chandigarh, India

issues of feature extraction and high-dimensional indexing mechanism are crucial in visual information retrieval (VIR) due to the massive amount of data collections. A typical VIR system (Wang et al. 2016) operates in three phases namely feature extraction phase, high-dimensional indexing phase, and retrieval system design phase. Potential applications (Datta et al. 2005, 2008) include digital libraries, commerce, medical, biodiversity, copyright, law enforcement and architectural design. Figure 1, below, displays the block diagram of the query by visual example.

The most important aspect of any indexing technique is to make a quick comparison between the query and object in the multidimensional database (Bohm et al. 2001). Multi-dimensional databases may have a lot of redundant data also. The indexed data, therefore, need to be aggregated to decrease the size of the index which further eliminates unnecessary comparisons.

1.1 Basic concepts

Feature and feature extraction Feature corresponds to the overall description of the image contents. 'Local' and 'global' are the terms used in the context of image features. Shape, color, and texture individually describe contents of an image, but that information is not descriptive enough. In this regard Histograms, SIFT, and CNN based computer vision techniques are developed to extract more informative contents. Feature aggregation techniques like Bag-of-visual-words (BoVW), VLAD, and Fisher vector produces fixed length vector which helps to approximate the performance of similarity metrics.

Index The image index is a data structure which stores data in a suitably abstracted and compressed form to facilitate rapid processing by an application. Feature-based indexing is found to be quite useful to speed up retrieval and is currently needed in this generation. Typically, any information retrieval system demands the following principle requirements (Téllez et al. 2014): size of the index, parallelism, the speed of index generation and speed of search.

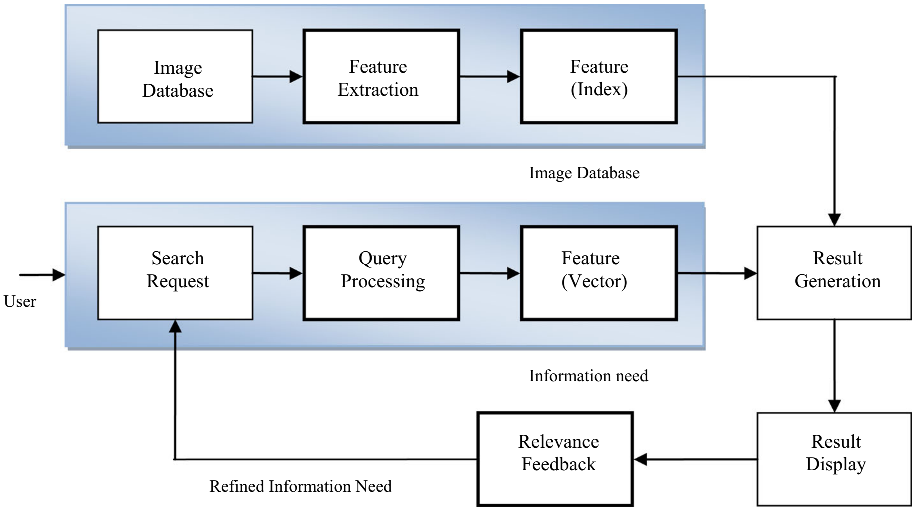


Fig. 1 Overview of visual information retrieval processes

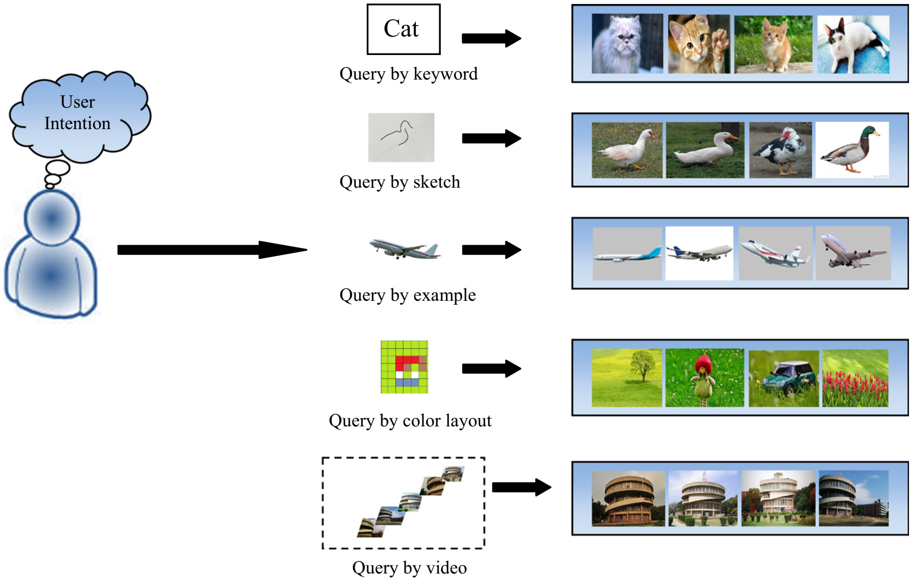


Fig. 2 Illustration of different query schemes

Query processing The retrieval process starts with feature extraction for a query image. The primary aim is to extract and match the corresponding query features with pre-computed image dataset features under issue such as scalability of image descriptors and user intent to search. A query can be processed in a number of ways, depending on the type of indexing and extracted features.

Query formation Query formation is an attempt to define user's precise needs and subjectivity. It is very difficult to capture the human perception and intention into a query. There is different query formation schemes proposed in literature such as query by text, query by image example, query by sketch, query by color layout etc. In Fig. 2 different query formation techniques are presented.

Relevance feedback The different user intent may contain image clarity, quality, and associated meta-data. With the use of earlier user logs and semantic feedback; query refinement and iterative feedback techniques are highly recommended to satisfy the user. The ultimate goal is to optimize the interaction between system and user during a session. Feedback methods may range from short-term techniques, that directly modify the queries to long-term methods, that make the use of query logs.

1.2 Indexing techniques

Indexing phase is one of the most important aspects of any VIR system. Indexing techniques are a powerful means to make a quick comparison between the query and object in the multidimensional database (Bohm et al. 2001). Multidimensional databases may have a lot of redundant data also. The indexed data, therefore, need to be aggregated to decrease the size of the index which further eliminates unnecessary comparisons. Most of the VIR systems are designed for specific types of image such as sports, species, architecture design, art galleries, fashion, etc. Generally, in VIR four kinds of techniques are adopted for indexing: inverted files, hashing based, space partitioning and neighborhood graph construction. To facilitate the indexing and image similarity there is need to pack the visual features. The different techniques of image similarity and indexing are highly affected by the feature extraction/aggregation and query formation as discussed above. For hashing techniques we refer readers to previous survey (Gani et al. 2016; Wang et al. 2016, 2018). In Gani et al. 2016 surveyed indexing techniques in the past 15 years from the perspective of big data. They categorize the indexing techniques in three different categories viz. Artificial Intelligence, Non-artificial Intelligence and Collaborative Artificial Intelligence on the basis of time and space consumption. The main intent of the paper is to identify indexing techniques in cloud computing. Other recent surveys on hashing are reported in Wang et al. (2016, 2018). The concept of learning to hash are delineated there. The survey in Wang et al. (2016) only categorizes and emphasizes on the methodology of data sensitive hashing techniques with big data perspective without any performance analysis. Wang et al. (2018) analyze and compare a comprehensive list of learning to hash algorithms in brief. In this work, they consider similarity preserving and quantization to group the hashing techniques exists in the literature. In addition, they have analyzed the query performance of pair-wise, multi-wise and quantization techniques for limited datasets. The recently conducted survey has been focused only on hash based indexing techniques for which there are only a limited support for experimental analysis and applications.

Therefore, we found that there is a need of survey article which has to cover detail of applications, datasets, performance analysis, evaluation challenges and non-hash based indexing techniques as well. In this regard, we analyze and compare a comprehensive list of indexing techniques. Here, we are focusing on overview of hash and non-hash based indexing techniques of the recent years. We provide more categorical detail of indexing techniques in Sect. 4. In comparison to earlier surveys (Gani et al. 2016; Wang et al. 2016, 2018), the contributions of our article are as follows:

1. It provides splendid image retrieval application areas.
2. It provides an immense description and categorization of hash and non-hash based indexing techniques.
3. It extensively discussed and listed 36 different datasets in detail.
4. It presents a separate evaluation section to examine the performance of different hashing techniques for different datasets.
5. It presents details of multimedia evaluation programs covering top conferences related to indexing techniques.

1.3 Organization of the article

The rest of the article has been organized as follows: Sect. 2 describes the significant challenges in VIR system; Sect. 3, presents promising application areas followed by Sect. 4, which categorizes the indexing techniques. Section 5 provides an evaluation framework for different hash based indexing techniques followed by Sect. 6, that presents some evaluation campaign organized for multimedia indexing. Section 7 addresses the future work, and in the last section, we draw a conclusion.

2 Major challenges in VIR systems

2.1 Similarity search

Content-based search extends our capability to explore/search the visual data in different domains. This operation relies on the notion of similarity for search e.g. to search for images with content similar to a query image (Kurasawa et al. 2010). Translating the similarity search into the nearest neighbor (NN) (Uysal et al. 2015), search problem finds many applications for information retrieval, machine learning, and data mining. The context of large-scale unstructured data envisages finding approximate solutions. Approximate similarity (Pedreira and Brisaboa 2007; Hjaltason and Samet 2003) search relaxes the search constraints to get acceptable results at a lower cost (e.g. computation, memory, time).

2.2 Curse of dimensionality

All the research works have a common concern of scaling up indexing from low dimensional feature space to high dimensional feature space in getting good results, and it is a significant problem due to the phenomenon so called "the curse of dimensionality" (Wang et al. 1998). Recent studies show that most of the indexing schemes even become less efficient than sequential indexing for high dimensions. Such degradations and shortcomings prevent a widespread usage of such indexing structures, especially on multimedia collections.

2.3 Semantic gap

The field of semantic based image retrieval first received active research interest in the late 2000s (Zhang and Rui 2013). Both the single feature and the combination of multiple features are lacking in capturing the high-level concept of images. It is essential to understand the discrepancy between low-level image features and high-level concept to design good applications for VIR. The disparity leads to the so-called semantic gap (Sharif et al. 2018) in the VIR context. Describing images in semantic terms is the highest level of visual information retrieval, and it is a challenging task (Wang et al. 2016).

3 Applications

Indexing is widely used in visual information retrieval systems to make fast offline and online comparisons among data items. With the increase of storage devices as well as progress on the internet, image retrieval is growing with diverse application domains. In the literature a very few fully operational VIR systems are available but the importance of image retrieval has been highlighted in many fields. Though these certainly represent only the tip of the mountain, some potentially productive areas at the end of 2017 are as follows:

a. **Medical applications** A rapid evolution in diagnostic techniques results in a large archives of medical images. At present this area is largely publicized as the prime users of VIR systems. This area has great potentials to be developed as huge markets for VIR system as it has unique ingredients (feature set viz. shape, texture etc.) for feature selection and indexing. The use of VIR can result in valuable services that can benefit biomedical information systems. The retrieval, monitoring, and decision-making should be integrated seamlessly to design an efficient medical information system for radiologist, cardiologist, and others.

b. **Biodiversity information systems** Researchers in the life sciences are becoming increasingly concerned about to detect various diseases related to agricultural plants and to understand habitats of species. The in-time gathering and monitoring of visual data consistently achieve objectives as well as minimize the effect of diseases in plants/animals and monitors the lack of nutrients in plants.

c. **Remote sensing applications** VIR system can be used to retrieve images related to fire detection, flood detection, land sliding, rainfall observation in agriculture, etc. For the query "show all forest area having less rainfall in last ten years" system replies with images having a region of interest. From military applications point of view probably this area is well developed and less publicized.

d. **Trademarks and copyrights** This is one of the mature areas and on the advanced stage of development. In recent years, illegal use of logos and trademarks of noted brands has been emerged for business benefits. VIR is used as a counter mechanism in the identification of duplicate/similar trademark symbols which further helps in law enforcement and copyright violation investigation.

e. **Criminal investigation** As an application this is not a truly VIR system as it purely supports identity matching rather than similarity matching. The VIR systems have a big significance in the criminal investigation. The identification of mugshot images, tattoos, fingerprint and shoeprint can be supported by these systems. Practically a large number of systems are used throughout the world for criminal investigation.

f. **Architectural and interior design** Images that visually model the inner and outer structure of a building are containing more diagrammatic information. The use of VIR can result in important services that can benefit interior design or decorating and floor plan of a building.

g. **Fashion and fabric design** The fashion and fabric industry have a predominant position among other industries all over world. For the product development purpose, designer of cloths has to refer previous designs. For the online shopping purpose, the user has to retrieve similar product options. As an application the aim of VIR system is to search the similar fabrics and products for designers and buyers respectively.

h. **Cultural heritage** In comparison to other areas, image retrieval in art galleries and museum highly depends upon the creativity of user as images have heterogeneous specifics. In digitized art gallery and museum, the feature set is of high dimensionality which in turn requires advanced VIR systems.

i. **Education and manufacturing** The main paradigm for performing 3D model retrieval has been using query-by example and query-by-sketch approach. The 3D image retrieval can be seen, as a toolbox for computer aided design, video game industry, teaching material and different manufacturing industries.

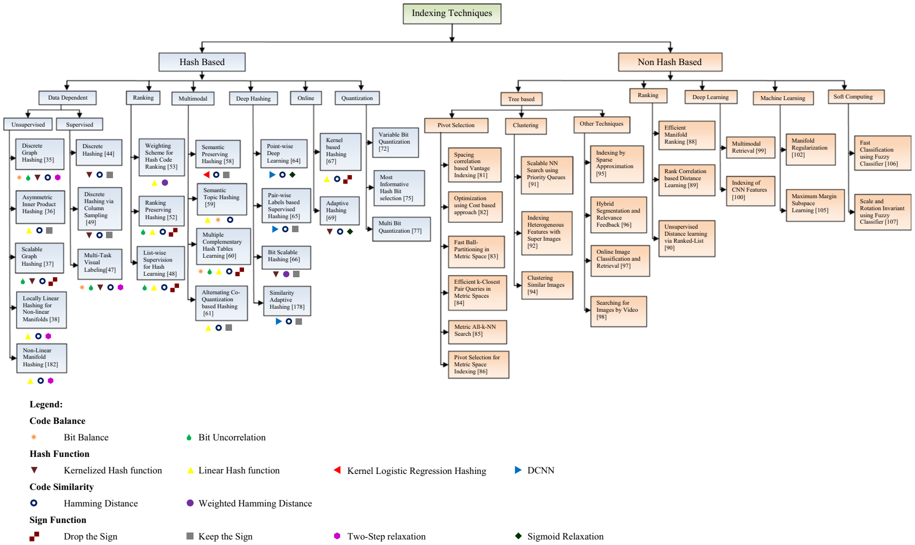
Other examples of database applications where visual information retrieval and indexing is useful: Personal Archives, Scientific Databases, Journalism and advertising, Storage of fragile documents, Biometric identification and Sketch-based Information Retrieval.

4 Categorization of indexing techniques

This section provides a background on indexing techniques and how they facilitate visual information retrieval and visual query by example. Many of the existing indexing techniques may range from the simple tree based (Robinson 1981; Lazaridis and Mehrotra 2001; Uhlmann 1991; Baeza-Yates et al. 1994) approaches to complex approaches that include deep learning (Babenko et al. 2014; Donahue et al. 2014; Dosovitskiy et al. 2014; Fischer et al. 2014) and hashing based (Andoni and Indyk 2008; Baluja and Covell 2008; He et al. 2011; Zhuang et al. 2011; Mu and Yan 2010; Liu et al. 2012). By approximate nearest neighbor search there exist hash and non-hash based indexing techniques and methodology of both turns around various concepts in the literature. But our finding says it is limited to some quality concepts. The hash based techniques are basically turnaround these concepts: Graph based, Matrix Factorization, Column Sampling, Weight Ranking, Rank Preserving, List-wise Ranking, Quantization, Semantic similarity (Text/image), Bit Scalability, Variable bit etc. whereas the non hash based techniques contains concepts namely Pivot Selection, Ball Partitioning, Pruning Rules, Semantic Similarity, Queue based Clustering, Manifold Ranking, Hybrid Segmentation, Approximation etc. Out of these significant concepts related methods have been detailed in this section. The categorization of these indexing techniques is presented in Fig. 3.

4.1 Hash based indexing

The Hashing has its origins in different fields including computer vision, graphics, and computational geometry. It was first introduced as locality sensitive hashing (Datar et al. 2004) in an approximate nearest neighbor (ANN) (Muja and Lowe 2009) search context. Any hash based ANN search works in three basic steps: figuring out the hash function, indexing the database objects and querying with hashing. Most ordinary hash functions are of the form



Indexing Techniques
Hash Based
Non Hash Based
Data Dependent
Ranking
Multimodal
Deep Hashing
Online
Quantization
Unsupervised
Supervised
Discrete Graph Hashing [35]
Asymmetric Inner Product Hashing [36]
Scalable Graph Hashing [37]
Locally Linear Hashing for Non-linear Manifolds [38]
Non-Linear Manifold Hashing [182]
Discrete Hashing [44]
Discrete Hashing via Column Sampling [49]
Multi-Task Visual Labeling[47]
Weighting Scheme for Hash Code Ranking [53]
Ranking Preserving Hashing [52]
List-wise Supervision for Hash Learning [48]
Semantic Preserving Hashing [58]
Semantic Topic Hashing [59]
Multiple Complementary Hash Tables Learning [60]
Alternating Co-Quantization based Hashing [61]
Point-wise Deep Learning [64]
Pair-wise Labels based Supervised Hashing [65]
Bit Scalable Hashing [66]
Similarity Adaptive Hashing [178]
Kernel based Hashing [67]
Adaptive Hashing [69]
Variable Bit Quantization [72]
Most Informative Hash Bit selection [75]
Multi Bit Quantization [77]
Tree based
Ranking
Deep Learning
Machine Learning
Soft Computing
Pivot Selection
Clustering
Other Techniques
Spacing correlation based Vantage Indexing [81]
Optimization using Cost based approach [82]
Fast Ball-Partitioning in Metric Space [83]
Efficient k-Closest Pair Queries in Metric Spaces [84]
Metric All-k-NN Search [85]
Pivot Selection for Metric Space Indexing [86]
Scalable NN Search using Priority Queues [91]
Indexing Heterogeneous Features with Super Images [92]
Clustering Similar Images [94]
Indexing by Sparse Approximation [95]
Hybrid Segmentation and Relevance Feedback [96]
Online Image Classification and Retrieval [97]
Searching for Images by Video [98]
Efficient Manifold Ranking [88]
Rank Correlation based Distance Learning [89]
Unsupervised Distance learning via Ranked-List [90]
Multimodal Retrieval [99]
Indexing of CNN Features [100]
Manifold Regularization [102]
Maximum Margin Subspace Learning [105]
Fast Classification using Fuzzy Classifier [106]
Scale and Rotation Invariant using Fuzzy Classifier [107]
Legend:
Code Balance
Bit Balance
Bit Uncorrelation
Hash Function
Kernelized Hash function
Linear Hash function
Kernel Logistic Regression Hashing
DCNN
Code Similarity
Hamming Distance
Weighted Hamming Distance
Sign Function
Drop the Sign
Keep the Sign
Two-Step relaxation
Sigmoid Relaxation

Fig. 3 Categorization of Indexing Techniques

$$h_k = \mathrm{sgn}\left(f\left(w_k^T x + b_k\right)\right)$$

here $f()$ is nonlinear function. The projection vector w and the corresponding bias b are estimated during the training procedure. $sgn()$ is the element wise function which returns 1 if element is positive number and return − 1 otherwise. In addition, the choice of $f()$ varies with type of hashing under consideration. Further, the choice of Hashing technique is highly dependent and mostly effected by the following factors: type of hash function, hash code balancing, similarity measurement and code optimization. The hash functions are of following forms: Linear, Non-linear, Kernelized, Deep CNN. The concept of code balancing is transfigured to bit balance and bit uncorrelation. Bit $\mathrm{b_i}$ is balanced iff it is set to 1 in one half of all hash codes. Bit uncorrelation means that all pair of bits of hash code B is uncorrelated. Hash codes similarity is measured by hamming distance and its variants. Further, in optimization process sign function keeping, dropping and other relaxation methods are available. All these factors are denoted by legend in Fig. 3. To improve the search performance by fast hash function learning, researchers come up with new hashing methods of different flavors:

4.1.1 Data-dependent hashing (unsupervised)

The design of hash functions subject to analysis of available data and to integrate different properties of data. The main aim is to learn features from the particular dataset and to preserve similarity among the various spaces viz. data space and Hamming space. The unsupervised data dependent method uses unlabeled data to learn hash codes and committed to maintaining Euclidean similarity between the samples of training data. A representative method includes Kernelized LSH (Kulis and Grauman 2012), Spectral Hashing (Weiss et al. 2008), Spherical Hashing (Heo et al. 2012) and much more. Some of these techniques are discussed below.

