# Multimedia Indexing, Search and Retrieval in Large Databases of Social Networks

Theodoros Semertzidis, Dimitrios Rafailidis, Eleftherios Tiakas, Michael G. Strintzis and Petros Daras

**Abstract** Social networks are changing the way multimedia content is shared on the web, by allowing users to upload their photos, videos, and audio content, produced by any means of digital recorders such as mobile/smart-phones, and web/digital cameras. This plethora of content created the need for finding the desired media in the social media universe. Moreover, the diversity of the available content, inspired users to demand and formulate more complicated queries. In the social media era, multimedia content search is promoted to a fundamental feature towards efficient search inside social multimedia streams, content classification, context and event based indexing. In this chapter an overview of multimedia indexing and searching algorithms, following the data growth curve is presented in detail. The chapter is thematically structured in two parts. In the first part pure multimedia content retrieval issues are presented, while in the second part, the social aspects and new, interesting views on multimedia retrieval in the large social media databases are discussed.

## 1 Introduction

Social networking sites enabled multimedia content sharing in large volumes, by allowing users to upload their photos, videos, and audio data, produced by any means of digital recorders. Moreover, the huge volumes of information in the social me-

T. Semertzidis e-mail: theosem@iti.gr and M.G. Strintzis e-mail: strintzi@eng.auth.gr
are with the Information Processing Laboratory, Electrical and Computer Engineering Dept., Aristotle University of Thessaloniki, Greece, and the Information Technologies Institute, Centre For Research and Technology Hellas, Thessaloniki, Greece.

D. Rafailidis e-mail: drafail@iti.gr, E. Tiakas e-mail: tiakas@csd.auth.gr and P. Daras e-mail: daras@iti.gr
are with the Information Technologies Institute, Centre For Research and Technology Hellas, Thessaloniki, Greece.

dia inspired users to formulate new types of queries that pose complex questions to these heterogeneous databases.

An attempt to index and retrieve multimedia content shared through the social media, using techniques of the Content Based Multimedia Retrieval community, shows clearly the inability to do so. The parameters and constraints posed from the social media aspect reformed the multimedia indexing and retrieval processes in a new problem seeking for new solutions. The major problems of social multimedia indexing and retrieval exist due to a) the enormous volumes of information that push existing techniques to their edges; b) the well known semantic gap [55] between the low-level multimedia descriptors and the higher level concepts that exist in each multimedia content. These facts do not imply that previous knowledge and tools are totally useless, rather that they should be used in a different way.

This chapter aims to structure the concept of multimedia indexing, search and retrieval, based on the data growth curve from the small, locally stored, multimedia collections to the huge, heterogeneous and context-rich social multimedia collections and to present interesting works along the way.

Multimedia indexing, search and retrieval is a multi-step process that deeply depends on the content type and its characteristics. The typical content-based multimedia indexing (CBMI) and retrieval methods apply the well studied query by example (QBE) paradigm, where a multimedia (MM) object is used as a query to retrieve similar multimedia objects (See Fig. 1). In this chapter we mainly focus on image indexing, search and retrieval, yet the concepts, workflow and conclusions are valid also for other multimedia objects such as video, audio, 3D, etc. A common initial step of multimedia indexing is the extraction of characteristic features from the content in order to describe it in a more compact and discriminative manner. A plethora of works has been published on the field, and many robust and well evaluated multimedia content descriptors exist that encode dominant features such as color information, texture and edge, spectral characteristics, motion, etc (e.g. SIFT, Self-Similarity, CEDD [41, 40, 79, 80]) targeted to specific application areas [78]. The next step is to define a distance function (such as L1, L2, Mahalanobis, etc.) between these descriptors in order to compare the similarity between multimedia objects. As it is obvious, when volumes of data increase dramatically (as in the case of social media data) this part of the process becomes extremely time-consuming to be performed in real-time. Indexing structures (discussed in Sect. 2) are data structures aiming to reduce comparisons and consequently reduce the search time. However, content based multimedia retrieval restricts user queries to the QBE approach which is not sufficient for the environment of the large social multimedia databases for various reasons. The responsiveness of the services in a timely manner, the inability to map with the *subjective* user semantics to enhance the quality of the retrieved results and the new, complex types of user queries are some of them. In the context of large Social Multimedia databases, the available social metadata are used to give answers to these challenging problems.

Sect. 2 and 3 present in a compact way, indexing structures aiming to address the problem of searching inside large collections of multimedia objects. Sect. 2 discusses multidimensional indexing structures, by classifying them either in exact or