Discrete graph hashing Liu et al. (2014) proposed a mechanism for nearest neighbor search. To introduce a graph- based hashing approach, the author uses anchor graph to capture neighborhood structure in a dataset. The anchor graph provides nonlinear data-to-anchor mapping, and they are easy to build in linear time which is directly proportional to a number of data points. The proposed discrete graph hashing is an asymmetric hashing approach as it has different ways to deal with queries and in and out of sample data points (Table 1). The objective function written in matrix form is:

$$\min_B \frac{1}{2} \sum_{i,j=1}^{n} \left\| b_i - b_j \right\|^2 S_{ij} = tr\left(B^T E\ B\right)$$

Define graph Laplacian L as anchor graph Laplacian $L = I_n - S$, then the objective function is rewritten as:

$$\max_B tr\left(B^T S B\right)$$

By softening the constraints and ignoring error prone relaxation the objective function can be transformed as

$$\max_{B,Y} Q(B, E) = tr\left(B^T S B\right) + \rho tr\left(B^T E\right)$$

$$s.t.\ E \in \mathbb{R}^{n\times c},\ 1^T E = 0,\ E^T E = nI_c$$

Table 1 Summary of Notations

Notation	Meaning
N	Number of data points
D	Dimensionality of data points
C	Number of hash bits or code length
M	Number of kernel bases
$W \in \mathbb{R}^{nxn}$	Weight matrix
w_k	Weight vector
$w_{ij} \in \mathbb{R}^{d}$	The *i j*th element of weight matrix
$x_i, x_j \in \mathbb{R}^{d}$	The *i*th and *j*th data point
$X = [x_1,\dots, x_n] \in \mathbb{R}^{dxn}$ and $Y = [y_1, \dots, y_n] \in \mathbb{R}^{dxn}$	Two set of data points as column
$S \in \mathbb{R}^{nxn}$	Similarity matrix of data points
$\hat{S} \in \mathbb{R}^{nxn}$	Sampled sub-matrix of similarity
$S_{ij} = \mathrm{sim}\,(x_i, x_j)$	Similarity between data points x_i and x_j
$b_i, b_j \in \{1,-1\}^c$ or $\{0,1\}^c$	Binary hash codes of data points x_i and x_j
$B = [b_1,\dots, b_n]^T \in \{1, -1\}^{nxc}$	Hash codes of data X
$q_i \in \mathbb{R}^{d}$	A query point
$Q = \{q_i\}_{i=1}^{M}$	A query set
$H = [h_1, \dots, h_c] : \mathbb{R}^{d} \rightarrow \{-1,1\}^c$	Hash function for binary values
$\widehat{H} = [h_1, \dots, h_c] : \mathbb{R}^{d} \rightarrow \{1,2,3,\dots\}^c$	Hash function for real values
$h_k : \mathbb{R}^{d} \rightarrow \{-1,1\}$	*k*th hash function
$r_i^q, r_j^q \in \{1,2,3,\dots,n\}$	Ranking list elements
$\|\cdot\|_F$, $\|\cdot\|_2$	Matrix Frobenius and ℓ_2-norm
$\mathrm{tr}(\cdot)$	Matrix trace norm

Asymmetric inner product hashing Fumin Shen et al. (Shen et al. 2017) address the asymmetric inner product hashing to learn binary code efficiently. In another context, it maintains inner product similarity of feature vectors. With the help of asymmetric hash function the Maximum Inner Product Search (MIPS) problem is formulated as

$$\min_{h,z}\left\| h(Y)^T z(X) - S \right\|_F^2$$

Here $h(\cdot)$ and $z(\cdot)$ are the hash functions, $\cdot$ is the Frobenius norm and S is the similarity matrix computed as $S = A^T X$. Further with linear form of hash functions the proposed Asymmetric Inner-product Binary Coding problem is formulated as (in matrix form):

$$\min_{W,R}\left\| sgn\left(W^T Y\right)^T sgn\left(R^T X\right) - S \right\|_F^2$$
$$s.t.\ h(x) = sgn\left(W^T X\right), z(x) = sgn\left(R^T X\right)\ and\ W, R \in \mathbb{R}^{d\times c}$$

The author incorporates the discrete variable as a substitute for sign function to optimize the bit generation approach.

Scalable graph hashing Jiang and Li (2015) get inspiration from asymmetric LSH (Datar et al. 2004) to propose an unsupervised graph hashing. The proposed scheme formulates the approximation of the whole graph through feature transformation. Here approximation of graph is made without computing pair-wise similarity graph matrix. It learns hash functions bit by bit. The objective function is:

$$\min_{\{b_l\}_{l=1}^{n}} \sum_{i,j=1}^{n} \left(\tilde{S}_{i,j} - \frac{1}{2} b_i^T b_j \right)^2$$

$$s.t.\ \tilde{S}_{i,j} = 2S_{i,j} - 1 \text{ and } \tilde{S}_{i,j} \in (-1, 1]$$

In particular, it uses concept of kernel bases to learn hash function and the objective function is defined as (in matrix form):

$$\min_{W} \left\| c\ \tilde{S} - \text{sgn}\left(K(X)W^T\right)\text{sgn}\left(K(X)W^T\right)^T \right\|_F^2$$

Subject to: $K(X) \in \mathbb{R}^{n\times m}$ is the kernel feature matrix.

Locally linear hashing to capture non-linear manifolds Irie et al. (2014) suggest a locally linear hashing to obtain the manifolds structure concealed in visual data. To identify the nearest neighbor in the same manifold related to the query, a local structure preserving scheme is proposed. In particular, it uses Locally Linear Sparse Reconstruction to capture locally linear structure:

$$\min_{w_t} \lambda \left\| S_i^T w_i \right\| + \frac{1}{2} \left\| x_i - \sum_{j \in N_E(x_i)} w_{ij} x_j \right\|_F^2$$

$$s.t.\ N_E(\text{x}) \text{ be a set of nearest neighbours of x}$$

The proposed model maintains the linear structure in a Hamming space by simultaneously minimizing the errors and losses due to reconstruction weights and quantization respectively. Therefore, the optimization problem is defined as:

$$\min_{B, R \in \mathbb{R}^{c\times c} Z \in \mathbb{R}^{n\times c}} tr\left(Z^T M Z\right) + \eta ||B - ZR||_F^2$$

$$s.t.\ R^T R = I_c,\ Z \text{ is continuous embedding and R isan orthogonal transformation that rotates Z.}$$

Non-linear manifold hashing to learn compact binary embeddings Shen et al. (2015) address the embedding to learn Nonlinear Manoifolds efficiently. The proposed method is entirely unsupervised which is used to uncover the multiple structures obscured in image data via Linear embedding and t-SNE (t-distributed stochastic neighbor embedding). For the construction of embedding a prototype algorithm has been proposed. For a given data point x_q, the embedding y_q can be generated as

$$C\left(y_q\right) = \sum_{i=1}^{n} w\left(x_q, x_i\right) ||y_q - y_i||^2$$

$$s.t. w\left(x_q, x_i\right) = \exp\left(-\frac{||x_q - x_i||^2}{\sigma^2} \right), \quad if\ x_i \in N_k\left(x_q\right) otherwise\ it\ is\ 0.$$

$$Here\ N_k\left(x_q\right)\ be\ a\ set\ of\ nearest\ neighbours\ of\ x_q.$$

4.1.2 Data-dependent hashing (supervised)

Another category of learning to hash technique is supervised hashing. The supervised data dependent method uses labeled data to learn hash codes and committed to maintaining semantic similarity constructed from semantic labels of training samples. In comparison to unsupervised methods, supervised methods are slower during learning of large hash codes and labeled data. Further, it is limited to applications as it is not possible to get semantic labels always. The level of supervision further categorizes supervised methods in point-wise, triplets and list-wise approaches. Representative method includes Supervised Discrete hashing (Shen 2015), Minimal loss hashing (Norouzi and Fleet 2011) and many more (Lin et al. 2013; Ding et al. 2015; Ge et al. 2014; Neyshabur et al. 2013). Some of these techniques are listed and discussed below.

Discrete hashing Shen et al. (2015) proposed a new learning-based data dependent hashing framework. The main aim is to optimize the binary codes for linear classification. This method jointly learns bit embedding and linear classifier under the optimization. For the optimization of hash bits, the discrete cyclic coordinate descent (DCC) algorithm is proposed. The objective function is defined as (in matrix form):

$$\min_{\mathrm{B,W,F}} \left\| \mathrm{Y} - \mathrm{W}^{\mathrm{T}}\mathrm{B} \right\|^2 + \lambda \| W \|^2 + v \| B - H(X) \|_2^2$$

here λ and v are the regularization parameter and penalty term respectively. They use hinge loss and l_2 loss for linear classifier. The proposed method showed improvement in results when it is compared with state-of-the-art supervised hashing methods (Zhang et al. 2014; Lin et al. 2014; Deng et al. 2015; Wang et al. 2013) and addressed their limitations. Further, in (Shen et al. 2016) they presented fast optimization of binary codes for linear classification and retrieval. Supervised and Unsupervised losses are considered for the development of scalable and computationally efficient method.

Discrete hashing by applying column sampling approach Kang et al. (2016) presented discrete supervised hashing method based on column sampling for learning hash codes generated from semantic data. The proposed scheme is an iterative scheme where sampling of similarity matrix columns has been done through a column sampling technique (Li et al. 2013) i.e. the technique sample current available training data point into a single iteration. A randomized approach is used to sample data points i.e. few of the possible data points are selected for random sampling. Further, it partitions the sample space into two unequal halves and objective function is formulated as:

$$\min_{B^{\Omega} B^{\Gamma}} \left\| c\tilde{S}^{\Gamma} - B^{\Gamma}\left[B^{\Omega}\right]^{T} \right\|^2 + \left\| c\tilde{S}^{\Omega} - B^{\Omega}\left[B^{\Omega}\right]^{T} \right\|_F^2$$

$s.t.\ \tilde{S} \in \{-1, 1\}^{\Gamma \times \Omega}$, Ω and Γ are two halves of sample space, $\Gamma = N - \Omega$ with N = {1, 2,… n}.

It is a multi-iterative approach where the sample spaces updated with each iteration on an alternate basis.

Adaptive relevance based multi-task visual labeling Deng et al. (2015) developed an image classification approach to overcome issues like data scarcity and scalability of learning technique. A use of hashing based feature dimension reduction reported much better image classification and stepped down required storage. The proposed method is a two-pronged multi-task hashing learning. Firstly, in learning step, each task suggests learning the defined model for a particular label. Further, this learning step executed in two levels viz. tasks and features simultaneously. The idea suggests that the complicated structure of processed features efficiently handle by task relevance scheme. Secondly, in the prediction step test datasets and trained model simultaneously classify and predicts the multiple labels. Outcomes reveal the algorithm potential of enhancing the quality of classification across multiple modalities.

4.1.3 Ranking-based hashing

With the advent of supervised, unsupervised and semi-supervised learning algorithms it is easy to generate optimized compact hash codes. Nearest neighbor search in large dataset under data-dependent methods produces suboptimal results. By exploring the ranking order and accuracy, it is easy to evaluate the quality of hash codes. Associated relevant values of hash codes help to maintain the ranking order of search results. A representative method includes Ranking-based Supervised Hashing, Column Generation Hashing and Rank Preserving Hashing and much more. Some of these methods are listed and discussed below.

Weighting scheme for hash code ranking Ji et al. (2014) presented a ranking method involving weighting system. They aim to make an improved hamming distance ranking. However hamming distance ranking loses some valuable information during quantization of hash functions. To get highly efficient hamming distance measurement the proposed scheme learns weights [Qsrank (Zhang et al. 2012) and Whrank (Zhang et al. 2013)] during similarity search. To cope with expensive computing of higher order independence among hash function, this method uses mutual information of hash bit to propose mutual independence among hash bits. The neighbour preservation weighting scheme is defined as:

$$w_k = exp\left[\gamma \sum_{p \in NN(q)} s(q,p) h_k(q) h_k(p)\right]$$

Here p and q are two weight vectors to capture the shared structure among task parameters, and variations specific to each task respectively. The objective function is maximized as follows:

$$\max_{\pi} w_i^* w_j^* a_{ij}$$

s.t. $1^T\pi = 1, \pi \geq 0, \gamma \geq 0$, a_{ij} is mutual independence between bit variables and $w_i^* = w_k \pi_k$.

The anchor graph is used to represent sample to make similarity measurement useful for various datasets.

Ranking preserving hashing Wang et al. (2015) proposed ranking preserving hashing to improve the ranking accuracy named Normalized Discounted Cumulative Gain (NDCG) which is the quality measure for hashing codes. The main aim is to learn new devised hashing codes that can maintain the ranking order and relevance values both for data examples. The ranking accuracy is calculated as follows:

$$NDCG = \frac{1}{Z} \sum_{i=1}^{n} \frac{2^{r_i} - 1}{\log(1 + \pi(x_i))}$$

here ranking position is defined as , $\pi(x_i) = 1 + \sum_{k=1}^{n} I(b_q^T(b_k - b_i) > 0)$, Z is the normalization factor and r is the relevance value of data item x. It's hard to optimize NDCG directly due to its dependency on the ranking of data standards. Optimization of NDCG is done through linear hashing function which evaluates the expectation of NDCG.

Use of list-wise supervision for hash codes learning Wang et al. (2013) presented an interesting variant for learning hash function. To this end, ranking order is used for learning procedure. The proposed approach implemented in three steps: (a) Firstly, transforms Ranking lists of queries into triplet matrix. (b) Secondly, the inner product has been used to compute the hash codes similarity which further derives the rank triplet matrix in Hamming space. (c) Finally, triplets are set to minimize inconsistency. The formulation of Listwise supervision is based on ranking triplet defines as:

$$S\big(q_m; x_i, x_j\big) = \begin{cases} 1 & : r_i^q < r_j^q \\ -1 & : r_i^q > r_j^q \\ 0 & : r_i^q = r_j^q \end{cases}$$

The objective is to measure the quality of ranking list, which is formally calculated through loss function written in matrix form:

$$\begin{aligned} L_H &= -\sum_m \sum_{i,j} q_m^T C C^T \big[x_i - x_j\big] S_{mij} \\ &= \sum_m q_m^T C C^T p_m = -tr\Big(C C^T G\Big) \end{aligned}$$

s.t. $C \in \mathbb{R}^{\mathrm{dxk}}$ is the coefficient matrix, $p_m = \sum_{i,j}\big[x_i - x_j\big] S_{mij}$ and $\mathrm{G} = \sum_m p_m q_m^T$.

The measurement of Loss function differentiates two ranking lists. They used Augmented Lagrange Multipliers (ALM) for optimization to reduce the computation time.

4.1.4 Multi-modal hashing

Developing a new retrieval model that is focusing on different types of multimedia data is a challenging task. Cross view or multimodal hashing techniques map different high dimensional data modalities into a small dimensional Hamming space. The main issue in performing joint modality hashing is preserving of similarity among inter and intra modalities. Utilization of information further categorize these methods in real-valued (Rasiwasia et al. 2010) and binary representation learning (Song et al. 2013) approaches. Some of the representative techniques are listed and discussed below.

Semantic-preserving hashing Lin et al. (2015) proposed semantic preserving hashing named SPH for multimodal retrieval. The proposed scheme formulate the procedure in two steps: Firstly, it transforms semantic labels of data into a probability distribution. Secondly, it approximates data in Hamming space by minimizing its Kullback-Leiber divergence. The objective function is written as follow:

$$\Psi = \min_{\widehat{H \in \mathbb{R}^{n \times c}}} \sum_{i \neq j} p_{i,j} \log \frac{p_{i,j}}{q_{i,j}} + \frac{\alpha}{No} \big\| \, |\widehat{H}| - I \big\|_2^2$$

s.t. α is the parameter for balancing the KL divergence and No is the normalizing factor

$$\text{s.t. } q_{i,j} = \frac{\left(1+\frac{1}{4}\left\|\widehat{H}_{i,}\cdot-\widehat{H}_{j,\cdot}\right\|^{2}\right)^{-1}}{\sum_{k\neq m}\left(1+\frac{1}{4}\left\|\widehat{H}_{k,\cdot}-\widehat{H}_{m,\cdot}\right\|^{2}\right)^{-1}} \text{ and } p_{i,j} = \frac{A_{i,j}}{\sum_{i\neq j}A_{i,j}}$$

here the probabilities $p_{i,j}$ and $q_{i,j}$ are the probability of observing the similarity among training data and similarity among data items in hamming space respectively. I is the matrix having all entries 1. Quantization loss is measured as $\left|\widehat{H}\right| - I^2$. They used kernel logistic regression boosted by sampling technique to introduce projection, the regression is done through k-mean and random sampling similarly.

Semantic topic hashing via latent semantic information Wang et al. (2015) addresses the issues with graph-based and matrix decomposition based multimodal hashing methods. The long training time, decrease in mapping quality, and large quantization errors are the generic drawbacks of above-mentioned techniques. In the proposed work, the discrete nature of hash code has been considered. The overall objective function for text and image concept (L_T, L_I, L_C) is defined as:

$$\min_{\mathbf{F},\mathbf{H},\mathbf{U},\mathbf{V},\mathbf{P}} L = \lambda L_T + (1-\lambda)L_I + \mu L_C + \gamma R(\mathrm{F},\mathrm{U},\mathrm{V},\mathrm{P})$$

$$\text{s.t. } h_{ij} \in \{0,1\}, \sum_{i=1}^{c} h_{ij} = const$$

Here, F is a set of latent semantic topic, P is the correlation matrix between text and image, U and V are the set of semantic concepts, λ, μ, γ are the tradeoff parameters and R$(\cdot)$ is the regularization term to avoid over fitting.

Multiple complementary hash tables learning Liu et al. (2015) switches research direction from compact hash codes to multiple additional hash tables. The author claims that this is the first approach which takes into account the multi-view complementary hash tables. In this method, additional hash tables are considered as clusters that use exemplars-based feature fusion. They extend exemplar based approximation techniques by adopting a new feature representation $[z_i]_k = \frac{\delta_k K(x_i,u_k)}{\sum_{k'=1}^{M}\delta_k K(x_i,u_{k'})}$ here $\delta_k \in \{0,1\}$, u_k are exemplar points $\in \{\mathbb{R}^{\mathrm{d}}\}_{k=1}^{M}$ and $\kappa(\cdot,\cdot)$ is kernel function. The overall objective is to minimize:

$$\min_{\mu,\mathrm{B}} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} S_{ij}\left\|b_i - b_j\right\|_F^2$$

Here μ is the weight vector. The strength of presented scheme is that they assume a few cluster centers to measure the similarity in between entire data points. They defend linear weighting and bundling of multiple features in one vector using nonlinear transformation.

Alternating co-quantization based hashing Irie et al. (2015) proposed an improved multimodal retrieval which is based on the binary hash codes. The primary goal is to minimize the binary quantization error. To reduce the errors, the proposed model learns hash functions that provide a uniform mapping to one standard hash space with minimum distance among projected data points. The overall objective function (similarity preserving + quantization error) is formulated as:

$$\max_{A,B,U,V} tr\left(A^T C_{xx} A + D^T C_{yy} D + 2\alpha A^T C_{xy} D\right) + tr\left(2\lambda U X^T A + 2\eta V Y^T D\right)$$

Here λ, η are the balancing parameter, A and D are parameters of the binary quantizer, C is used to define inter and intra modal correlation matrix between X and Y and $U, V \in \{\pm 1\}$ are binary codes for X and Y. Later in quantization phase, two different quantizers are generated for two separate modalities viz. image and data which leads to end-to-end binary optimizers.

4.1.5 Deep hashing

In all hashing methods, the quality of extracted image features will affect the quality of generated hash codes. To ensure good quality and error free compact hash codes a joint learning model is needed to incorporate feature learning and hash value learning simultaneously. The model consists of different stages of learning and training deep neural networks. Deep hashing model has three simple steps to generate individual hash codes: (a) Image input (b) generation of intermediate features using convolution layer (c) a divide and encode module to distribute intermediate features into different channels further each channel is encoded into the hash bit. Representative methods (Abbas et al. 2018) are based on Recurrent Neural Networks (RNN) and Convolutional Networks (CNN). Some of these methods are listed and discussed below.

Point-wise deep learning Lin et al. (2015) suggested a deep learning framework for fast ANN search. The main aim is to generate compact binary codes driven by CNN. Instead of applying learning separately on image representation and binary codes, simultaneous learning has been adopted with the assumption of labeled data. The proposed approach implemented in three steps: (a) supervised pre-training on the dataset (b) fine tuning the network (c) image retrieval. For computationally cheap large scale learning with compact binary codes, multiple visual signatures are converted to binary codes. Following the data independence approach, the proposed approach is highly scalable which lead to very efficient and practical large scale learning algorithms.

Pair-wise labels based supervised hashing Li et al. (2016) addresses the issue of optimal compatibility among handcrafted features and hashing function used in various hashing methods. Simultaneous learning has been adopted, instead of applying learning separately on feature and hash code. The proposed end-to-end learning framework contains three main elements: (a) deep neural network for learning (b) hash function, which takes care of mapping between two spaces (c) loss function that is used to grade the hash code led by the pairwise label. The overall problem (feature learning + objective function) is formulated as:

$$\min_{\mathrm{B,W,v,\theta}} \mathcal{I} = -\sum_{s_{ij}\in S}\left(s_{ij}\theta_{ij} - \log\left(1 + \mathrm{e}^{\theta_{\mathrm{ij}}}\right)\right) + \eta\sum_{i=1}^{n}\left\|\mathrm{b_i} - \left(\mathrm{W^T}\emptyset(\mathrm{x_i};\theta) + \mathrm{v}\right)\right\|_2^2$$

$$s.t.\ \theta_{ij} = \frac{1}{2}u_i^T u_i,\ u_i = \left(\mathrm{W^T}\emptyset(\mathrm{x_i};\theta) + \mathrm{v}\right)$$

Here θ represents the all parameters of the 7 layers, $\emptyset(\mathrm{x_i};\theta)$ denotes the output of the last Full layer, v is the bias vector and η is the hyper-parameter. Following the principles, deep architecture integrates all three components which further permit the cyclic chain of feedback among different parts.

Regularized learning based bit scalable hashing Zhang et al. (2015) incorporate the concept of bit scalability to compute similarity among the images. For rapid and efficient image retrieval the author group training images into a pack of triplet sample. The author assumes that each sample consists of two images with a similar label and one with the dissimilar label. In particular, the learning algorithm has been implemented in a batch-process fashion that makes use of stochastic gradient descent (SGD) for minimizing the objective function in large-scale learning. The objective function is formally written as follows:

$$\min_{\mathrm{w},\omega}\sum_{i=1}^{N}\max\left\{D_{\mathbf{w}}\left(r_i, r_i^+, r_i^-\right), C\right\} + \lambda\ \mathrm{tr}\left(RLR^T\right)$$

Here $D_{\mathbf{w}}\left(r_i, r_i^+, r_i^-\right) = \mathrm{M}\left(r_i, r_i^+\right) - \mathrm{M}\left(r_i, r_i^-\right)$ with r_i, r_i^+, r_i^- are the approximated hash codes,M is weighted Euclidean distance, $C = -\frac{\mathrm{q}}{2}$ for q bit hashing code, $R = \left[r_1\widehat{\mathrm{w}}^{\frac{1}{2}}, r_2\widehat{\mathrm{w}}^{\frac{1}{2}}, \ldots, r_T\widehat{\mathrm{w}}^{\frac{1}{2}}\right]$ for T number of images and ω and λ are the parameter of hashing function and hyper-parameter for balancing respectively.

Similarity-adaptive deep hashing For learning similarity-preserving binary codes Shen et al. (2018) proposed an unsupervised deep hashing method. To introduce similarity preserving binary codes, the author uses anchor graph to propose pair-wise similarity graph. The anchor graph provides nonlinear data-to-anchor mapping, and they are easy to build in linear time which is directly proportional to a total number of data points. The proposed approach implemented in three steps: (a) Firstly, Euclidean loss layer has been introduced to train the deep model for error control (b) Secondly, the pair-wise similarity graph has been updated to make deep hash model and binary codes more compatible. (c) Finally, alternating direction method of multipliers (Boyd et al. 2010) is used for optimization. The updating of similarity graph is highly dependent on the output representations of deep hash model which subsequently improves the code optimization.

4.1.6 Online hashing

Existing data dependent and independent hashing schemes follows batch mode learning algorithms. That is learning based hashing methods demands massive labeled data set in advance for training and learning. Providing gigantic data in advance is infeasible. A new hashing technique that is focusing on updating of a hash function with a continuous stream of data has been developed known as online hashing. Some of the representative methods are listed and discussed below.

Kernel-based hashing Huang et al. (2013) proposed a real-time online kernel based hashing which has its origins in the problem of large data handling in existing batch-based learning schemes. In other contexts, the online learning is known as the passive aggressive learning. It is first introduced by Crammer et al. (2006). The objective function for updating projection matrix is inspired by structured prediction and given as below:

$$W^{t+1} = \arg\min_{W}\frac{1}{2}\left\|W - W^T\right\|_F^2 + C\xi$$

Here C is the aggressiveness parameter and $\xi \geq 0$.

Adaptive hashing Cakir and Sclaroff (2015) presented adaptive hashing, a new hashing approach that makes use of SGD for minimizing the objective function in large-scale learning. The objective function is defined as:

$$l\left(f(x_i), f\left(x_j\right); W\right) = \left(f(x_i)^T f\left(x_j\right) - Bs_{ij}\right)^2 + \frac{\lambda}{4}\left\| W^T W - I \right\|_F^2$$

An update strategy is utilized for deciding in what amount any hash function need to update and which of the hash function need to be corrected. W is updated as follows: $W^{t+1} = W^t - \eta^t \nabla_W l\left(f(x_i), f\left(x_j\right); W^t\right)$. Here η^t is the constant value, ∇_W is obtained approximating the sgn() function with sigmoid. During online learning orthogonality regularization is required to break the correlation among decision boundaries. Adaptive hashing is highly flexible and iterative as it updated the hash function with the speed of streaming data.

4.1.7 Quantization for hashing

Every approximate nearest neighbor searching techniques comprise two stages: projection and quantization. Initially, the data points are mapped into low dimensional space. Next, each assigned value quantized into binary code. During quantization information loss is instinctive, but in any bit selection approach for quantization, similarity preservation of Hamming and Euclidean distance and independence between bits are other major issues. A representative method includes single bit quantization (Indyk and Motwani 1998), double bit quantization (Kong and Li 2012), multiple bit quantizations (Moran et al. 2013) and much more. Some of these methods are listed and discussed below.

Variable bit quantization Moran et al. (2013) proposed a data driven variable bit allocation per locality sensitive hyperplane hashing for quantization stage of hashing based ANN search. In previous widely popular approaches SBQ (Indyk and Motwani 1998), DBQ (Kong and Li 2012), NPQ (Moran et al. 2013) and MQ (Kong et al. 2012) it is taken for granted that each hyperplane has been assigned with 1 or 2 bit respectively and in case of any method violate the defined assignment principle then either bits are discarded or other hyperplanes serves with lesser bits accordingly. Initially, in order to allocate a variable number of bits to each hyperplane, the F-measure has been computed for each hyperplane. The principle idea behind F-measure calculation is that large informative hyper-planes results in higher F-measure.

Most informative hash bit selection Liu et al. (2013) proposed a bit selection method named NDomSet which unify different selection problem into a single framework. The author presents a new family of hash bit selection from a pool of hashed bits which further grows into the discovery of standard dominant set in the bit graph. In this approach, firstly an edge-weighted graph is made, representing the bit pool. The proposed approach consisted of bit selection as quadratic programming to deal with similarity preservation and non-redundancy properties of bits. The experimental results show that the proposed non-uniform bit selection strategy perform well while using hash bits generated by different hashing methods viz. ITQ (Gong et al. 2013), SPH (Weiss et al. 2008), RMMH (Joly and Buisson 2011).

Exploring code space through multi bit quantization Wang et al. (2015) address the issue of quantization error and neighborhood structure of raw data. The author introduced an innovative multi-bit quantization scheme to use available code space at its maximum. To depict the similarity preservation among Hamming and Euclidean distance space, a distance error function has been introduced. They also proposed an O(n^3) algorithm for optimization to reduce the computation time. Results obtained by experiments demonstrate a possible improvement in search accuracy due to proposed quantization method. They also demonstrate the effectiveness of Hamming, Quadra and Manhattan distance on multi-bit quantization approach.

Table 2 compares various hash based indexing techniques regarding their pros, cons, dataset, feature used, evaluation measure and experimental results. Table 3 lists a brief introduction to different datasets used in hash based indexing techniques.

4.2 Non-hash based indexing

Non-hash based methods are classified in various categories viz. tree, bitmap, machine learning, deep learning and soft computing based. To maximize the scope of non-hash based methods here, we consider every technique one by one in order.

4.2.1 Tree based techniques

Earlier nearest neighbor searching methods are tree-based, and there is a need for indexing structure to partition the data space. Further different similarity measurement metrics, space partitioning, and pivot selection techniques are adopted to compute the nearest neighbor among image features. Due to these joint efforts, a large class of tree-based indexing techniques is available in literature such as R-Tree (Guttman 1984), KD-Tree (Friedman et al. 1977), VP-Tree (Markov 2004) and M-tree (Ciaccia et al. 1997). Studies below stressed on some essential techniques in large image datasets.

Pivot selection based Indexing schemes based on reference (pivot) objects results in minor distance computation and disk accesses. The different pivot selection algorithms compete for selection of right pivots, the number of pivots, pre-computed distances, and distribution of pivots. Pivot selection techniques are classified into two categories: Pivot partitioning and Pivot filtering. Further partitioning can be done in two different ways ball partitioning and hyperplane partitioning. Some of these techniques are listed and discussed below.

(i) Use of Spacing-Correlation Objects Selection for Vantage Indexing

Van Leuken et al. (2011) propose an algorithm to select a set of pivots carefully. The proposed vantage indexing makes use of a randomized incremental algorithm for the selection of a set of pivots. The two-pronged scheme firstly proposes criteria to measure the quality of pivots and secondly provides a pivot selection scheme with the condition of no random pre-selection. They have proposed two new quality criteria for variance of spacing and correlation, defined as:

$$\sigma_{sp}^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}\left(\left(d\left(A_{i+1}, V_j\right) - d\left(A_i, V_j\right)\right) - \mu\right)^2 \text{ and}$$

$$C(V_1, V_2) = \frac{\sum_i (d_{1i} - d_{2i}) - \sum_i d1i \sum_i d2i}{\sqrt{n\sum_i (d_{1i})^2 - \left(\sum_i d_{1i}\right)^2}\sqrt{n\sum_i (d_{1i})^2 - \left(\sum_i d_{2i}\right)^2}}$$

Here μ is the average spacing σ be the variance of spacing, A and V are objects and vantage objects respectively, $d(\cdot, \cdot)$ be the distance function and C be the linear correlation coefficient.

(ii) Cost-Based Approach for Optimizing Query Evaluation

Erik and Hetland (2012) proposed a cost-based approach to evaluating pivot selection dynamically. The main aim is to find selective pivots and an exact number of pivots required to assess a query. Initially, to perform a sequential search by skipping the searching in indexes, the cost

Table 2 Comparison of Hash Based Indexing Techniques

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Data-dependent hashing (unsupervised hashing)						
Discrete Graph Hashing (Liu et al. 2014)	Hashing with Graphs (Liu et al. 2011)	GIST Feature Deep Convolution Activation Feature LBP Feature	CIFAR SUN397 YouTube Faces Tiny	*Pros* The addition of graph based hashing, binary optimization and maximization algorithm for solving sub-problems enhances search accuracy The use of anchor graph results in good separation and balanced partitioning *Cons* The Training Time and Testing Time are not fast enough in comparison to state-of-the-art	Total Training Time Total Testing Time (encoding a query to hash code) Mean Precision Hash Lookup success rate and F-measure	*For YouTube Faces* Mean precision = 0.7805 (@ 128 hash bits) *For Tiny* Mean precision = 0.5358 (@ 128 hash bits) Nearly 100% Hash lookup success rates for all four dataset
Asymmetric Inner Product Hashing (Shen et al. 2017)	Supervised Hashing with Kernels (Liu et al. 2012)	Deep Convolution Activation Feature LBP Feature	SUN397 YouTube Faces ImageNet	*Pros* This approach generates high-quality binary hash codes The approach is used to learn binary hash codes and hash function simultaneously The correlation among inner products in this method is maximum, the key to generating high-quality codes	Total Training Time Total Test Time Precision@500 Mean Average Precision (MAP) Hamming Distance 2 precision Precision-Recall Curve	*For SUN397* MAP = 0.4348, Precision @ 500 = 0.4642 and Hamming distance 2 Precision = 0.0522 (@ 128 hash bits) *For YouTube Faces* MAP = 0.8400, Precision @ 500 = 0.9625 and Hamming distance 2 Precision = 0.9329 (@ 128 hash bits) *For Tiny* MAP = 0.5756, Precision @ 500 = 0.6658 and Hamming distance 2 Precision = 0.2243 (@ 128 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Scalable Graph Hashing (Jiang and Li 2015)	Discrete Graph Hashing (Liu et al. 2014)	GIST Feature	Tiny MIRFLICKR	*Pros* This approach suggests that there is no need to compute pair-wise similarity graph matrix Bit by bit sequential learning method results in accurate learning and retrieval The time and space complexity of proposed method is O(n) which sets a benchmark for accuracy and scalability.	Top K Precision (Hamming Ranking) Training Time	*For Tiny* Top-1000 precision = 0.6737 (@ 128 hash bits) and 0 .7357 (@ 256 hash bits) *For MIRFLICKR* Top-1000 precision = 0.6985 (@ 128 hash bits) and 0 .7584 (@ 256 hash bits)
Locally Linear Hashing to Capture Non-linear Manifolds (Irie et al. 2014)	Unsupervised Learning of Low Dimensional Manifolds (Saul and Roweis 2003)	GIST Feature VLAD	Yale Face Images USPS MNIST CIFAR ImageNet MIRFLICKR	*Pros* The proposed method is entirely unsupervised which is used to uncover the multiple structures obscured in image data This approach considers quantization loss and reconstruction error at the same time to map high dimension learned structure into a low dimension Hamming space	Binary Code Generation Time Hamming Ranking Time Computation Time Mean Precision Query Time	*For USPS* Precision@500 = 0.7815 (@ 64 hash bits) *For MNIST* Precision@500 = 0.8672 (@ 64 hash bits) *For CIFAR* Precision@500 = 0.2722 (@ 64 hash bits) *For ImageNet* Precision@500 = 0.0812 (@ 64 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Non-Linear Manifold Hashing (Shen et al. 2015)	Non-parametric function Induction (Delalleau et al. 2005)	GIST features SIFT features	MNIST CIFAR	*Pros* The supervised and unsupervised method is used to uncover the multiple structures obscured in images A non linear feature extraction method reported much better image classification and stepped down required storage	MAP Recall and Precision	*For CIFAR* MAP = 0.2085 (@ 64 hash bits) *For MNIST* MAP = 0.8774 (@ 64 hash bits)
Data-dependent hashing (supervised hashing)						
Discrete Hashing (Shen 2015)	Fast Supervised Hashing with Decision Trees (Lin et al. 2014)	GIST features Bag of Words (BoW) Features extracted from CNN model	MNIST CIFAR ImageNet NUS-WIDE	*Pros* This method jointly learns bit embedding and linear classifier under the optimization At the same time, it considers effectiveness, efficiency, and scalability issues The use of semantic information results in good multiclass classification separation and balanced partitioning	MAP Precision and Recall of hamming distance 2 Mean Classification Accuracy	*For CIFAR* Precision = 0.4229 and MAP = 0.4555 (@ 64 hash bits) *For MNIST* Precision = 0.8224 and Recall = 0.4977 (@ 128 hash bits) *For NUSWIDE* Precision = 0.5483 and MAP = 0.5316 (@ 128 hash bits) *For ImageNet* Precision = 0.5863 @ 128 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Discrete Hashing by Applying Column Sampling Approach (Kang et al. 2016)	Supervised Hashing with Latent Factor Models (Zhang et al. 2014)	GIST feature	CIFAR NUS-WIDE	*Pros* This approach suggests using an iterative method which could sample current available training data point into a single iteration With proposed binary optimization it is easy to avoid errors caused by conventional continuous optimization *Cons* Some state-of-the-art is much scalable than the proposed scheme Very few evaluation measurement metric taken into consideration	MAP Training Time	*For CIFAR* MAP = 0.6371 (@ 64 hash bits) For NUSWIDE MAP = 0.6329 (@ 64 hash bits)
Adaptive Relevance Based Multi-Task Visual Labeling (Kong and Li 2012)	Convex Multitask Learning with Flexible Task Clusters (Zhong and Kwok 2012)	SIFT Feature Bag of Words (BoW)	MIR-FLICKR TRECVID NUS-WIDE	*Pros* A use of hashing based feature dimension reduction reported much better image classification and stepped down required storage The complicated structure of processed features efficiently handles by task relevance scheme Direct use of task relationship in feature space and utilization of neighbor structure of complete training data helps in improvement of performance	MAP Training Time Mean Classification Accuracy	*For MIRFLICKR* MAP = 0.2920 and MCA = 0.7082 (@ 32 hash bits) *For NUSWIDE* MAP = 0.2145 and MCA = 0.2808 (@ 32 hash bits) *For TRECVID* MAP = 0.2115 and MCA = 0.8146 (@ 32 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Ranking-based hashing						
Weighting Scheme for Hash Code Ranking (Ji 2014)	QsRank (Zhang et al. 2012) and WhRank (Zhang et al. 2013)	SIFT Feature Bag of Words (BoW)	MNIST NUS-WIDE	*Pros* To get highly effective Hamming distance measurement the proposed scheme learns weights during similarity search They use mutual information of hash bit to propose mutual independence among hash bits The anchor graph is used to represent samples to make similarity measurement useful for various datasets	MAP Precision Recall Time cost (weight computation Time + distance ranking Time)	*For MNIST* MAP = 0.4588, Precision@5000 = 0.5426 and Recall = 0.3874 (@ 96 hash bits) *For NUSWIDE* MAP = 0.2962 and precision = 0.3896 (@ 96 hash bits)
Ranking Preserving Hashing (Wang et al. 2015)	Direct Optimization of Information Retrieval Measures (Qin et al. 2010)	GIST Feature	NUS-WIDE MIRFLICKR	*Pros* This approach suggests improving the ranking accuracy named Normalized Discounted Cumulative Gain (NDCG) which is the quality measure for hashing codes Optimization of NDCG is done through linear hashing function which evaluates the expectation of NDCG	Normalized Discounted Cumulative Gain (NDCG) using Hamming Ranking Average Cumulative Gain (ACG) using Hash Lookup with HM2	*For NUSWIDE* NDCG@20 = 0.234 (@ 64 hash bits) *For MIRFLICKR* NDCG@20 = 0.283 (@ 64 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Use of List-wise Supervision for Hash Codes Learning (Wang et al. 2013)	Learning to Rank (Li 2011)	Bag of Words (BoW) GIST Feature	CIFAR NUS-WIDE Tiny	*Pros* The proposed approach is shown to be efficient and scalable as dimensionality directs the complexity and the number of bits used This approach computes the quality of unsorted ranking lists by considering them as rank triplets	Normalized Discounted Cumulative Gain (NDCG) using Hamming Ranking Average Cumulative Gain (ACG) using Hash Lookup with HM3 Hash Code generation Time	*For CIFAR* NDCG = 0.923 and ACG = 0.1527(@ 32 hash bits) *For NUSWIDE* NDCG = 0.2317 and ACG = 0.4930 (@ 32 hash bits) *For Tiny* NDCG = 0.2827 and ACG = 0.3270 (@ 36 hash bits)
Multi-modal hashing						
Semantic-Preserving Hashing (Lin et al. 2015)	Collective Matrix Factorization Hashing for Multimodal Data (Ding et al. 2014)	Edge Histograms SIFT Features Bag of Words (BoW)	NUS-WIDE MIRFLICKR Wiki	*Pros* The Nonlinear projection and generated probability quality of kernel logistic regression help in good learning of projections Reduction in training and forecasting of hash function further increase the sampling size which maintains the search performance *Cons* It mentions about the extension of scheme for more than two modalities of data but does not provided any experimental proof of it	MAP (to retrieve text with images and to retrieve images with text)	*For Wiki* MAP = 0.6709 (@ 128 hash bits) *For MIRFLICKR* MAP = 0.7354 (@ 128 hash bits) *For NUSWIDE* MAP = 0.6580 (@ 128 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Semantic Topic Hashing via Latent Semantic Information (Wang et al. 2015)	Latent Semantic Sparse Hashing for Cross-Modal Similarity Search (Zhou et al. 2014)	SIFT Features Bag of Words (BoW)	NUS-WIDE Wiki	*Pros* A high-level abstraction results in the removal of redundant and noisy information A small number of the topic can be used to describe the data which further improves the semantic similarity Discrete nature of hashing provides practical learning of hash codes and naturally saves the memory cost and retrieval time *Cons* It mentions about the extension of scheme for more than two modalities of data but does not provided any experimental proof of it	MAP TopN- precision Recall and Precision	*For Wiki* MAP = 0.6623 (@ 112 hash bits), Recall- Precision curve = 0.1256 (@ 32 hash bits) and Top 2000 Precision = 0.1468 (@ 64 hash bits) *For NUSWIDE* MAP = 0.6984 (@ 112 hash bits), Recall- Precision curve = 0.3349 (@ 32 hash bits) and Top 2000 Precision = 0.5838 (@ 64 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Multiple Complementary Hash Tables Learning (Liu et al. 2015)	Fast Search in Hamming Space with Multi-Index Hashing (Norouzi et al. 2012)	GIST Features Spatial Pyramid Bag of Words (BoW) Wavelet Texture Blockwise Color Movement SIFT-based BoW Histograms	CIFAR TRECVID NUS-WIDE	*Pros* They defend linear weighting and bundling of multiple features in one vector using nonlinear transformation Due to the sparse similarity table, it is easy to optimize the objective function directly on behalf of exemplar initiated feature fusion The exemplar reweighting further improves the similarity among previously mis-separated neighbors for new table learning and also addresses the complementarity and low-rank similarity *Cons* The offline Training Time is high in comparison to state-of-the-art	Average Precision Training Time Search Time Hamming Distance Ranking Hash Table Lookup	*For CIFAR* Avg Precision @500 = 0.2760 (@ 16 lookup tables) *For TRECVID* Avg Precision @500 = 0.2511 (@ 16 lookup tables) *For NUSWIDE* Avg Precision @500 = 0.3507 (@ 16 lookup tables)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Alternating Co-Quantization based Hashing (Irie et al. 2015)	Neighborhood Preserving Embedding. (He et al. 2005)	Features extracted from CNN model Histogram of Oriented Gradient (HOG), Edge, Color and Texture Features	Wiki a-Pascal Microsoft COCO	*Pros* Two different quantizers are generated for two separate modalities viz. image and data which leads to end-to-end binary optimizers The author claims with evidence that their method can be extended to three or more modalities *Cons* The use of generalized multi-view analysis put a restriction on the proposed approach regarding coupling with other dimensionality reduction methods It mentions about the extension of scheme for more than two modalities of data but does not provided any experimental proof of it	MAP (for Image to text, image to image and Text to Image search) Precision Hash table Lookup Image Description (Image Annotation and Image Search) with Recall	*For Wiki* MAP = 0.309 (@ 64 hash bits) *For Pascal* MAP = 0.472 (@ 64 hash bits) *For COCO* MAP = 0.562 (@ 64 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Deep hashing						
Point-wise Deep learning (Lin et al. 2015)	Convolutional Architecture for Fast Feature Embedding (Jia et al. 2014)	Features extracted from CNN model	MNIST CIFAR Yahoo-1 M	*Pros* Instead of applying learning separately on image representation and binary codes, simultaneous learning has been adopted with the assumption of labeled data Following the data independence, the proposed approach is highly scalable Useful directions are provided for coarse level and excellent level search	Image Retrieval Precision MAP Error Rate of Classification	*For MNIST* Precision = 0.998 (@ 48 hash bits) *For CIFAR* Precision = 0.893 (@ 48 hash bits) *For Yahoo-1 M* Precision = 0.755 (@ 48 hash bits)
Pair-wise Labels based Supervised Hashing (Li et al. 2016)	Simultaneous Feature Learning and Hash Coding (Lai et al. 2013), Mat-Convolutional Neural Networks (Vedaldi and Lenc 2015), Supervised Hashing via Image Representation Learning (Xia et al. 2014)	GIST Feature SIFT Feature Wavelet Texture Blockwise Color Movement Color Histograms Edge Direction Histogram Features extracted from CNN model	CIFAR NUS-WIDE	*Pros* Simultaneous learning has been adopted, instead of applying learning separately on feature and hash code Deep architecture integrates all three components which further permit the cyclic chain of feedback among different components	MAP	*For CIFAR* MAP = 0.757 (@ 48 hash bits) *For NUSWIDE* MAP = 0.851 (@ 48 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Regularized Learning based Bit Scalable Hashing (Zhang et al. 2015)	Hyper-graph Laplacian Sparse Coding (Gao et al. 2013)	Features extracted from CNN model GIST Feature	MNIST CIFAR NUS-WIDE CUHK03	*Pros* For rapid and efficient image retrieval the author group training images into a pack of triplet sample The unequal weighting of each hash bit controls the code length by truncating the extra bits Simultaneous optimization has been adopted for image features and hash functions	MAP Top k-Precision Cumulative Matching Characteristics Curve	*For MNIST* MAP = 0.9809 and precision = 0.987 (@ 64 hash bits) *For CIFAR* MAP = 0.6326 and precision = 0.685 (@ 64 hash bits) *For NUSWIDE* MAP = 0.6414 and precision = 0.639 (@ 64 hash bits)
Similarity-Adaptive Hashing (Shen et al. 2018)	Hashing with Graphs (Liu et al. 2011), ADMM (Wu and GhanemB 2016)	Features extracted from CNN model (VGGNet) GIST Feature	CIFAR NUS-WIDE MNIST GIST	*Pros* The Iterative training procedure effectively generates semantic graphs with the surety to preserve the updated semantic similarity Better performance has been observed by imposing the constraints bit uncorrelation and balance simultaneously	MAP Precision Precision-Recall	*For CIFAR* MAP = 0.3030 and precision = 0.3202 (@ 64 hash bits) *For MNIST* MAP = 0.410 and precision = 0.4573 (@ 64 hash bits) *For NUSWIDE* MAP = 0.5633 and precision = 0.7504 (@ 64 hash bits) *For GIST* MAP = 0.3025 and precision = 0.4750 (@ 64 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Online hashing						
Kernel-based Hashing (Huang et al. 2013)	Minimal Loss Hashing for Compact Binary Codes (Norouzi and Fleet 2011)	GIST Feature	Photo Tourism LabelMe	*Pros* The system performs learning on the current round of data as the system does not require complete data in advance Kernel Mapping is the key to making this approach, produce results in a reasonable time Training Time and MAP is nearly linear	MAP Training Time	*For Photo Tourism* MAP = 0.781 (@ 96 hash bits) **For** *LabelMe* MAP = 0.302 (@ 96 hash bits)
Adaptive Hashing (Cakir and Sclaroff 2015)	Online Hashing (Huang et al. 2013)	GIST Feature	Half Dome LabelMe Tiny	*Pros* An updated strategy is utilized for deciding in what amount any hash function has to be updated and also to determine which of the hash function need to be corrected Adaptive hashing is highly flexible and iterative as it updated the hash function with the speed of streaming data	MAP	*For LabelMe* MAP = 0.71 (@ 256 hash bits) *For Half Dome* MAP = 0.83 (@ 256 hash bits) *For Tiny* MAP = 0.52 (@ 256 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Quantization for hashing						
Variable Bit Quantization (Moran et al. 2013)	Neighborhood Preserving Quantization (Moran et al. 2013), Double-Bit Quantization (Kong and Li 2012), Manhattan Hashing (Kong et al. 2012)	GIST Feature	CIFAR Reuters TDT-2	*Pros* The value of F-measure score directly decided by information on a hyperplane. That is more informative hyperplane results in a high F-measure score Objective function decides the lesser number of bits to allocate *Cons* Here only hamming ranking based scenarios are taken into account	Area under Precision and Recall Curve	*For CIFAR* Area under the Precision-Recall Curve = 0.219 (@ 32 hash bits for image and 128 bit for text) *For TDT-2* Area under the Precision-Recall Curve = 0.374 (@ 32 hash bits for image and 128 bit for text) *For Reuter* Area under the Precision-Recall Curve = 0.389 (@ 32 hash bits for image and 128 bit for text)
Most Informative Hash Bit selection (Liu et al. 2013)	Semantics-Aware Query-Adaptive Hashing	GIST Feature SIFT-based BoW Histograms Wavelet Texture Blockwise Color	GIST CIFAR NUS-WIDE	*Pros* Three new bit selection methods are proposed to avoid redundancy and to preserve similarity These methods extend the concept of quantization from single method to a common framework supporting multiple bit selection techniques The pools of bits are processed through edge weighted graph concept	MAP Recall Hamming Lookup Precision Precision Recall Curve	*For CIFAR* MAP = 0.1564 (@ 64 hash bits) *For NUSWIDE* MAP = 0.2774 (@ 64 hash bits) *For GIST* MAP = 0.0893 (@ 64 hash bits)

Table 2 continued

Technique	Baseline approach	Features used	Dataset used	Pros and cons	Evaluation measure	Results
Exploring Code Space through Multi Bit Quantization (Wang et al. 2015)	Manhattan Hashing (Kong et al. 2012), Hashing with Graphs (Liu et al. 2011)	GIST Feature Block-wise Color	LabelMe CIFAR NUS-WIDE	*Pros* It is a multi-bit quantization scheme to use available code space at its maximum A distance error function has been introduced to depict the similarity preservation among Hamming and Euclidean distance space	Retrieval Time cost with Hamming/Manhattan/Quadratic Distance Mean Average Precision Impact of Training Sample Size	*For LabelMe* MAP = 0.647 (@ 256 hash bits) *For CIFAR* MAP = 0.659 (@ 256 hash bits) *For NUSWIDE* MAP = 0.883 (@ 128 hash bits)

Table 3 List of dataset used in hash based indexing techniques

Year	Name	Institute	Type	No. of Images	General Description	Dimensions	Features/ Descriptors	License, Content, and Accessibility
1990	USPS (Le Cun et al. 1990)	New York University	Handwritten Images of Digits 0 to 9	11 K	Dataset with a total number of 10 classes of 8-bit grayscale images partitioned into a training and testing set with 7291 and 2007 samples	256	CNN for visual feature	© ⬇
1998	MNIST (LeCun et al. 1998)	New York University	Handwritten Images of Digits 0 to 9	70 K	The digits have been size-normalized and centered in a fixed-size image	784	CNN for visual feature	© ⬇
2005	Yale Face Image (zip Accessed 27 May 2017)	Yale University	Face Images	5.7 K	Dataset consists of 5760 grayscale images of 10 peoples with nine different poses. Mainly built for manifold learning	1080	CNN for visual feature	© ★ ⬇
2006	Photo Tourism (Snavely et al. 2006)	University of Washington	Tourism images	300 K	Dataset is reconstructed from 3 different tourist spots with a total number of 100 K patches of bitmap images	512	GIST	🅭 ⬇
2006	Half Dome (Accessed 27 May 2017)	University of Washington	Tourism images from California Yosemite National Park	107 K	The dataset is partitioned into a training and testing set with 105 K and 2 K samples with a total number of 100 K patches of bitmap images	512	GIST	🅭 ⬇
2008	Tiny (Torralba et al. 2008)	New York University	General Images	80 M	Dataset composed of 256 classes in total and all images have been ranked and labeled	384	GIST	© ★ ⬇ ✍

Table 3 continued

Year	Name	Institute	Type	No. of Images	General Description	Dimensions	Features/ Descriptors	License, Content, and Accessibility
2008	LabelMe (Torralba et al. 2008)	MIT Computer Science and Artificial Intelligence Laboratory	General Images in the Form of Polygons	22 K	Dataset consists of 111,490 polygons images; images are annotated online and offline	512	GIST	© ★ ⬇ @
2009	NUS-WIDE (Chua et al. 2009)	National University of Singapore	Images from Flickr and Associated Text Labels	270 K	A dataset with a total number of 5000 unique tags six types of low-level features extracted from these images, ground-truth for 81 concepts that can be used for evaluation	500	Color Histogram, Color Correlogram, Edge Direction Histogram, Wavelet Texture and BoW based on SIFT	🅭 ★ ⬇ ✎ 🌐 📌 📼 ⌛
2009	CIFAR (Krizhevsky 2009)	University of Toronto	General Images	60 K	Dataset consists of 10 classes with 6000 images per class. There are 50,000 training images and 10,000 test images	512	GIST, Color Histograms, and Bag-of-Visual-Words Histograms	© ★ ⬇ ✎
2009	ImageNet (Deng et al. 2009)	Princeton University	General Images	1.2 M	Dataset composed of 1000 categories in total and each image in the collection has been labeled with tags i.e. five captions per image and contains multiple objects	512 out of 4096$^{\#}$ (reduced by PCA)	SIFT, LBP, CNN	© ★ ⬇ ✎ 🗐 @

Table 3 continued

Year	Name	Institute	Type	No. of Images	General Description	Dimensions	Features/ Descriptors	License, Content, and Accessibility
2009	Pascal (Farhadi et al. 2009)	Pascal VOC Challenge	General Images viz. Person, Animal, Vehicle, etc.	23 K	Dataset composed of 20 classes in total and all scenes and objects have been labeled	128	HOG, Edge, and Color	© ★ ⬇ 🗐 @ 🏆
2010	SUN397 (Xiao et al. 2010)	Princeton University	Different Scene Images	130 K	Dataset composed of 397 categories in total and all scenes and objects have been labeled	1600 out of 12,288$^{\#}$ (reduced by PCA)	SIFT, GIST, HOG, SSIM, LBP	© ★ ⬇ ✍
2010	MIRFLICKR (Huiskes et al. 2010)	Leiden University	Images from Flickr	1 M	Each image in the collection has been labeled with tags. The average number of tags per image is 8.94	512	MPEG-7, Edge Histogram, and Homogeneous Texture Features	🅭 ★ ⬇ ✍ 🌍 📌 🚹 📷 ⌛
2011	Youtube Faces (Wolf et al. 2011)	Open University of Israel	Face Images	370 K* from 3425 videos	Dataset contain 1595 different people greyscale face images	1770	LBP, CSLBP and Four-Patch LBP vector	© ★ ⬇
2012	TRECVID (Yu et al. 2012)	Columbia University	General Images and Associated Text Labels	260 K*	Largest multi-attribute image dataset It contains 126 uniquely labeled query attributes and 6000 weak attributes.	1500	Spatial Pyramid Bag of Words (BoW), GIST and SIFT	🅭 ★ ⬇ ✍ 🗐 @ 🏆

Table 3 continued

Year	Name	Institute	Type	No. of Images	General Description	Dimensions	Features/ Descriptors	License, Content, and Accessibility
2014	Microsoft COCO (Lin et al. 2014)	Microsoft	Person Images	123 K	Dataset composed of 80 categories in total and each image in the collection has been labeled with tags i.e. five captions per image and contains multiple objects with ten ground truth values	128	CNN for visual feature, Skip-gram word vectors for text features	🅭 ★ ⬇ ⌛ 📌
2014	Yahoo-1 M (Lin et al. 2015)	Yahoo	Shopping Product Images	1 M	Dataset composed of 116 clothing specific categories in total has been collected from different yahoo shopping sites	4096	SIFT, LBP, CNN	© ★ ⬇
2014	CUHK (Li et al. 2014)	Chinese University of Hong Kong	Person Images	13 K images from 474 video clips	Dataset collected from streets, shopping malls, airports, and parks and some clips are collected from Pond5 and Getty Image	4096	LBP, HSV color histogram, and Gabor feature	© ★ ⬇
2014	Wiki (Pereira et al. 2014)	University of California	General Images and Associated Text Labels	3 K	Dataset constructed from different Wikipedia page. The dataset contains multiple objects with ten ground truth values and associated text view as 100 dimensions skip-grams	128	GIST and Bag-of-Visual-Words	© ★ ⬇

CNN Features
* Video Data

© Dataset contents are copyrighted.
✎ Dataset contains pre-computed features
🌍 Dataset comprises location information
⌛ Dataset comprises timestamp
🏆 Must to participate in the challenge to get the data

★ Dataset is annotated with class labels.
📌 Dataset comprises tags
📷 Dataset comprises camera information
@ Must to create an account to get the data
🅭 Dataset contents has a Creative Commons license

⬇ Dataset has to be downloaded.
☯ Dataset powered by ground-truth
🚹 Dataset comprises user info
🗐 Licence agreement is must get the data

model has been used. Quadratic Form Distance is used to compare histograms, and Euclidean distance has been used for experimental measurements. This approach believes in the static use of pivots (Traina et al. 2007) and the principle of maximizing the distance between pivots strengthen the approach.

(iii) Improving Node Split Strategy for Ball-Partitioning

De Souza et al. (2013) described ball-partitioning-based metric access methods that able to reduce the number of distance calculation and fast execution of distance-based queries. Node split strategies of M-tree and slim tree are too complex. The main aim is to propose the modified node split strategies. For better pivot selection, to avoid unbalanced splits and to categorize the nodes in different sets, three different algorithms viz. maximum dissimilarity, path distance sum based on prim's algorithm and reference element have been proposed. The proposed methods are shown to be efficient in the number of distance calculations and the time spent to build the structures.

(iv) Efficient k-closest pair queries by considering Effective Pruning Rules

Gao et al. (2015) proposed several algorithms for closest pair query processing by developing more effective pruning rule. The contribution of this paper is twofold: minimizes the number of distance computations as well as the number of node accesses. The proposed approach consisted of depth-first and best-first traversals to deal with duplicate accesses. Query efficiency is achieved by the employment of new pruning rules based on metric space. Experimental results on different data sets proved that the proposed scheme reached minimum distance computation and minimized I/O overhead.

(v) Metric all-kNN search by Considering Grouping, Reuse and Progressive Pruning Techniques

Chen et al. (2016) proposed a novel method for All-k-Nearest-Neighbor Search named Count M-tree (COM). The contribution of this paper is twofold: minimizes the number of distance computations as well as some node accesses. The indexing method relies on dynamic disk-based metric indexes which use different pruning rules, grouping, recycle and pruning methods. The strength of presented scheme is that the query set and the object set share the same dataset as no different dataset required to train them separately.

(vi) Radius-Sensitive Objective Function for Pivot Selection

Mao et al. (2016) proposed an improved metric space indexing which is based on the selection of several pivots. The system performs following functions: Firstly, they present importance of pivot selection. Their criterion for pivot selection is based on relevance and distance among pivot objects. An extended pruning mechanism has been presented with a framework to fix and select some relevant pivots. Radius-sensitive objective function for pivot selection is to maximize:

$$P = \underset{T}{\operatorname{argmax}} \left| (x, y) | x, y \in S, L_{\infty}\left(F_{T,d}(x), F_{T,d}(y)\right) \geq r \right|, |T|$$

Here S is the dataset in metric space, L_{∞} is the distance.

Clustering based The grouping of semantically similar images into clusters suggests a novel framework for nearest neighbor search in image retrieval. Instead of matching large part of image data set with query image it is meaningful to match a representative image(s) from a cluster. Following the principles, clustering based techniques work for a particular dataset. Varieties of clustering methods are available in the literature. Some of these methods are listed and discussed below.

(i) Priority Queues and Binary Feature based Scalable Nearest Neighbor Search

Muja and Lowe (2014) proposed a priority queue based algorithms for approximate nearest neighbor matching and proposed an algorithm for matching binary features also. The focus of this algorithm is to extend in finding a large number of closest neighbors. They have developed an extended version using a well-known best-bin-first strategy. A small number of iterations considerably cut down the tree build time which further maintains the search performance. The author also comes up with an open source library called the fast library for approximate nearest neighbors (FLANN) for the use of research community.

(ii) Indexing and Packaging of Semantically Similar Images into Super-Images

Luo et al. (2014) work on the packaging of semantically similar images into super-images. The fact behind this proposal is the strong relevance of the images into a dataset. The concept of super-image effectively bundle the multiple images into a single unit of same relevance and significantly decreases the size of the index. Semantic attribute extraction is the main issue in index construction. The attributes are extracted during packaging of one super-image i.e. during off-line indexing to make fast index structure. Visual compactness of a superimage candidate is calculated as:

$$\mathrm{VC}\big(\widehat{\mathrm{SI}}\big) = 1 - \frac{\sum_{i,j\in\widehat{\mathrm{SI}},i\neq j} \mathrm{dist}\big(\mathrm{TF_i},\mathrm{TF_j}\big)}{0.5 \times \big|\widehat{\mathrm{SI}}\big| \times \big(\big|\widehat{\mathrm{SI}}\big| - 1\big)}$$

Here TF is the normalized term frequency vector and dist() is the cosine distance.

(iii) Image Discovery through Clustering Similar Images

Zhang and Qiu (2015) proposed a scheme to discover landmark images in large image datasets. For rapid and efficient image retrieval the author group semantically similar images into clusters. One landmark image with different viewpoints adequately packed into a sub-cluster. Clusters can be partially overlapped. Each sub-cluster contains a center called as bundling center. Further, the bundling center of sub-cluster acts as a representative of the sub-cluster, to avoid exact image matching the scheme performs image matching by placing the bundling center.

Other techniques This half contains the approximation, relevance feedback and some other techniques for VIR. Retrieval of images by approximation, relevance feedback or by some other means viz. online techniques and query as video demands extra efforts but results in fine-grained results. Some of the representative methods are listed and discussed below.

(i) Indexing via Sparse Approximation

Borges et al. (2015) propose a high-dimensional indexing scheme based on sparse approximation techniques. The focus of this scheme is to improve the retrieval efficiency and to reduce the data dimensionality by designing a dictionary for mapping the feature vectors onto sparse feature vectors. The proposed scheme switches its direction to compute the high-dimensional sparse representations based on regression with the condition of preserved maximum locality. They showed that traversal of the data structure would be independent of metric function, low storage space required for efficient encoding of sparse representation and search space pruned efficiently.

(ii) Use of Hybrid Segmentation and Relevance Feedback for Colored Image Retrieval

Bose et al. (2015) proposed a new Relevance Feedback (RF) based VIR. The first advantage of the proposed scheme is feature-reweighting for relevance feedback i.e. to compute relevance score and weights of features the combination of feature-reweighting and instance based cluster density approach are used. The second advantage is a good utilization of image and shape contents. This scheme extracts the color and shape information through the color co-occurrence matrix (CCM) and segmentation of the image respectively. Here, unrestricted segmentation (k-mean) is used to segment the images. The relevance feedback scheme is initialized with three different approaches: intersection approach, union approach, and a combination approach. They have proposed two new measures retrieval efficiency, false discovery respectively to address the accuracy of retrieval.

(iii) Parallelism based Online Image Classification

Xie et al. (2015) propose a united algorithm for classification and retrieval named ONE (Online Nearest-neighbor Estimation). They observed that image classification and retrieval fundamentals are same and similarity measurement function could launch both of them. Its overall aim is to utilize the GPU parallelization to make the fast computation fully. The dimension reduction scheme is initiated with the help of both PCA and PQ. The scheme relies on feature extraction, training, quantization and an inverted index structure.

(iv) Query by Video Search

Yang et al. (2013) proposed a priority queue based algorithms for feature description and proposed a cache based bi-quantization algorithm also for information retrieval concept implementation. This method considers a short video clip as a query. Further, to find stable and representative good points among SIFT features, scheme perform feature tracking (Ramezani and Yaghmaee 2016) within video frames. The calculation of good point is formulated as:

$$G\left(p_i^j\right) = \propto \times \frac{Len\left(S\left(p_i^j\right)\right)}{Frame\ Count} + (1 - \propto) \times Cent\left(p_i^j\right)$$

$$s.t.\ Cent\left(p_i^j\right) = -\frac{\sum_{p \in S\left(p_i^j\right)} d(p, c)}{Len\left(S\left(p_i^j\right)\right) \times d(0, c)}$$

Here $S\left(p_i^j\right)$ denotes stableness of the point, Len() represents number of frames being tracked, and Framecount denotes the total number of frames in video query. Query representation is initiated by combining good points into a histogram.

Table 4 compares various tree-based indexing techniques regarding their pros, cons, dataset, feature used, evaluation measure and experimental results. Table 5 lists a brief introduction to different datasets used in tree-based indexing techniques.

4.2.2 Ranking based

Another category of tree-based indexing technique exists in literature to rectify the distance computation cost and index building cost by the use of different ranking strategies. In comparison to other methods, most of the ranking based methods are independent of distance measures. By exploring the ranking order and post processing, it is easy to the accurate and fast construction of index. The ranking scheme can be further extended to graph based, manifold ranking, supervised and unsupervised techniques.

4.2.3 Deep learning based

The need for full utilization of feature extraction, processing, and indexing in VIR shifted the research direction towards deep learning. The recently proposed models map low-level

Table 4 Comparison of Tree-Based Indexing Techniques

Technique	Baseline approach	Features used	Dataset	Pros and cons	Evaluation measure
Pivot selection based					
Use of Spacing-Correlation Objects Selection for Vantage Indexing (Van Leuken et al. 2011)	Vantage Point Indexing (Vleugels and Veltkamp 2002)	Color Histograms 64 BIN Shape Feature MPEG-7	Three Dataset of 1400 shape contour images, 50,000 color photographs, 500,000 fragments of music notation	*Pros* It generates good quality vantage objects based on spacing correlation Here no random pre-selection is made in the selection of vantage objects The proposed model establishes a good mapping in a feature space by selecting vantage objects and building the index simultaneously Their method is set to work well for multimodal retrieval	Average Distance Error Average Precision False positive ratio
Cost-based approach for optimizing query evaluation (Erik and Hetland 2012)	R* Tree, B + Tree Pivot-based LAESA Approach (Micó et al. 1994)	Not mentioned	NASA: It is a set of feature vectors	*Pros* A cost-based approach is developed for creating pivots; it identifies lesser selective pivots and query radii for the evaluation of each query It dynamically remove and create new pivots This approach is capable of indexing both memory (main) and disk resident data	Query time with pivots Query time with buffer Cost of sequential scan

Table 4 continued

Technique	Baseline approach	Features used	Dataset	Pros and cons	Evaluation measure
Improving node split strategy for ball-partitioning (De Souza et al. 2013)	Metric-Tree (Ciaccia et al. 1997)	Scalable Color	COPhIR dataset	*Pros* New split strategies used in this approach helps in lowering the overlap of the structure They emphasize on fast construction of trees by equally dividing the number of elements between the tree nodes Some useful directions are provided for computing distance	Global overlap Number of distance calculations Time spent to build the structures
Efficient k-closest pair queries by considering Effective Pruning Rules (Gao et al. 2015)	Metric-Tree (Ciaccia et al. 1997)	Objects are represented in metric space	SF and LA: Dataset of locations in California and San Francisco Color: Histograms of image database NASA dataset Signature dataset: each object is a string containing 64 alphabets	*Pros* The combined efforts of DFS and BFS traversals techniques and pruning rules results in reduced processing and I/O costs The improvement in performance is assured by controlling redundancy among object pairs through a novel concept The proposed technique claim better retrieval accuracy with no false positive and negative Proposed method is flexible with other existing indexes as well	Total cost the sum of the I/O time and CPU time) Number of node accesses and the selectivity

Table 4 continued

Technique	Baseline approach	Features used	Dataset	Pros and cons	Evaluation measure
Metric all-kNN search by Considering Grouping, Reuse and Progressive Pruning Techniques (Chen et al. 2016)	Metric-Tree (Ciaccia et al. 1997)	Objects are represented in metric space	SF and Color Signature dataset FFT: forest cover type data	*Pros* The addition of pruning rules, grouping, reuse, pre-processing enhances the performance of indexing This approach lower down the total distance computations and total node/page accesses The algorithm is flexible with other existing indexing techniques as well	Total cost (i.e., the sum of the I/O time and CPU time) Number of node accesses (NA) and the selectivity Efficiency of pruning rules
Radius-sensitive objective function for pivot selection (Mao et al. 2016)	Metric-Space Indexing	Objects are represented in metric space	US cartographic boundary data of Texas and Hawaii Protein sequence fragments	*Pros* A new objective function has been developed for range query It identifies lesser selective pivots and query radii for the evaluation of each query The proposed scheme needs lesser experimental efforts to select good pivots *Cons* It mention about variety of data but not discussed in the end	Objective function strength Estimation of intrinsic dimension

Table 4 continued

Technique	Baseline approach	Features used	Dataset	Pros and cons	Evaluation measure
Clustering based					
Priority queues based scalable nearest neighbor search (Muja and Lowe 2014)	k-d Forest and k-Means Tree (Muja and Lowe 2009)	Vector Binary Object SIFT Features SURF BRIEF ORB	UKBench SIFT features of different sizes obtained from the CD cover Tiny	*Pros* A small number of iterations considerably cuts down the tree build time which further maintains the search performance The idea of including best-bin-first strategy improves in the closest neighbors search	Search precision Search speedup (ratio of linear search and approximate search time)
Indexing and packaging of semantically similar images into super-images (Luo et al. 2014)	k-Means (Murthy et al. 2010), Scalable Recognition via Vocabulary Tree (Nistér and Stewénius 2006), Maximal cliques (Tomita et al. 2006), Semantic-aware Co-indexing (Zhang et al. 2015)	HSV histogram GIST Classemes Meta Class feature	UKBench dataset INRIA Holidays dataset	*Pros* The concept of super-image lower down the storage and retrieval time as index is constructed by lesser number of images of the whole dataset The embedding of semantic cues in super-images enhances retrieval performance An improvement in performance is assured by controlling redundancy among object pairs through a novel concept *Cons* It does not provide any strategy to unpack the super-image	MAP Recall

Table 4 continued

Technique	Baseline approach	Features used	Dataset	Pros and cons	Evaluation measure
Image discovery through clustering similar images (Zhang and Qiu 2015)	Tree Partitioning Min-Hash (Zhang et al. 2013)	Bag of visual word SIFT feature	Oxford Paris dataset Oxford building dataset	*Pros* The idea of the exclusion of query expansion process, helpful in improving recall rate With higher degree of freedom, the scheme perform efficient image discovery *Cons* A small size datasets are used to validate the performance results	MAP Recall False positive
Other techniques					
Indexing via sparse approximation (Borges et al. 2015)	Inverted Index based on Regression	GIST and SIFT Features	Tiny Billion vectors dataset	*Pros* The concept is helpful in search space pruning It identifies a non-metric data structure traversal There has been an increase in precision and scheme takes the only ¼ of time in comparison to linear search *Cons* A little storage overhead due to optimized dictionary	Searching time Precision

Table 4 continued

Technique	Baseline approach	Features used	Dataset	Pros and cons	Evaluation measure
Use of Hybrid Segmentation and relevance feedback for colored image retrieval (Bose et al. 2015)	k-Means (Murthy et al. 2010)	Color, Shape and Texture Features the color co-occurrence matrix	Dataset of 2000 images Dataset of 2020 images	*Pros* The approach has reduced tree build time which further maintains the search performance It introduces an open source library called FLANN for the use of research community *Cons* The important factor of dimensionality is not considered here The scheme is not fit for large data sets	Retrieval Efficiency with False Discovery Precision Recall
Parallelism based online image classification (Xie et al. 2015)	Bag of Visual Words (BoVW) (Arulmozhi and Abirami 2016), Nearest-Neighbor based Image Classification (Boiman et al. 2008)	Features extracted from CNN model VLAD	For image classification: LandUse Indoor SUN Oxford Pet Oxford Flower Caltech-UCSD Bird-200-2011 dataset For image retrieval: UK Bench Holiday dataset	*Pros* The addition of PCA, approximation, and parallelism enhances classification and retrieval results The novelty of concepts is assured by observing that image classification, and retrieval fundamentals are same *Cons* The proposed technique searching cost is more in comparison to state-of-the-art	Time and memory costs Descriptor Extraction Time

Table 4 continued

Technique	Baseline approach	Features used	Dataset	Pros and cons	Evaluation measure
Query by video search (Yang et al. 2013)	Inverted File Structure, Approximate k-Means, Hierarchical k-Means Tree (Muja and Lowe 2009)	SIFT	Oxford building dataset 11 landmark names suitable videos in YouTube	*Pros* It observes an improvement in retrieval efficiency. It shows accurate results comparable to methods using image as query *Cons* More features should be considered for video query as SIFT features are taken into account only The issue of quality and capturing of the video query are unanswered	MAP

Table 5 List of Dataset used in Tree-Based Indexing Techniques

Year	Name	Institute	Type	No. of Images	General Description	Dimensions	Features/ Descriptors	License, Content, and Accessibility
1966	Brodatz (Accessed 22 May 2017)	University of Southern California	Texture Images	1.7 K	A well known gray scale image dataset of 111 different textures divided into 16 blocks.	Not Mentioned	LBP, CCOM, and LAS	© ⬇
2000	MPEG-7 (Latecki et al. 2000)	Temple University	Images of Objects Shape	1.4 K	A popular dataset consists of 70 classes of 20 images per class to evaluate unsupervised distance learning approaches.	Not Mentioned	SS, BAS, IDSC, CFD, ASC and AIR	© ⬇
2001	SIMPLIcity (Wang et al. 2001)	Stanford University	General Image from Corel Photo Gallery	200 K	Dataset is part of collection of COREL dataset	Not Mentioned	Color and spatial features	© ★ ⬇
2003	UCID (Schaefer and Stich 2003)	Nottingham Trent University	Images of Natural Scenes and Objects	1.3 K	Each TIFF image in the collection has been labeled with tags and stored in compressed form.	Not Mentioned	Different MPEG-7 feature: CM, CCV, ACC, SCH	© ★ ⬇ 📼 ☯
2005	ALOI (Geusebroek et al. 2005)	University of Amsterdam	Images of Objects shape	110 K	A dataset of PNG format images composed of 1000 objects or scenes captured from different viewpoints and distances.	Not Mentioned	ACC, BIC, GCH, CCV and LCH	© ★ ⬇ 📼

Table 5 continued

Year	Name	Institute	Type	No. of Images	General Description	Dimensions	Features/ Descriptors	License, Content, and Accessibility
2006	UKBench (Nistér and Stewénius 2006)	Center for Visualization and Virtual environment, University of Kentucky	Images of popular music CD, indoor images	10 K	Dataset composed of 2550 objects or scenes captured from different viewpoints and distances.	Not Mentioned	ACC, BIC, CEED, FCTH, JCD and SIFT	© ★ ⬇
2006	Soccer (van de Weijer and Schmid 2006)	Joint team of Inria and Laboratories Jean Kuntzmann	Soccer Images	280	Dataset composed of images from 7 soccer teams, containing 40 images per class	Not Mentioned	BIC, ACC, and GCH	© ★ ⬇
2007	Oxford (Philbin et al. 2007)	University of Oxford	Images of Building	5 K	The dataset contains 11 different landmark images. Each image in the collection has been labeled with tags	Not Mentioned	SIFT with 16,334,970 number of descriptor	© ★ ⬇
2008	Paris (Philbin et al. 2008)	University of Oxford	Images of Landmark	6 K	Dataset composed of different landmark images. Each image in the collection has been labeled with tags	128	SIFT with 20,219,488 number of descriptor	© ★ ⬇
2008	Holiday INRIA (Jegou et al. 2008)	Joint project of INRIA and the Advestigo	Personal Holidays Photos	1.5 K	Dataset consists of a very large variety of scenes like natural, human-made, water and fire effects having pre-computed features and visual descriptors.	128	SIFT with 4,455,091 number of descriptor	© ★ ⬇ ✍

Table 5 continued

Year	Name	Institute	Type	No. of Images	General Description	Dimensions	Features/ Descriptors	License, Content, and Accessibility
2008	Oxford Flower (Nilsback and Zisserman 2008)	Oxford University	Flowers Images	8.1 K	The dataset contains 102 different classes of flower images.	Not Mentioned	HOG and SIFT	© ★ ⬇ ☯
2009	COPHIR (Bolettieri et al. 2009)	SAPIR European project managed by ISTI-CNR research institute in Pisa	General Images from Flickr	106 M	Each image in the collection has been labeled with tags. The average number of tags per image is 5.02, and 0.52 comments per image and total tag count are 105,999,880.	64	Different MPEG-7 Features: SC, CS&L, EH, HT	© ★ ⬇ 📌 🌍 🗐 @ 🚹
2009	Indoor (Quattoni and Torralba 2009)	MIT	Indoor Images	15 K	Dataset of JPG format images composed of 67 categories in total.	Not Mentioned	GIST, SIFT, and ROI	© ★ ⬇
2010	Landuse (Yang and Newsam 2010)	United States Geological Survey	Land Images	100	A dataset of various USA regions containing 21 classes of land images.	Not Mentioned	SIFT, BOVW and Color histogram	© ★ ⬇
2011	Caltech-UCSD Bird (Wah et al. 2011)	University of California	Birds Images	12 K	Dataset composed of 200 categories in total and all images have been annotated with 15 Part Locations, 312 Binary Attributes, 1 Bounding Box.	Not Mentioned	RGB color histograms and histograms of vector quantized SIFT descriptors	© ★ ⬇

Table 5 continued

Year	Name	Institute	Type	No. of Images	General Description	Dimensions	Features/ Descriptors	License, Content, and Accessibility
2011	Billion Vector (Jegou et al. 2011)	Institute for Research in Computer Science and Random Systems	General Images	1 M	The dataset is supplied with ground truth for each set of images in the form of the pre-computed k nearest neighbors and their square Euclidean distance.	960	SIFT and GIST	© ★ ⬇
2012	Oxford Pet (Parkhi et al. 2012)	Oxford University	Cat and Dog Images	7 K	A gray scale image dataset of 37 different breeds of cats and dogs in total and all scenes and objects have been labeled.	Not Mentioned	SIFT Spatial Histogram	© ★ ⬇ ☯

HOG: Histogram of gradient orientations
EH: Edge Histogram
ACC: Auto Color Correlograms
GCH: Global Color Histogram
CFD: Contour Features Descriptor
LAS: Local Activity Spectrum
CCOM: Color Co-Occurrence Matrix
FCTH: Fuzzy Color and Texture Histogram

SC: Scalable Color
HT: Homogeneous Texture
SS: Segment Saliences
SCH: Spatial-chromatic histograms
LBP: Local Binary Patterns
AIR: Articulation-Invariant Representation
SIFT: Scale-Invariant Feature Transform
IDSC: Inner Distance Shape Context

CS&L: Color Structure and layout
CCV: Colour coherence vectors
BAS: Beam Angle Statistics
CM: Colour moments
ASC: Aspect Shape Context
JCD: Joint Composite Descriptor
CEED: Color and Edge Directivity Descriptor
BIC: Border/Interior Pixel Classification

features into a high level with the help of nonlinear mapping techniques. Different feature extraction networks are available in literature viz. Alexnet (Krizhevsky et al. 2012), Goognet (Szegedy et al. 2015) and VGG (Simonyan and Zisserman 2014). Issues like number of layers in a network, distance metric, indexing techniques and much more are still unanswered and need to be benchmarked. Representative methods are based on RNN and CNN.

4.2.4 Machine learning based

It is essential to understand the discrepancy between low-level image features and high-level concept to design good applications for VIR. This leads to the so-called semantic gap in the VIR context. To reduce the semantic gap different classification and clustering techniques under machine learning are available. The support vector machine (SVM) and manifold learning are used to identify the category of the images in the dataset. Relevance feedback is also a good alternative. The level of learning and feedback further categorize learning methods in 'active' and 'passive' learning and 'long' and 'short' term learning approaches.

4.2.5 Soft computing based

Soft computing is combined effort of reasoning and deduction that employ development of membership and classification. The key to any productive soft computing based CBIR technique is to choose the best feature extraction scheme. Some of the soft computing techniques are Artificial Neural Network, Fuzzy Logic, and Evolutionary Computation. Above listed methods are discussed in Table 6.

In the previous half we present a detailed review of hash and non-hash based indexing techniques and we found that hash and non hash (tree) based techniques are totally different in nature. Generally speaking, in comparison to hash based techniques, the tree based techniques have following serious issues:

(1) Tree based methods are in need of large storage requirement in comparison to hashing based methods (sometimes more than the size of dataset itself) and the situation becomes worse when managing high dimensional datasets.
(2) For high dimensional datasets, in comparison to hashing based methods the retrieval accuracy of tree based methods approaches to linear search as backtracking takes long search time.
(3) The use of branch and bound criteria in tree based method makes them computationally more expensive as they are unable to stop after finding the optimal point and continuing in search of other points whereas in hashing based methods the criteria is to stop the search once they find a good enough points.
(4) On behalf of partitioning the entire dataset the hashing based methods repeatedly partition the dataset to generate a one 'bit' hash from each partitioning whereas tree based methods uses the recursive one.

On the application side tree based methods are applicable and useful when we have low dimensional datasets and user wanting exact nearest neighbor search. In the era of big data and deep learning, hashing based techniques are more suitable for high dimensional datasets and nearest neighbor search with low online computational overheads. Further, different intents and needs of users bring up unheard challenges as discussed in Sect. 7 later, all these challenges are in the scope of hash based methods. So we surely handle all of these issues by employing advanced hashing techniques. Therefore, to ensure fair comparisons a summary of their potential is presented in Table 7.

Table 6 Non Hash-based approaches (other than tree)

Technique	Description	Pros and Cons
Ranking based		
Graph based manifold ranking	Xu et al. (2011) developed a manifold ranking approach to overcome issues like high computation cost in graph construction, ranking calculation and out of sample queries. To introduce a graph-based manifold ranking, the author uses anchor graph to propose a novel adjacency matrix design. The anchor graph provides nonlinear data-to-anchor mapping, and they are easy to build in linear time which is directly proportional to a total number of data points. To get anchors the author opt k-mean algorithm. The concept of anchor graph is compelling as the addition of new samples does not force to update the anchors. Experimental results on different data sets proved that the proposed scheme achieved fast retrieval accuracy with reduced computational time and storage space. The author claims that their new adjacency matrix has been set to work well for other graph-based algorithms	*Pros* The addition of scalable graph and out-of-sample query results in efficient computation A significant response time is observed with out-of-sample retrieval
Rank correlation based unsupervised distance learning	Okada et al. (2015) incorporate the concept of ranking to compute similarity among the images. Instead of using the distance information Ranking-List based indexing schemes uses only ranking information. For query image, the ranked-list supposes to divide into three different sections. Each divided section determines some part of the ranked list. The dividing policy further helps in the individual processing of each section. In this article, the author derives six different rank correlation measures. Experimental results show that this is one of the best techniques for fast distance computation	*Pros* The proposed method is entirely unsupervised which is used to redefine the initial distance It uses ranking information to computes the similarity between ranked lists which makes it independent from distance information Two new rank correlation measures has been proposed

Table 6 continued

Technique	Description	Pros and Cons
Unsupervised distance learning via ranked-list	Valem et al. (2015) put forward the ranked-list recommendation. They present a novel unsupervised distance learning method. The proposed indexing scheme takes advantage of top positions of ranked lists. The position of an image in a ranked list plays a vital role in distance or weight computation. The individual position of images is not fixed in a related ranked list here because they strike up in the ranked list as the distance between images decreases accordingly. The experimental results show that with little computation attempts proposed indexing scheme effective and efficient in index construction. The shortcoming of the study is that it does not consider different descriptors and different modalities	*Pros* It jointly encounters the issues like effectiveness, efficiency, and scalability Low computation efforts are observed for different size of ranking lists The proposed approach uses a part of ranked lists which makes it flexible with other indexing techniques as well
Deep learning		
An indexing scheme for CNN features	Liu et al. (2015) proposed an indexing scheme for CNN features. The system makes use of BoW model and inverted table to deal with high dimensional global features. The focus of this method is to improve the computational efficiency and to reduce the data storage cost by mapping CNN (global) features to visual (local) features. Depending on semantic data of available CNN features space two different dictionary construction methods (product quantization based) are presented to map features from global space to local space. The tradeoff of information loss during mapping has been compensated by different strategies viz. binary embedding, multiple link, and assignment. The scheme has been performed on four different image databases and results show that the scheme improves the computational efficiency with little drop in precision	*Pros* The use of multiple linking strategies to index CNN features reported much better image retrieval Faster in comparison to brute force methods as the scheme is narrow down to inverted index table *Cons* The use of global features and inverted tables simultaneously results in quantization error Further multiple representations of global features require optimization at different stages

Table 6 continued

Technique	Description	Pros and Cons
Deep Learning based Multimodal Retrieval	Wang et al. (2016) proposed two novel learning techniques to improve the multi-modal image retrieval performance. To introduce general learning objective, the author presented unsupervised and supervised approach inspired by auto-encoders and deep CNN respectively. The associated semantic label information with image data finalize the selection of learning technique i.e. the presence of semantic labels advocates the use of supervised approach compared to unsupervised approach. The contribution of this paper is threefold: a nonlinear function for mapping, lesser information required for training input and a memory efficient training process for large dataset. Here VA files (Weber et al. 1998) are used to index the data.. Further, they build a distributed training platform named SINGA to support supervised and unsupervised learning approach. Results obtained by experiments demonstrate that the scheme performs efficient image discovery	*Pros* The method jointly implements unsupervised and supervised learning used to uncover the multiple structures obscured in different multimedia data *Cons* It stabilizes learning at the level of mapping and thus helps in training large scale models
Machine learning		
Maximum margin subspace learning	He et al. (2008) developed a dimensionality reduction approach named Maximum Margin Subspace. The author suggests local manifold structure discovery instead of the global structure as in PCA (Duda et al. 2000) and LDA (Swets and Weng 1996). Due to insufficient samples, it's hard to discover global structure (both geometrical and discriminative). In this approach, nearest neighbor graphs are constructed to model the local geometric structure. Depending on class and neighborhood information, the nearest neighbor graphs are split into two categories: within the class and between class graphs. Further, the labeling information has been used to maintain the within and between class graphs, for two similar labels within class graph connect these labels whereas for non-similar labels between class graph is used to connect them. The precision rate showed that the method is deemed successful	*Pros* It is easy to apply the standard classification or clustering techniques in proposed data representation space *Cons* The type of relevance feedback mechanism used is unanswered

Table 6 continued

Technique	Description	Pros and Cons
Manifold regularization for active learning	Zhang et al. (2017) addresses the issues with optimum experimental design (OED) and Transductive Experimental design (TED) based active learning methods. The ignorance of unlabelled database samples and a limited selection of these samples are the generic drawbacks of techniques mentioned above. In this work, the intrinsic manifold of unlabelled samples has been considered. They proposed a new manifold deformed learning method to select most representative and informative samples simultaneously. To introduce learning, they used data-dependent kernel function boosted by labeled and unlabeled data samples. The independence of training samples from labeled data helps to alleviate the sensitivity problem. They showed that for active learning it is possible to select multiple samples iteratively. This method has been evaluated on synthetic and real image dataset to learn most informative samples further these samples are used to train a classifier and unselected samples are used for testing. The author claims that their method is set to work well for a bunch of application viz. image retrieval, face recognition, image classification, and annotation	*Pros* The proposed approach is shown to be effective and scalable as the learning is directed by selecting multiple features iteratively The proposed kernel functions help to discover manifold structure from labeled and unlabeled samples both
Soft computing		
Fuzzy Logic based Improved Scale and Rotation Invariant	Mukane et al. (2014) presented a retrieval method involving fuzzy logic classifier. They aim to make an improved scale and rotation invariant (Ionescu et al. 2018) for retrieval analysis. To get highly efficient retrieval the proposed scheme group scale and rotation invariant in four groups viz. no rotation no scale, only scale, only rotation, both scale, and rotation. The Gaussian membership function is used to develop fuzzy logic classifier. Clusters have been formed for selected features of training samples before passing testing samples to the classifier. Results obtained by experiments demonstrate that with different invariant schemes the method achieves good retrieval rate, but the limitation of the study is that the method has been performed on few number of images.	*Pros* The approach is shown to be effective against the complex scale and rotation operation on texture features Useful directions are provided to develop fuzzy logic classifier *Cons* The retrieval performance affected with the increase in a total number of features

Table 6 continued

Technique	Description	Pros and Cons
Classification of Images via Fuzzy Classifiers and Boost learning	Korytkowski et al. (2016) developed an image classification approach for object identification and searching. A combined use of fuzzy logic and boosting learning reported much better image classification and stepped up searching speed. In the process of weak classifier building, initially, a feature has been randomly selected from an image class and selected feature's weight is computed using the Adaboost algorithm. The proposed technique automatically builds local feature based fuzzy rules for image classification, and further, it is free from initial parameters. The idea suggests that the addition of new classes have been done by adding fuzzy rules directly into the system. Results obtained by experiments demonstrate a possible improvement in learning and testing time compared to state-of-the-art. The author claims that their method is set to work well for all kind of fuzzy membership functions	*Pros* The model establishes a better image classification by combining fuzzy logic and boosting learning The method observed fast classification and learning The proposed approach is easy to expand as the addition of new classes has been done through fuzzy rule addition

Table 7 Comparison of Indexing Techniques

	Dimensionality	Nearest Neighbor	Storage Requirement	Retrieval Performance	Computational Complexity	Dataset Compression
Non Hashing	Low	Exact	High	Low	High	No
Hashing	High	Exact/ Approximate	Low	High	Low	Yes

5 Evaluation framework

In this section, we evaluate hashing techniques by careful analysis of results in the literature and approaches surveyed in this paper. By focusing on experimental works, we make an analysis of large number of notable hash based indexing techniques whose codes are available online. The experiments are run on an Intel Core i7 (4.20 GHz) with 32 GB of RAM, and Windows 10 OS. All the strategies are implemented in Matlab R2017b using the same framework to allow a fair comparison (Table 8).

5.1 Description of data

For the experiments regarding large scale similarity search and image retrieval we resorted to the five data sets: NUS-WIDE, CIFAR, SUN397, LabelMe, Wiki. The five datasets are chosen for their qualities viz. diverse, accessible, large size, and a rich set of descriptor considering different properties of the image. A large number of datasets are available in the literature as listed in Table 3 with their license, content, and accessibility issues. We have only opted general image datasets as they are largely opted by state-of the-art.

5.2 Evaluation metrices

The majority of datasets and techniques use Mean Average Precision (MAP) as the central evaluation metric for experiments. Along with MAP, they consider Mean Precision, Mean Classification Accuracy (MCA), Precision and Recall of Hamming distance 2. They also use two different metrics for evaluating retrieval: (i) Normalized Discounted Cumulative Gain (NDCG) using Hamming Ranking (ii) Average Cumulative Gain (ACG) using Hamming Ranking.

5.3 Evaluation mechanisms

The evaluation of system decides how far the system accomplish user's needs and technically which methodology is best for feature selection feature weighting, and hash function generation to make efficient and accurate retrieval process. Accordingly, researchers have explored a variety of ways to assessing user satisfaction and general evaluation of Image Retrieval system. It has always been a challenging and difficult matter for image retrieval importantly due to semantic gap (Wang et al. 2016) and further it is more problematic to pick out relevant set in the image database. There exist different ways of evaluating Image Retrieval systems in the literature are described below.

Table 8 Description of Hashing Techniques and abbreviations used in analysis

Abbreviation	Meaning	Abbreviation	Meaning
LSH$	Locality Sensitive Hashing (Datar et al. 2004)	ALSH$	Asymmetric Locality Sensitive Hashing (Shrivastava and Li 2014)
SH$	Spectral Hashing (Weiss et al. 2008)	IsoHash$	Isotropic Hashing (Kong and Li 2012)
ITQ$	Iterative Quantization (Gong et al. 2013)	AGH$	Anchor Graph Hashing (Liu et al. 2011)
InnerKSH$	Inner Product + Kernel Based Supervised Hashing (Liu et al. 2012)	BRE$	Binary Reconstructive Embedding (Kulis and Darrell 2009)
AIBC$	Asymmetric Binary Coding (Shen et al. 2017)	TSH$	Two step Hashing (Lin et al. 2013)
SDH$	Supervised Discrete Hashing (Shen 2015)	MLH$	Minimal Loss Hashing (Norouzi and Fleet 2011)
SPLH$	Sequential Projection Learning Hashing (Wang et al. 2010)	FastHash$	Fast Hash (Lin et al. 2014)
KSH$	Kernel Based Supervised Hashing (Liu et al. 2012)	COSDISH$	Column sampling based discrete supervised hashing (Kang et al. 2016)
LFH$	Latent Factor Model Hashing (Zhang et al. 2014)	DBQ$	Double bit Quantization for Hashing (Kong and Li 2012)
SSH$	Semi-Supervised Hashing (Wang et al. 2012)	DPSH$	Deep Supervised Hashing (Li et al. 2016)
CMFH$	Collective Matrix Factorization Hashing (Ding et al. 2014)	ADH$	Adaptive Hashing for Fast Similarity Search (Cakir and Sclaroff 2015)
LSSH$	Latent semantic sparse hashing (Zhou et al. 2014)	SPH$	Semantics-preserving hashing (Lin et al. 2015)
KLSH$	Kernelized Locality sensitive Hashing (Kulis and Grauman 2012)	STMH#	Semantic topic multimodal hashing (Wang et al. 2015)
SCM#	Linear cross-modal hashing (Zhang and Li 2014)	OKH#	Online kernel-based Hashing (Huang et al. 2013)
ACQ#	Alternating Co-Quantization (Irie et al. 2015)	RSH#	Ranking-based Supervised Hashing (Wang et al. 2013)
RPH#	Ranking Preserving Hashing (Wang et al. 2015)	CVH#	Cross view Hashing (Kumar and Udupa 2011)
CMSSH#	Cross Modality Similarity Sensitive Hashing (Bronstein et al. 2010)	PDH#	Predictable Dual View Hashing (Rastegari et al. 2013)

$ Implemented by using the online open source code
# Results are directly cited from papers

(1) **Precision** It is a measure of exactness which pertains to the fraction of the retrieved images that is relevant to the query.
(2) **Recall** It is the measure of completeness which refers to the fraction of relevant images that is responded to the query.
(3) **Average precision** This is the ratio of relevant images to irrelevant images in a specific number of retrieved images.
(4) **Mean average precision** This is the average of the average precision value for a set of queries.
(5) **Normalized discounted cumulative gain (*NDCG)***: It is the measure of uniformity between ground-truth relevance list to a query and estimated ranking positions.
(6) **F-measure** It is combined measure that assesses precision/recall tradeoff.

Besides these evaluation measures there exist some measures which can strengthen the evaluation procedure despite semantic gap type of challenges.

(1) **The size of index** It determines the storage utilization of generated Index. Practically the size of the index must be a fraction of dataset size.
(2) **Index compression** Some indexing techniques generate short hash codes or other similar codes for image data thereby reducing the storage requirement of Index.
(3) **Multimodal indexing** This refers to the ability of the index to support cross-media retrieval. As per current scenario the user intent to search via query by keywords, query by image or combination of both. Practically being multimodal, a system must support text to text, text to image and image to image search.

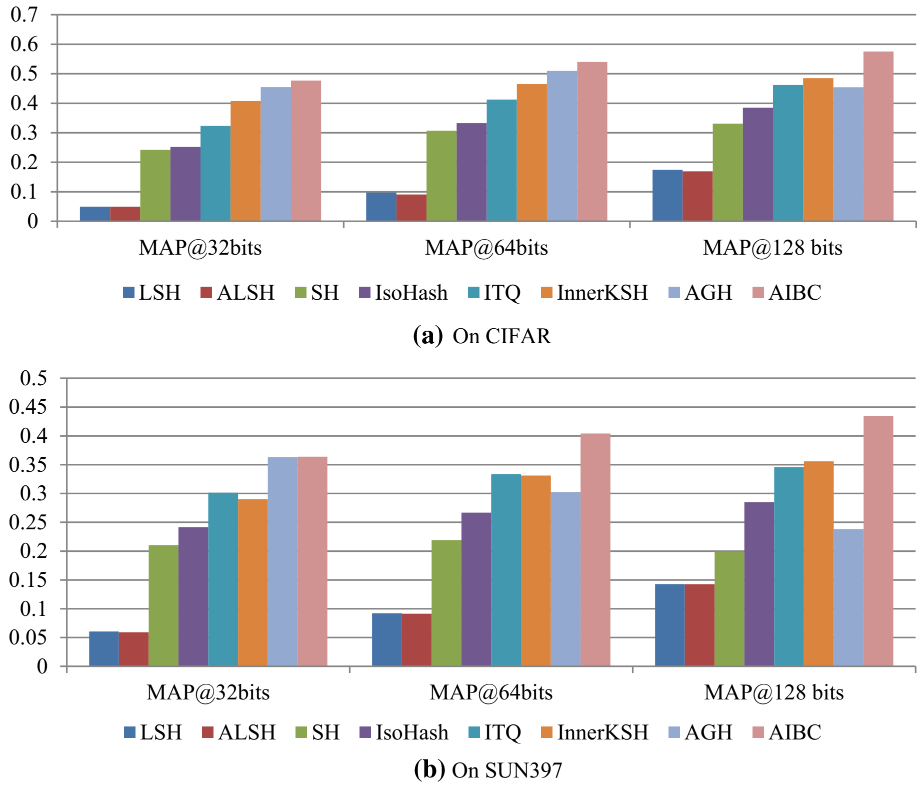
The metrics and measures mentioned above do not quantify user requirements. Other than image semantic, the different user intent may contain image clarity, quality, and associated meta-data. The satisfaction of user highly depends on the following factors:

(1) **User effort** This factor decides the role of the user and their efforts in devising queries, conducting the search, and viewing the output.
(2) **Visualization** This refers to the different ways to display the results to the users either in linear list format or 2-D grid format. Further, it influences the user's ability to employ the retrieved results.
(3) **Outcome coverage** This factor decides to which level the relevant images (by agreed relevant score) are included in the output.

5.4 Evaluated techniques and results

To evaluate the large-scale similarity search accuracy and effectiveness, we compare some hashing methods. To allow the comparison the most important aspects to evaluate is the algorithm performance metric as discussed in Sect. 5.2. Figure 4 below shows the comparison of various unsupervised data dependent techniques on CIFAR (GIST features) and SUN397 (CNN features) datasets.

It is evident from the Fig. 4a that AIBC (Shen et al. 2017) performed better than other unsupervised techniques. The AIBC improved Mean Average Precision nearly 2%, 4% and 9% for code length 32, 64 and 128 bits. The MAP results for various unsupervised data dependent are examined on the SUN397 dataset. Best score of MAP are obtained by AIBC because the correlation among inner products in this approach is maximum, the key to generating high-quality codes. Figure 4b displays the comparison of various techniques with AIBC regarding MAP values. The AIBC improved the MAP nearly 8% for code length 64 and 128 bits.



(a) On CIFAR

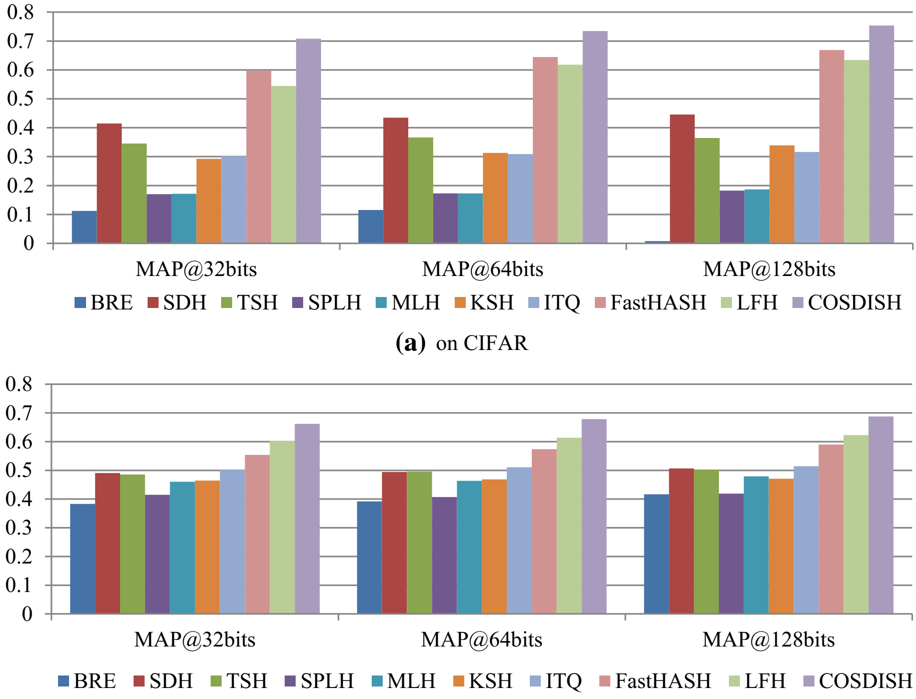
(b) On SUN397

Fig. 4 Comparison of Data-Dependent Hashing (Unsupervised) methods

In Fig. 5 we compare the performance of Column Sampling based Discrete Supervised Hashing (Kang et al. 2016) with state-of-the-art. Figure 5a presents the MAP obtained with values of code length ranging from 32, 64 and 128 bits. It can be seen that the column sampling based discrete hashing improved the MAP nearly 16%, 12% and 12% for code length 32, 64 and 128 bits respectively on CIFAR (GIST features) dataset. Figure 5b displays the comparison regarding MAP value for NUSWIDE (GIST features) dataset. The COSDISH improved MAP nearly 6% for code length 32, 64 and 128 bits with the use of column sampling technique to sample similarity matrix columns iteratively.

The ranking quality of retrieved results for Ranking based Hashing on the NUS-WIDE (GIST features) dataset is displayed in Fig. 6. NDCG@K is used to evaluate the ranking quality. Here K represents the value of retrieved instance. The figure presents the value of NDCG obtained with values of K ranging from 5 to 20 for 64 hashing bits. The Ranking preserving hashing improved NDCG nearly 1% under NDCG@5, 10 and 20 measures.

Next, the comparison of various Multi-Modal techniques on WiKi (CNN features for image data and skipgrams for text data) and NUS-WIDE (SIFT features for image data and binary tagging vector for text data) datasets has been represented. Figure 7a depicts the MAP values results on Wiki dataset for the image to image search. The MAP value results are observed for 32, 48 and 64-bit code length. The ACQ (Irie et al. 2015) improved the values approximately 1.5%, for three different code lengths respectively. Figure 7b also displays the comparison regarding MAP values among various techniques on Wiki dataset for the text to



(a) on CIFAR

(b) on NUSWIDE

Fig. 5 Comparison of Data-Dependent Hashing (Supervised) methods

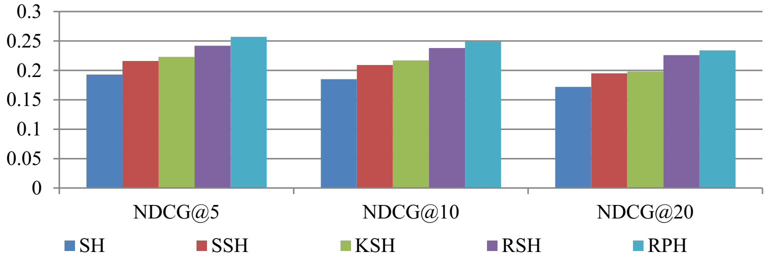
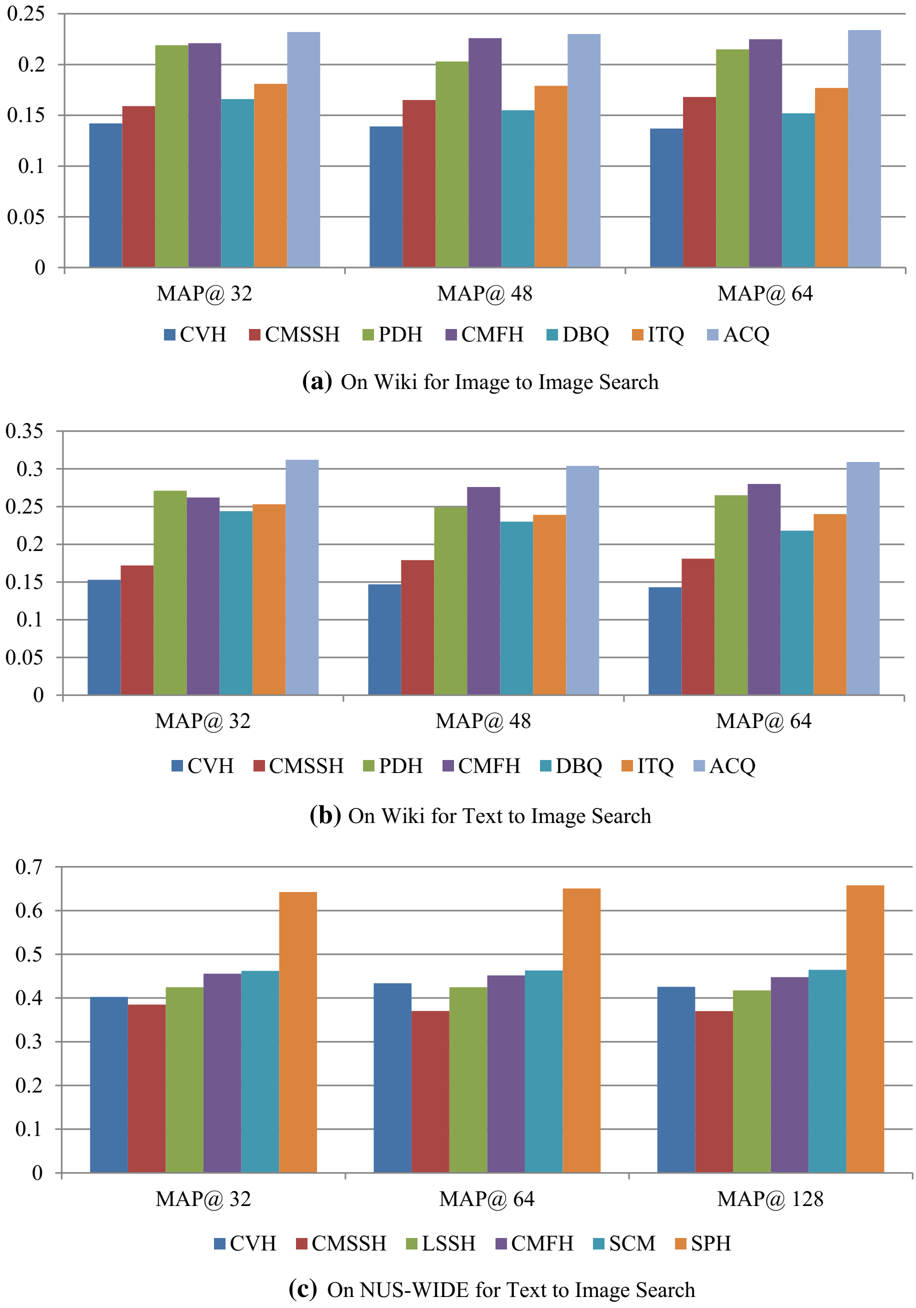


Fig. 6 Comparison of Ranking Based Hashing methods on NUS-WIDE

image search. Here we show the comparison of ACQ (Irie et al. 2015) with state-of-the-art multi-modal techniques. The ACQ improved the values nearly 4, 3 and 2% for 32, 48 and 64-bit code length respectively. Figure 7c depicts the MAP values results on NUS-WIDE dataset for the text to image search. Here we show the comparison of Semantics-Preserving Hashing (Lin et al. 2015) with state-of-the-art multi-modal techniques. The probability based SPH showed an average maximum improvement of 19% for 32, 48 and 64-bit code length respectively.



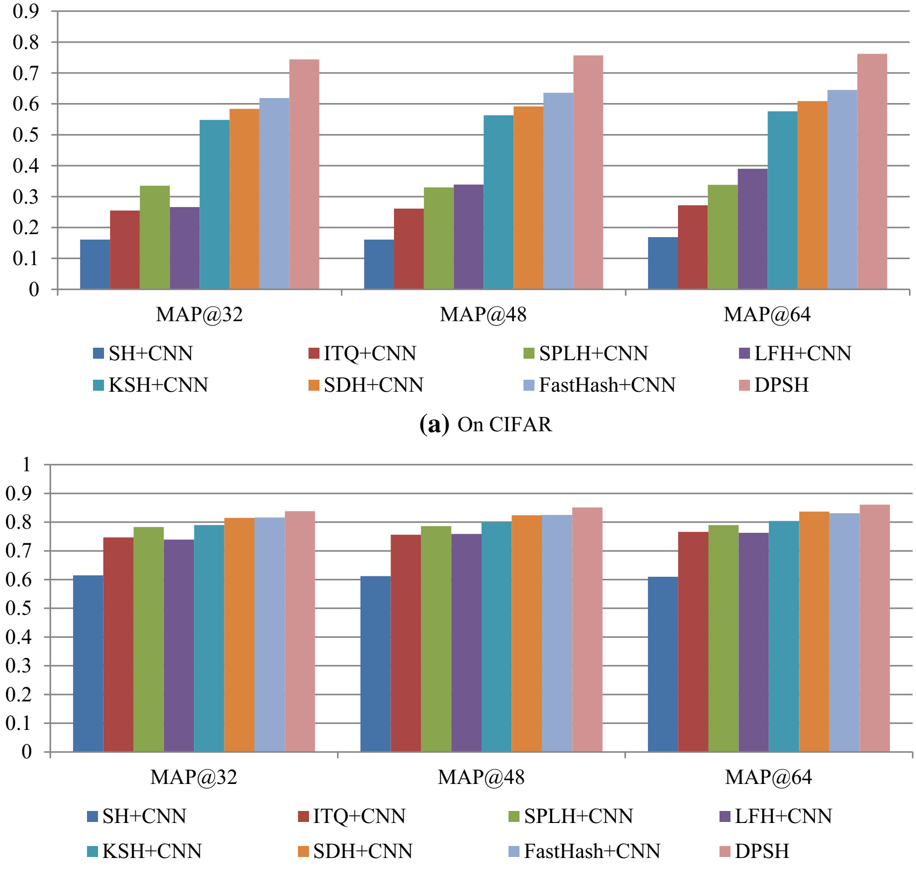
(a) On Wiki for Image to Image Search

(b) On Wiki for Text to Image Search

(c) On NUS-WIDE for Text to Image Search

Fig. 7 Comparison of MultiModal Hashing methods

The image retrieval results for Deep Hashing on CIFAR and NUSWIDE dataset are displayed in Fig. 8. The figure presents the value of MAP obtained for different values of hash code length viz. 32, 48 and 64 bits. Figure 8a depicts the MAP values results on CIFAR



(a) On CIFAR

(b) On NUS-WIDE

Fig. 8 Comparison of Deep Hashing methods

dataset. The Pair-wise Labels based Supervised Hashing (Li et al. 2016) improved MAP by 13, 12 and 12% respectively. Figure 8b also displays the comparison regarding MAP values among various techniques on NUS-WIDE dataset. The DPSH (Li et al. 2016) improved the values nearly 2, 3 and 3% respectively.

Figure 9 list the retrieval results of Adaptive Hashing, Online hashing and other Batch Hashing techniques are examined on LabelMe (GIST features) dataset. The MAP values for different hash code length generated by Adaptive hashing are not very promising. BRE (Kulis and Darrell 2009) showed an average maximum improvement of 6% in MAP on other batch techniques and 4% as compared to online methods. Further, it is observed that the concept of adaptive hashing (Cakir and Sclaroff 2015) does not put much impact on MAP values of kernel-based online hashing (Huang et al. 2013).

From the above evaluations, we draw the following conclusions:

(1) Supervised learning methods mostly attain good performance in comparison to unsupervised learning methods. As, the supervised method uses labeled data to learn hash codes and committed to maintaining semantic similarity constructed from semantic labels. In

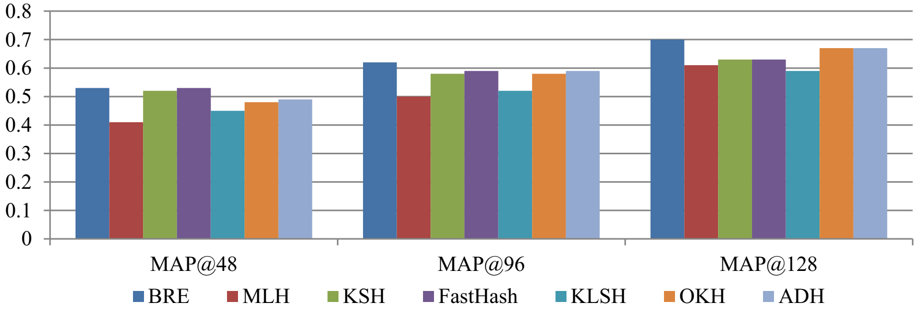


Fig. 9 Comparison of Adaptive Hashing, Online hashing, and other Batch Hashing techniques on LabelMe

comparison to unsupervised methods, supervised methods are slower during learning of large hash codes due to labeled data. This slowness can be overcome by incorporating deep learning concept.

(2) The performance of multimodal retrieval methods is totally guided by the quality of feature set. The results of text to image search are better than image to image search for multimodal datasets. The direct extension of two- modality algorithm into three or more modality is not possible.

(3) The methods like adaptive hashing and online hashing does not provide promising results. Even, the concept of adaptive hashing does not put much impact on performance of kernel-based online hashing.

(4) Nearest neighbor search on optimized compact hash codes of large dataset induces suboptimal results. By exploring the ranking order and accuracy, it is easy to evaluate the quality of hash codes. Associated relevant values of hash codes help to maintain the ranking order of search results.

6 Multimedia indexing evaluation programs

There are several well-instituted evaluation campaigns and meetings which provide test-bed and metric based environment to compare different proposed solutions in image retrieval domain. In this section, we describe various evaluation campaigns in image retrieval organized by the University of Sheffield, Stanford University, Princeton University and various research groups.

6.1 MediaEval

MediaEval[1] Benchmarking Initiative is a benchmarking activity organized by various research groups devoted to evaluating a new algorithm for multilingual multimedia content access and retrieval. It set out initially to benchmark some tasks related to the image, video, and music viz. Tagging Task for videos including Social Event Detection, Subject Classification, affect task, later from 2013 it sets to benchmark Diverse Images task.

[1] http://www.multimediaeval.org.

Table 9 Evolution of Mediaeval tasks

Task/Year	2010	2011	2012	2013	2014	2015	2016
Image related task							
Visual privacy task			✓	✓	✓		
Search and hyperlinking			✓	✓	✓	✓	
Crowd-sourcing				✓	✓		
Synchronization of multi-user event media					✓	✓	
Placing Task	✓	✓	✓	✓	✓	✓	✓
Automatic detection of manipulation (verification task)						✓	✓
Diverse images				✓	✓	✓	✓
Audio/video related task							
Tagging task	✓	✓	✓				
Affect task	✓	✓	✓	✓	✓	✓	
Rich speech retrieval task		✓					
Spoken web search		✓	✓	✓	✓	✓	✓
Social event detection		✓	✓	✓	✓		
User account matching			✓	✓			
MusiClef: multimodal music tagging			✓	✓			
QA task for spoken web				✓			
Emotion in music/video*				✓	✓	✓	✓*
Social speech				✓			
QA on classical music scores task					✓	✓	✓
Person discovery in TV						✓	✓
Context of experience task						✓	✓
Mini-drone video privacy						✓	

*Video dataset

The goal of Diverse Images task is to retrieve images from tourist images dataset that is participants has to refine (reorder) provided a ranked list of photos by maintaining diversity and representativeness. As a novelty in 2015, the diverse image task has been extended to multi-concept, ad-hoc, queries scenario. The dataset used for campaign is of small size having general-purpose visual/textual descriptors. It contains 95,000 images, splits into 50% for designing/training and 50% for evaluating. Different metric are considered to evaluate the system: Cluster Recall, Precision and F-measure. The evaluation of ranking (F-measure score) has been done by visiting the first page of the displayed outcome only. Other popular evaluation metrics are intent-aware expected reciprocal rank and the Normalized discounted cumulative gain (NDCG). Table 9 shows the evolution of MediaEval tasks over the year.

6.2 ImageCLEF

This evaluation forum[2] was initiated by University of Sheffield in 2003 nowadays it is run by individual different research groups. Initially, it is launched as a part of CLEF.[3] As mentioned on the website: It is a series of challenges to promote concept based annotation of images and multimodal and multilingual retrieval both. The first image retrieval track was included in 2003, where the objective was to perform similarity search and to find relevant images related to a topic in the cross-language environment. In 2004, Visual features were included the first time in any image CLEF track. The task was to perform image (tagged by English captions) search with text queries and visual features based medical image retrieval and classification. Over the year, it considers a broad range of topics related to multimedia retrieval and analysis. Different tracks under imageCLEF challenge are listed in Table 10.

6.3 ILSVRC

ILSVRC is in its 8th year in 2017 and is governed by a research team from Stanford and Princeton University respectively. The ImageNet[4] large scale task organized (since 2010) some object category classification and category detection task to promote the evaluation of proposed retrieval and annotation methods. Over the year ILSVRC consists of following tasks: Image classification (2010–2014), Single-object localization (2011–2014) and Object detection (2013–2014) and different dataset for testing and training and evolution tasks are listed in Tables 11 and 12 respectively.

7 Open issues and future challenges

Over the years, image retrieval has come a long way from simple linear scan techniques to more traditional learning and hashing techniques such as rank-based, deep learning, multimodal and online hashing techniques. Interest in topics such as quantization, supervised and unsupervised hashing is also increasing. The field of multimedia retrieval has witnessed different indexing techniques for data analysis. Further, different intents and needs of users bring up unheard challenges. We discuss some open and unresolved issues as follows:

7.1 A collection of big multimodal dataset

Instead of the uni-modal retrieval system, multiple unimodal systems can be combined to obtain multimodal retrieval system. To study and retrieve information across various modalities have been widely adopted by different research communities. There is an urgent need of large, annotated, easily available, and benchmarked multimodal dataset to train, test and evaluate multimodal algorithms.

[2] http://www.imageclef.org/.

[3] http://www.clef-campaign.org/.

[4] http://www.image-net.org/.

Table 10 Evolution of ImageCLEF tasks

Task/Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Ad hoc text retrieval	✓	✓	✓	✓	✓	✓	✓							
Domain-specific document retrieval	✓	✓	✓	✓	✓	✓								
Interactive cross-language retrieval	✓	✓	✓	✓	✓	✓	✓							
Cross-language question answering	✓	✓	✓	✓	✓	✓	✓	✓						
Cross-Language image retrieval/imageCLEF*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓*	✓*
Cross-language spoken document retrieval	✓	✓												
Cross-language geographical retrieval			✓	✓	✓	✓								
Multilingual web track			✓	✓	✓	✓								
Cross-language speech retrieval			✓	✓	✓									
Cross-language video retrieval						✓	✓							
Multilingual information filtering						✓	✓							
Information retrieval in intellectual property domain							✓	✓	✓	✓	✓			
Logfile analysis							✓	✓	✓					
Uncovering plagiarism, authorship, social software misuse								✓	✓	✓	✓	✓	✓	

Table 10 continued

Task/Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Cross-lingual expert search								✓						
Question answering for machine reading evaluation									✓	✓	✓			
Cultural heritage in CLEF									✓	✓	✓			
Evaluation of XML retrieval										✓	✓	✓		
Cross-language Evaluation of e-health document analysis										✓	✓	✓	✓	✓
Evaluation of online reputation management systems										✓	✓	✓		
Entity recognition											✓			
Question answering over linked data											✓	✓	✓	
LifeCLEF (bird, plant, and fish identification)												✓	✓	✓
Social book search													✓	✓
Digital text forensic														✓
News recommendation evaluation												✓	✓	✓
Cultural microblog contextualization														✓

*Video dataset

Table 11 Detail of datasets for ILSVRC challenge

	2010	2011	2012	2013		2014	
	For classification	For classification with localization	For classification with localization	For detection	For classification with localization	For detection	For classification with localization
Dataset name	Collected from Flickr and other search engines						
Validation and test data							
# of images	200 K	150 K	150 K	60 K	150 K	60 K	150 K
# of categories	1000	1000	1000	200	1000	200	1000
# of objects (not for testing)	NA	NA	NA	55 K	NA	55 K	NA
Training data							
# of images	1.2 M	1.2 M	1.2 M	395 K	1.2 M	456 K	1.2 M
# of categories	1000	1000	1000	200	1000	200	1000
# of objects	NA	NA	NA	345 K	NA	478 K	NA

	2015			2016			
	For detection	For classification with localization	Scene classification	For detection	For classification with localization	Scene classification	Scene parsing
Dataset name	Collected from Flickr and other search engines		Places 2 dataset	Collected from Flickr and other search engines		Places 2 dataset	ADE20 K Dataset
Validation and Test Data							
# of images	60 K	150 K	401 K	60 K	150 K	401 K	20 K
# of categories	200	1000	401	200	1000	401	150
# of objects (not for testing)	55 K	NA	NA	55 K	NA	NA	NA
Training data							
# of images	456 K	1.2 M	8.1 M	456 K	1.2 M	8.1 M	2 K
# of categories	200	1000	401	200	1000	401	150
# of objects	478 K	NA	NA	478 K	NA	NA	NA

Table 12 Evolution of ILSVRC tasks

Task/year	2010	2011	2012	2013	2014	2015	2016
Image classification	✓	✓	✓	✓			
Image classification with localization		✓	✓	✓	✓	✓	✓
Fine-grained classification			✓				
Object detection				✓	✓	✓	✓
Object detection from Video						✓	✓
Video scene classification						✓	✓
Scene parsing							✓

7.2 Utilization of labeled and unlabeled data samples to learn

Earlier VIR systems were simple, and were answerable to small datasets only i.e. they were independent or partially dependent on side (extra/labeled information. Simultaneous learning of labeled and unlabeled data samples put more complications in modern labeled information based systems. The ignorance of unlabelled database samples and a limited selection of these samples are the generic drawbacks of current approaches. Hence, need to utilize labeled and unlabeled data samples jointly for a better active learning.

7.3 Unsupervised deep learning

The need for full utilization of feature extraction, processing and indexing in VIR shifted the research direction towards deep learning. The recently proposed models map low-level features into a high level with the help of nonlinear mapping techniques. Researchers adopt the idea of supervised deep learning because it is a mature field and is in its middle stages of development. Human and animal learning is largely unsupervised which opens the door for researchers to develop future VIR system.

7.4 Multi feature fusion

To fulfil diverse user needs it becomes more challenging to develop fast and efficient multimodal VIR system as the traditional single feature or uni-modal based VIR system are lopsided. Better mechanism for fusing the multiple features for hash generation and learning are to be determined as assimilation of feature fusion concept can lessen the effect of well known Semantic gap.

7.5 Open evaluation program

Differences in technical capacities and data availability enlarge the VIR research gap between academics and industry based real application. A large number of researchers in academics bound to available resources and it is difficult to achieve industry based real application solution for them. It is necessary to organize open evaluation program to bridge this gap. The most felicitous part of organizing the open program is that they provide a common platform to industry and academic researchers to exchange more practical solutions. Further they jointly look into the key difficulties in different real time scenarios to develop application-independent methods.

The complete report of datasets, evaluation results, and the list of participants can be reviewed from the annual evaluation reports and websites of the challenges discussed above.

8 Conclusion

In this paper, we restrict ourselves to images and leave text and video indexing as a distinct topic. We propose to review two categories of indexing techniques developed for nearest neighbor search: (1) Hash based Indexing, and (2) Non-Hash based Indexing. These two categories contribute in many nearest neighbor search and similarity search techniques to provide efficient search capabilities. Further, the different methods of hash based and non-hash based indexing are categorized with brief details of the methodology employed including their pros and cons. Evaluation results are presented on various datasets for Hash-based techniques. A large number of potentially productive application areas and some relevant research domains, such as soft computing, clustering, ranking and deep learning have also been overviewed. We have summarized the open issues and future challenges. Our survey paper offers a number of practical guidelines to the readers:

1. It demonstrates the way to handle and process different types of queries.
2. The paper examines factors viz. query formation, image descriptors, type of hash function, code balancing, similarity measurement and code optimization can demonstrate the performance of hash based indexing techniques. The affect of above mentioned factors has been measured among standard 34 hash based approaches to derive the MAP and NDCG value.
3. Different types of low-level image representation and feature extraction criteria can coexist and be compared. (Multimodal hashing/Deep hashing methods).
4. There is an urgent need of large, annotated, easily available, and benchmarked multimodal dataset to train, test and evaluate multimodal algorithms.
5. There is a need.

   i. to utilize labeled and unlabeled data samples jointly for a better active learning.
   ii. of better mechanism for fusing the multiple features for hash generation and learning as assimilation of feature fusion concept can lessen the effect of well known Semantic gap.

Acknowledgements The authors thank the reviewers for their helpful comments. First author would like to thank Ministry of Electronics and IT, Government of INDIA, for providing fellowship under Grant number: PhD-MLA/4(61)/2015-16 (Visvesvaraya PhD Scheme for Electronics and IT) to pursue his Ph.D. work.

References

Abbas Q, Ibrahim MEA, Jaffar MA (2018) A comprehensive review of recent advances on deep vision systems. Int J Artif Intell Rev. https://doi.org/10.1007/s10462-018-9633-3

Andoni A, Indyk P (2008) Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. Int Mag Commun ACM 51(1):117–122

Arulmozhi P, Abirami S (2016) An analysis of BoVW and cBoVW based image retrieval. In: Proceedings of international symposium on big data and cloud computing challenges, pp 237–247

Babenko A, Slesarev A, Chigorin A, Lempitsky V (2014) Neural codes for image retrieval. In: Proceedings of European conference on computer vision, pp 584–599

Baeza-Yates R, Cunto W, Manber U, Wu S (1994) Proximity matching using fixed queries trees. Proceedings of Combinatorial pattern matching. Lecture notes in computer science 807:198–212

Baluja S, Covell M (2008) Learning to hash: forgiving hash functions and applications. Int J Data Min Knowl Discov 17(3):402–430

Bohm C, Berchtold S, Keim DA (2001) Searching in high-dimensional spaces: index structures for improving the performance of multimedia databases. Int J ACM Comput Surv 33(3):322–373

Boiman O, Shechtman E, Irani M (2008) In Defense of nearest-neighbor based image classification. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8

Bolettieri P, Esuli A, Falchi F, Lucchese C, Perego R, Piccioli T, Rabitti F (2009) CoPhIR: a test collection for content-based image retrieval. arXiv preprint arXiv: 0905.4627

Borges P, Mourão A, Magalhães J (2015) High-Dimensional indexing by sparse approximation. In: Proceedings of ACM international conference multimedia retrieval, pp 163–170

Bose S, Pal A, Mallick J, Kumar S, Rudra P (2015) A hybrid approach for improved content-based image retrieval using segmentation. arXiv preprint arXiv:1502.03215

Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2010) Distributed optimization and statistical learning via the lternating direction method of multipliers. Found Trends Mach Learn 3(1):1–122

Bronstein MM, Bronstein AM, Michel F, Paragios N (2010) Data fusion through cross-modality metric learning using similarity-sensitive hashing. In: Proceedings of computer vision and pattern recognition, pp 3594–3601

Cakir F, Sclaroff S (2015) Adaptive hashing for fast similarity search. In: Proceedings of IEEE international conference on computer vision, pp 1044–1052

Chen L, Gao Y, Chen G, Zhang H (2016) Metric all-k-nearest-neighbor search. Proc IEEE Int Conf Data Eng ICDE 28(1):1514–1515

Chua TS, Tang J, Hong R, Li H, Luo Z, Zheng Y (2009) NUS-WIDE: a real- world web image database from national university of Singapore. In: Proceedings of ACM international conference on image and video retrieval, pp 48–56

Ciaccia P, Patella M, Zezula P (1997) M-tree: an efficient access method for similarity search in metric spaces In: Proceedings of international conference on very large data bases, pp 426–435

Crammer K, Dekel O, Keshet J, Shalev-Shwartz S, Singer Y (2006) Online passive aggressive algorithms. J Mach Learn Res 7:511–585

Datar M, Immorlica N, Indyk P, Mirrokni VS (2004) Locality-sensitive hashing scheme based on p-stable distributions. In: Proceedings of annual symposium on computational geometry, pp 253–262

Datta R, Li J, Wang JZ (2005) Content-based image retrieval: approaches and trends of the new age. In: Proceedings of SIGMM international workshop on multimedia information retrieval, ACM, pp 253–262

Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: ideas, Influences, and Trends of the New Age. Comput Surv ACM 40(2):1–60

De Souza J, Razente H, Barioni M (2013) Faster construction of ball-partitioning-based metric access methods. In: Proceedings of ACM annual symposium on applied computing, pp 8–12

Delalleau O, Bengio Y, Le Roux N (2005) Efficient non-parametric function induction in semi-supervised learning. In: Proceedings of international workshop Artif. Intelli. Stat., pp 96–103

Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 248–255

Deng C, Liu X, Mu Y, Li J (2015) Large-scale multi-task image labeling with adaptive relevance discovery and feature hashing. Int J Signal Process ACM 112:137–145

Ding G, Guo Y, Zhou J (2014) Collective matrix factorization hashing for multimodal data. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 2083–2090

Ding K, Huo C, Fan B, Pan C (2015) kNN hashing with factorized neighborhood representation. In: Proceedings of IEEE international conference on computer vision, pp 1098–1106

Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T (2014) Decaf: a deep convolutional activation feature for generic visual recognition. In: Proceedings of international conference on international conference on machine learning, pp 647–655

Dosovitskiy A, Springenberg JT, Riedmiller M, Brox T (2014) Discriminative unsupervised feature learning with convolutional neural networks. In: Proceedings of international conference on neural information processing systems, pp 766–774

Duda RO, Hart PE, Stork DG (2000) Pattern classification, 2nd edn. Wiley, Hoboken

Erik S, Hetland L (2012) Dynamic optimization of queries in pivot-based indexing. Int J Multi Tools Apps 60(2):261–275

Farhadi A, Endres I, Hoiem D, Forsyth DA (2009) Describing objects by their attributes. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1778–1785

Fischer P, Dosovitskiy A, Brox T (2014) Descriptor matching with convolutional neural networks: a comparison to SIFT. arXiv Preprint arXiv: 1405.5769v2

Friedman JH, Bentley JL, Finkel RA (1977) An algorithm for finding best matches in logarithmic expected time. ACM Trans Math Softw 3(3):209–226

Gani A, Siddiqa A, Shamshirband S, Hanum F (2016) A survey on indexing techniques for big data: taxonomy and performance evaluation. Int J Knowl Inf Syst 46(2):241–284

Gao S, Tsang IW, Chia L (2013) Laplacian sparse coding, hypergraph laplacian sparse coding, and applications. IEEE Trans Pattern Anal Mach Intell 35(1):92–104

Gao Y, Chen L, Li X, Yao B, Chen G (2015) Efficient k-closest pair queries in general metric spaces. Int J Very Large Data Bases ACM 24(3):415–439

Ge T, He K, Sun J (2014) Graph cuts for supervised binary coding. In: Proceedings of international european conference on computer vision, pp 250–264

Geusebroek J-M, Burghouts GJ, Smeulders AWM (2005) The amsterdam library of object images. Int J Computer Vision 61(1):103–112

Gong Y, Lazebnik S, Gordo A, Perronnin F (2013) Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval. IEEE Trans Pattern Anal Mach Intell 35(12):2916–2929

Guttman A (1984) R-trees: a dynamic index structure for spatial searching. In: Proceedings of ACM SIGMOD international conference on management of data, pp 47–57

He X, Cai D, Yan S, Zhang HJ (2005) Neighborhood preserving embedding. In: Proceedings of IEEE international conference on computer vision, pp 1208–1213

He X, Cai D, Han J (2008) Learning a maximum margin subspace for image retrieval. IEEE Trans Knowl Data Eng 20(2):189–201

He J, Radhakrishnan R, Chang SF, Bauer C (2011) Compact hashing with joint optimization of search accuracy and time. In Proceedings of computer vision and pattern recognition, pp 753–760

Heo J, Lee Y, He J, Chang SF, Yoon S (2012) Spherical hashing. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 2957–2964

Hjaltason GR, Samet H (2003) Index-driven similarity search in metric spaces. ACM Trans Database Syst 28(4):517–580

Huang LK, Yang Q, Zheng W-S (2013) Online hashing. In: Proceedings of ACM international joint conference on artificial intelligence, pp 1422–1428

Huiskes MJ, Thomee B, Lew MS (2010) New trends and ideas in visual concept detection: The MIR-FLICKR retrieval evaluation initiative. In: Proceedings of ACM international conference on multimedia information retrieval, pp 527–536

Indyk P, Motwani R (1998) Approximate nearest neighbors: towards removing the curse of dimensionality. In Proceedings of ACM symposium on Theory of computing, pp 604–613

Ionescu RT, Ionescu AL, Mothe J, Popescu D (2018) Patch autocorrelation features: a translation and rotation invariant approach for image classification. Int J Artif Intell Rev 49:549–580

Irie G, Li Z, Wu X-M, Chang S-F (2014) Locally linear hashing for extracting non-linear manifolds. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 2123–2130

Irie G, Arai H, Taniguchi Y (2015) Alternating co-quantization for cross-modal hashing. In: Proceedings of IEEE international conference on computer vision, pp 1886–1894

Jegou H, Douze M, Schmid C (2008) Hamming Embedding and weak geometry consistency for large scale image search. In: Proceedings of european conference on computer vision, pp 304–317

Jegou H, Douze M, Schmid C (2011) Product quantization for nearest neighbor search. IEEE Trans Pattern Anal Mach Intell 33(1):117–128

Ji T (2014) Query-adaptive hash code ranking for fast nearest neighbor search. In: Proceedings of ACM international conference on multimedia, pp 1005–1008

Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093

Jiang Q, Li W (2015) Scalable graph hashing with feature transformation. In: Proceedings of international joint conference on artificial intelligence, pp 2248–2254

Joly A, Buisson O (2011) Random maximum margin hashing. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 873–880

Kang WC, Li WJ, Zhou ZH (2016) Column sampling based discrete supervised hashing. In: Proceedings of ACM AAAI conference on artificial intelligence, pp 1230–1236

Kong W, Li WJ (2012) Isotropic hashing. In: Proceedings of international conference advances in neural information processing systems, pp 1655–1663

Kong W, Li WJ (2012) Double-bit quantization for hashing. In: Proceedings of ACM AAAI conference on artificial intelligence, pp 634–640

Kong W, Li WJ, Guo M (2012) Manhattan hashing for large-scale image retrieval. In: Proceedings of the international ACM conference on research and development in information retrieval, pp 45–54

Korytkowski M, Rutkowski L, Scherer R (2016) Fast image classification by boosting fuzzy classifiers. Int J Info Sci 327(C):175–182

Krizhevsky A (2009) Learning multiple layers of features from tiny images. Technical report, University of Toronto

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Proceeding of international conference on neural information processing systems. pp 1097–1105

Kulis B, Darrell T (2009) Learning to hash with binary reconstructive embeddings. In: Proceedings of international conference advances in neural information processing systems, pp 1042–1050

Kulis B, Grauman K (2012) Kernelized locality-sensitive hashing for scalable image search. IEEE Trans Pattern Anal Mach Intell 34(6):1092–1104

Kumar S, Udupa R (2011) Learning hash functions for cross-view similarity search. In: Proceedings of international joint conference on artificial intelligence, pp 1360–1365

Kurasawa H, Fukagawa D, Takasu A, Adachi J (2010) Margin-based pivot selection for similarity search indexes. IEICE Trans Inf Syst E93(D6):1422–1432

Lai H, Pan Y, Liu Y, Yan S (2013) Simultaneous feature learning and hash coding with deep neural networks. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 3270–3278

Latecki LJ, Lakmper R, Eckhardt U (2000) Shape descriptors for non-rigid shapes with a single closed contour. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 424–429

Lazaridis I, Mehrotra S (2001) Progressive approximate aggregate queries with a multi-resolution tree structure. In: Proceedings of ACM SIGMOD international conference on management of data, pp 401–412

Le Cun Y, Boser B, Denker JS, Howard RE, HabbardW, Jackel LD, Henderson D (1990) Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems 2, Morgan Kaufmann Publishers Inc., pp 396–404

LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324

Li H (2011) A short introduction to learning to rank. IEICE Trans Inf Syst 94(10):1854–1862

Li X, Lin G, Shen C, van den Hengel A, Dick A (2013) Learning hash functions using column generation. In: Proceedings of international conference on international conference on machine learning, pp 142–150

Li W, Zhao R, Xiao T, Wang X (2014) Deepreid: deep filter pairing neural network for person re-identification. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 152–159

Li W, Wang S, Kang W (2016) Feature learning based deep supervised hashing with pairwise labels. In: Proceedings of ACM international joint conference on artificial intelligence, pp 1711–1717

Lin G, Shen C, Suter D, van den Hengel A (2013) A general two-step approach to learning-based hashing. In: Proceedings of IEEE international conference on computer vision, pp 2552–2559

Lin G, Shen C, Shi Q, van den Hengel A, Suter D (2014) Fast supervised hashing with decision trees for high dimensional data. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1971–1978

Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollar P, Zitnick L (2014) Microsoft coco: common objects in context. In: Proceedings of European conference on computer vision, pp 740–755

Lin Z, Hu M, Wang J (2015) Semantics-preserving hashing for cross-view retrieval. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 3864–3872

Lin K, Yang H, Hsiao J, Chen C (2015) Deep learning of binary hash codes for fast image retrieval. In: Proceedings of IEEE conference on computer vision and pattern recognition workshops, pp 27–35

Liu W, Wang J, Kumar S, Chang SF (2011) Hashing with graphs. In: Proceedings of international conference machine learning, pp 1–8

Liu W, Wang J, Mu Y, Kumar S, Chang SF (2012) Compact hyperplane hashing with bilinear functions. In: Proceedings of international conference on machine learning, pp 467–474

Liu W, Wang J, Ji R, Jiang YG, Chang SF (2012) Supervised hashing with kernels. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 2074–2081

Liu X, He J, Lang B, Chang SF (2013) Hash bit selection: a unified solution for selection problems in hashing. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1570–1577

Liu W, Mu C, Kumar S (2014) Discrete graph hashing. In: Proceedings of ACM international conference on neural information processing systems, pp 3419–3427

Liu X, Deng C, Lu J, Lang B (2015) Multi-view complementary hash tables for nearest neighbor search. In: Proceedings of IEEE international conference on computer vision, pp 1107–1115

Liu R, Zhao Y, Wei S, Zhu Z, Liao L, Qiu S (2015) Indexing of CNN features for large scale image search. arXiv preprint arXiv:1508.00217

Luo Q, Zhang S, Huang T, Gao W, Tian Q (2014) Indexing heterogeneous features with superimages. Int J Multimed Inf Retr 3(4):245–257

Mao R, Zhang P, Li X, Liu X, Lu M (2016) Pivot selection for metric-space indexing. Int J Mach Learn Cybern 7(2):311–323

Markov I (2004) VP-tree: content-based image indexing. In: Proceedings of IEEE international joint conference on neural networks, pp 45–50

Micó ML, Oncina J, Vidal E (1994) A new version of the nearest-neighbour approximating and eliminating search algorithm (AESA) with linear preprocessing time and memory requirements. Int J Pattern Recogn 15(1):9–17

Moran S, Lavrenko V, Osborne M (2013) Variable bit quantisation for LSH. In: Proceedings of annual meeting of the association for computational linguistics, pp 753–758

Moran S, Lavrenko V, Osborne M (2013) Neighbourhood preserving quantisation for LSH. In: Proceedings of ACM international conference on research and development in information retrieval, pp 1009–1012

Mu Y, Yan S (2010) Non-metric locality-sensitive hashing. In: Proceedings of AAAI conference on artificial intelligence, pp 539–544

Muja M, Lowe DG (2009) Fast approximate nearest neighbours with automatic algorithm configuration. In: Proceedings of international conference on computer vision theory and applications, pp 331–340

Muja M, Lowe DG (2014) Scalable nearest neighbor algorithms for high dimensional data. IEEE Trans Pattern Anal Mach Intell 36(11):2227–2240

Mukane Shailendrakumar M, Gengaje Sachin R, Bormane Dattatraya S (2014) A novel scale and rotation invariant texture image retrieval method using fuzzy logic classifier. Int J Comput Electr Eng 40(8):154–162

Murthy VSVS, Vamsidhar E, Swarup Kumar INVR, Sankara Rao P (2010) Content based image retrieval using hierarchical and k-means clustering techniques. Int J Eng Sci Technol 2(3):209–212

Neyshabur B, Srebro N, Salakhutdinov R, Makarychev Y, Yadollahpour P (2013) The power of asymmetry in binary hashing. In: Proceedings of international conference advances in neural information processing systems, pp 2823–2831

Nilsback M, Zisserman A (2008) Automated flower classification over a large number of classes. In: Proceedings of Indian conference on computer vision, graphics and image processing, pp 722–729

Nistér D, Stewénius H (2006) Scalable recognition with a vocabulary tree. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 2161–2168

Norouzi M, Fleet DJ (2011) Minimal loss hashing for compact binary codes. In: Proceedings of international conference machine learning, pp 353–360

Norouzi M, Punjani A, Fleet DJ (2012) Fast search in hamming space with multi-index hashing. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 3108–3115

Okada CY, Carlos D, Pedronette G, Torres RS (2015) Unsupervised distance learning by rank correlation measures for image retrieval. In: Proceedings of ACM international conference on multimedia retrieval, pp 331–338

Online. http://www.ux.uis.no/~tranden/brodatz.html. Accessed 22 May 2017

Online. http://phototour.cs.washington.edu/patches/default.htm. Accessed 27 May 2017

Online. http://vision.ucsd.edu/datasets/yale_face_dataset_original/yalefaces.zip Accessed 27 May 2017

Parkhi O, Vedaldi A, Zisserman A, Jawahar C (2012) Cats and dogs. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 3498–3505

Pedreira O, Brisaboa NR (2007) Spatial selection of sparse pivots for similarity. In: Proceedings of Springer international conference on current trends in theory and practice of computer science. Lecture notes in computer science, pp 434–445

Pereira J, Coviello E, Doyle G, Rasiwasia N, Lanckriet G, Levy R, Vasconcelos N (2014) On the role of correlation and abstraction in cross-modal multimedia retrieval. IEEE Trans Pattern Anal Mach Intell 36(3):521–535

Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2007) Object retrieval with large vocabularies and fast spatial matching. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8

Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2008) Lost in quantization: improving particular object retrieval in large scale image databases. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8

Qin Tao, Liu Tie-Yan, Li Hang (2010) A general approximation framework for direct optimization of information retrieval measures. Int J Inf Retr 13(4):375–397

Quattoni A, Torralba A (2009) Recognizing indoor scenes In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 413–420

Ramezani M, Yaghmaee F (2016) A review on human action analysis in videos for retrieval Applications. Int J Artif Intell Rev 4:485–514

Rasiwasia N, Costa Pereira J, Coviello E, Doyle G, Lanckriet GR, Levy R, Vasconcelos N (2010) A new approach to cross-modal multimedia retrieval. In: Proceedings of ACM international conference on multimedia, pp 251–260

Rastegari M, Choi J, Fakhraei S, Hal D, Davis LS (2013) Predictable dual-view hashing. In: Proceedings of international conference on machine learning, pp 1328–1336

Robinson JT (1981) The K–D–B–tree: a search structure for large multidimensional dynamic indexes. In: Proceedings of ACM SIGMOD international conference on management of data, pp 10–18

Saul L, Roweis S (2003) Think globally, fit locally: unsupervised learning of low dimensional manifolds. Int J Mach Learn Res 4:119–155

Schaefer G, Stich M (2003) UCID: an uncompressed color image database. In: Proceedings of international society for optical engineering, pp 472–480

Sharif U, Mehmood Z, Mahmood T, Javid MA, Rehman A, Saba T (2018) Scene analysis and search using local features and support vector machine for effective content-based image retrieval. Int J Artif Intell Rev. https://doi.org/10.1007/s10462-018-9636-0

Shen F (2015) Supervised discrete hashing. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 37–45

Shen F, Shen C, Shi Q, van den Hengel A, Tang Z, Shen HT (2015) Hashing on Nonlinear Manifolds. IEEE Trans Image Process 24(6):1839–1851

Shen F, Zhou X, Yang Y, Song J, Shen HT, Tao D (2016) A Fast Optimization Method for General Binary Code Learning. IEEE Trans Image Process 25(12):5610–5621

Shen F, Liu W, Zhang S, Yang Y, Shen HT (2017) Asymmetric binary coding for image search. IEEE Trans Multimed 19(9):2022–2032

Shen F, Xu Y, Liu L, Yang Y, Huang Z, Shen HT (2018) Unsupervised deep hashing with similarity-adaptive and discrete optimization. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/tpami.2018.2789887

Shrivastava A, Li P (2014) Asymmetric LSH (ALSH) for sublinear time maximum inner product search. In: Proceedings of international conference advances in neural information processing systems, pp 2321–2329

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556

Snavely N, Seitz S, Szeliski R (2006) Photo tourism: exploring photo collections in 3D. ACM Trans Gr 25(3):835–846

Song J, Yang Y, Yang Y, Huang Z, Shen H (2013) Inter-media hashing for large-scale retrieval from heterogeneous data sources. In: Proceedings of ACM SIGMOD international conference on management of data, pp 785–796

Swets DL, Weng J (1996) Using discriminant eigen features for image retrieval. IEEE Trans Pattern Anal Mach Intell 18(8):831–836

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–9

Téllez ES, Chávez E, Bejar JO (2014) Scalable proximity indexing with the list of clusters. In: Proceedings of international IEEE autumn meeting on power, electronics and computing, pp 1–6

Tomita E, Tanaka A, Takahashi H (2006) The worst-case time complexity for generating all maximal cliques and computational experiments. Int J Theor Comput Sci 363(1):28–42

Torralba A, Fergu R, Weiss Y (2008) Small codes and large image databases for recognition. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8

Torralba A, Fergus R, Freeman W (2008b) 80 million tiny images: a large data set for nonparametric object and scene recognition. IEEE Trans Pattern Anal Mach Intell 30(11):1958–1970

Traina CJ, Filho RF, Traina AJ, Vieira MR, Faloutsos C (2007) The omni-family of all-purpose access methods: a simple and effective way to make similarity search more efficient. Int J Very Large Data Bases 16(4):483–505

Uhlmann JK (1991) Satisfying general proximity/similarity queries with metric trees. Int J Inf Proc Lett 40(4):175–179

Uysal MS, Beecks C, Schmücking J, Seidl T (2015) Efficient similarity search in scientific databases with feature signatures. In: Proceedings of international conference on scientific and statistical database management, pp 1–12

Valem LP, Pedronette DCG, Torres RS, Borin E, Almeida J (2015) Effective, efficient, and scalable unsupervised distance learning in image retrieval tasks. In: Proceedings of ACM international conference multimedia retrieval, pp 51–58

van de Weijer J, Schmid C (2006) Coloring local feature extraction. In: Proceedings of European conference on Computer vision, pp 334–348

Van Leuken RH, Veltkamp RC, Typke R (2011) Selecting vantage objects for similarity indexing. ACM Trans Multimed Comput Commun Appl 7(3):453–456

Vedaldi A, Lenc K (2015) Mat-ConvNet—convolutional neural networks for MATLAB. In: Proceedings of ACM international conference on multimedia, pp 689–692

Vleugels J, Veltkamp RC (2002) Efficient image retrieval through vantage objects. Int J Pattern Recogn 35(1):69–80

Wah C, Branson S, Welinder P, Perona P, Belongie S (2011) The Caltech-UCSD birds-200-2011 dataset, computation & neural systems technical report, California Institute of Technology

Wang JZ, Wiederhold G, Firschein O, Wei SX (1998) Content-based image indexing and searching using Daubechies' wavelets. Int J Digit Libr 1(4):311–328

Wang James Z, Li Jia, Wiederhold G (2001) SIMPLIcity: semantics-sensitive integrated matching for picture libraries. IEEE Trans Pattern Anal Mach Intell 23(9):947–963

Wang J, Kumar S, Chang SF (2010) Sequential projection learning for hashing with compact codes. In: Proceedings of international conference on machine learning, pp 1127–1134

Wang J, Kumar S, Chang SF (2012) Semi-supervised hashing for large scale search. IEEE Trans Pattern Anal Mach Intell 34(12):2393–2406

Wang J, Liu W, Sun AX, Jiang YG (2013) Learning hash codes with list-wise supervision. In: Proceedings of IEEE international conference on computer vision, pp 3032–3039

Wang Q, Zhang Z, Si L, Lafayette W (2015) Ranking preserving hashing for fast similarity search. In: Proceedings of ACM international joint conference on artificial intelligence, pp 3911–3917

Wang D, Gao X, Wang X, He L (2015) Semantic topic multimodal hashing for cross-media retrieval. In: Proceedings of ACM international joint conference on artificial intelligence, pp 3890–3896

Wang Z, Duan L, Lin J, Wang X, Huang T, Gao W (2015) Hamming compatible quantization for hashing. In: Proceedings of ACM international joint conference on artificial intelligence, pp 2298–2304

Wang J, Liu W, Kumar S, Chang SF (2016a) Learning to hash for indexing big data: a survey. Proc IEEE 104(1):34–57

Wang W, Yang X, Ooi BC, Zhang D, Zhuang Y (2016b) Effective deep learning based multi-modal retrieval. Int J Very Large Data Bases 25(1):79–101

Wang J, Zhang T, Sebe N, Shen HT (2018) A survey on learning to hash. IEEE Trans Pattern Anal Mach Intell 40(4):769–790

Weber R, Schek HJ, Blott S (1998) A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In: Proceedings of international conference on very large data bases, pp 194–205

Weiss Y, Torralba A, Fergus R (2008) Spectral hashing. In: Proceedings of neural information processing systems conference, pp 1753–1760

Wolf L, Hassner T, Maoz I (2011) Face recognition in unconstrained videos with matched background similarity. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 529–534

Wu B, GhanemB (2016) lp-box admm: a versatile framework for integer programming. arXiv preprint arXiv: 1604.07666

Xia R, Pan Y, Lai H, Liu C, Yan S (2014) Supervised hashing for image retrieval via image representation learning. In: Proceedings of ACM AAAI conference on artificial intelligence, pp 2156–2162

Xiao J, Hays J, Ehinger KA, Oliva A, Torralba A (2010) Sun database: large-scale scene recognition from abbey to zoo. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 3485–3495

Xie L, Hong R, Zhang B, Tian Q (2015) Image classification and retrieval are onE. In: Proceedings of ACM international conference multimedia retrieval, pp 3–10

Xu B, Bu J, Chen C, Cai D, He X, Liu W Luo J (2011) Efficient manifold ranking for image retrieval. In: Proceedings of ACM international conference on research and development in information retrieval, pp 525–534

Yang Y, Newsam S (2010) Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of international conference on advances in geographic information systems, pp 270–279

Yang L, Hua X, Cai Y (2013) Searching for Images by Video. Int J Multimed Inf Retr 2(3):213–225

Yu F, Ji R, Tsai M-H, Ye G, Chang SF (2012) Weak attributes for large-scale image retrieval. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 2949–2956

Zhang D, Li W (2014) Large-scale supervised multimodal hashing with semantic correlation maximization. In: Proceedings of AAAI conference on artificial intelligence, pp 2177–2183

Zhang Q, Qiu G (2015) Bundling centre for landmark image discovery. In: Proceedings of ACM international conference multimedia retrieval, pp 179–186

Zhang L, Rui Y (2013) Image search—from thousands to billions in 20 year. ACM Trans Multimed Comput Commun Appl 9(1):36–55

Zhang X, Zhang L, Shum HY (2012) Qsrank: query-sensitive hash code ranking for efficient neighbor search. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 2058–2065

Zhang L, Zhang Y, Tang J, Lu K, Tian Q (2013) Binary code ranking with weighted hamming distance. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1586–159

Zhang Q, Fu H, Qiu G (2013) Tree partition voting min-hash for partial duplicate image discovery. In: Proceedings of IEEE international conference on multimedia and expo, pp 1–6

Zhang P, Zhang W, Li WJ, Guo M (2014) Supervised hashing with latent factor models. In: Proceedings of the international ACM SIGIR conference on research and development in information retrieval, pp 173–182

Zhang R, Lin L, Zhang R, Zuo W, Zhang L (2015a) Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification. IEEE Trans Image Process 24(12):4766–4779

Zhang S, Yang M, Wang X, Lin Y, Tian Q (2015b) Sematnic-aware co-indexing for image retrieval. IEEE Trans Pattern Anal Mach Intell 37(12):2573–2587

Zhang L, Shum HPH, Shao L (2017) Manifold regularized experimental design for active learning. IEEE Trans Image Process 26(2):969–981

Zhong LW, Kwok JT (2012) Convex multi task learning with flexible task clusters. In: Proceedings of international conference on machine learning, pp 41–48

Zhou J, Ding G, Guo Y (2014) Latent semantic sparse hashing for cross-modal similarity search. In: Proceedings of ACM international SIGIR conference on research and development in information retrieval, pp 415–424

Zhuang Y, Liu Y, Wu F, Zhang Y, Shao J (2011) Hypergraph spectral hashing for similarity search of social image. In: Proceedings of ACM international conference on multimedia, pp 1457–1460

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.