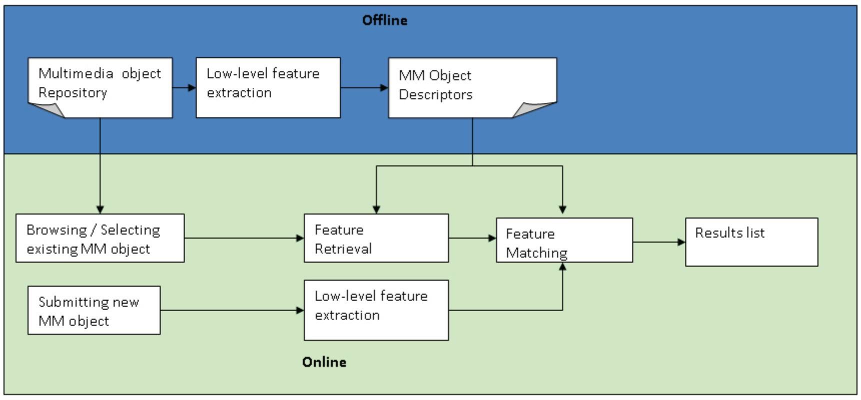


**Fig. 1** A typical content based multimedia retrieval process.

approximate approaches. Next, in Sect. 3, a use case of 1 Million images from the Flickr collection is examined to evaluate indexing in both exact and approximate approaches. In these two sections, the aim of the algorithms is to reduce search time and storage/memory overhead while keeping accuracy as close as possible to the baseline which is the exhaustive search. Exhaustive search is the process of comparing the query descriptor vector with the descriptor vectors of all multimedia objects of a database, i.e. one-to-one comparison without using any indexing structure. Although this gain is very important for the volumes of real world databases such as the social multimedia databases, the other major issue of bridging the semantic gap between the low-level multimedia descriptors and the concepts included in the multimedia objects remains unsolved. Towards these objectives, a new area of multimedia computing that clearly takes into consideration the social aspects of the social media data, has recently emerged [66, 65]. Sect. 4 examines exactly this part of social multimedia indexing and retrieval and categorizes the methods presented in three subsections: a) context-based multimedia retrieval, where the contextual information stored in social media is used to semantically enhance the retrieved results and thus improve the overall accuracy; b) event-based indexing, as another way of indexing multimedia content in a more user-oriented conceptualization; and c) time-related multimedia indexing and search for evolving social multimedia collections. Finally, Sect. 5 discusses the conclusions drawn from this work and presents future challenges towards efficient social multimedia indexing, search and retrieval in the ever growing social multimedia collections.

## 2 Content-based Multimedia Indexing

The availability of multimedia in the large databases of social networks emerged the imperative need to address the challenge of content-based searching, where users pose multimedia objects as queries, in order to find relevant content (See Fig. 1). However, exhaustive searching is infeasible for the large scale applications of social networks due to the extensive time consumption it requires. Thus, the large databases of social networks should be supported by indexing schemes which are able to provide: (a) low space requirements for storing the multimedia content within the indexing scheme; (b) efficient search time; and (c) high retrieval accuracy. Nevertheless, multimedia objects like compressed images, video and audio streams are usually described by sequences of descriptor vectors with over than a thousand dimensions. In this high dimensional space, the performance of existing indexing schemes deteriorate significantly, since content-based similarity search in high dimensions ($\geq$ 1000) is challenging, due the well known problem of Dimensionality Curse [4, 10]. In order to address the aforementioned challenges, the existing indexing schemes are divided into two main categories: those that follow (a) exact and (b) approximate similarity search strategies.

The family of exact similarity search, despite achieving identical retrieval accuracy to exhaustive search, fails to support the high dimensionality. Meanwhile, storage space and search time are dramatically increased. The family of indexing schemes that follow the approximate search strategy, despite reducing the space and search time requirements, fails to preserve the retrieval accuracy of the exhaustive search.

## *2.1 Exact Similarity Search*

In the family of exact similarity search indexing schemes, a state-of-the-art approach is the M-Tree and its variants [29, 15, 14]. The most efficient way to construct the M-Tree is using the bulk-load method [15]. The M-Tree structure manages the query processing according to the distances between multimedia objects, which are stored as nodes. Additionally, the M-Tree has been further extended, in order to support both exact and approximate strategies, while preserving the same indexing structure of the M-Tree file. Furthermore, M-Trees have been introduced to perform indexing and searching, not only in content-based multimedia applications, but also in other similarity search applications, due to their dynamic ability to support insertions and deletions efficiently [11]. Recently, M-Trees were evaluated in distributed environments, in order to support web-scale applications [3]. The idea was to build small M-Trees in each node of the distributed environment and perform the similarity search strategy to all relative nodes.

Moreover, a plethora of alternative exact similarity search indexing schemes have been proposed in the literature. The most important of them are presented in Table 1. However, in these schemes there are several constraints. In particular, the family of BKT, FQT, FHQT, FQA is constrained to metric spaces derived by a distance measure that necessary returns discrete values. In case that a continuous distance function is applied or a large amount of different discrete values are returned, it is infeasible to exploit these indexing schemes, as explained in [11]. The rest of the aforementioned methods support continuous distances, applicable to general metric spaces. However, the dynamic capabilities of insertions and deletions, and the input/output (I/O) cost must also be considered. The VPT, MVPT, VPF methods support latter insertions insufficiently [11]. Additional and more complicated problems appear in methods like GHT, BST, VT, GNAT, VPT, MVPT, VPF, PMT, NNG for deletion operations [11].

**Table 1** Exact similarity search indexing schemes.

| Method Name | Abbreviation | Reference |
|---|---|---|
| R-Tree | R-Tree | [21] |
| KD-Tree | KD-Tree | [5] |
| Quad-Tree | Quad-Tree | [36] |
| Burkhard-Keller Tree | BKT | [7] |
| Fixed Queries Tree | FQT | [1] |
| Fixed Height FQT | FHQT | [2] |
| Fixed Queries Array | FQA | [10] |
| Vantage Point Tree | VPT | [12] |
| Multi-Vantage Point Tree | MVPT | [8] |
| Vantage Point Forest | VPF | [45] |
| Bisector Tree | BST | [31] |
| Generalized Hyperplane Tree | GHT | [6] |
| Geometric NN Access Tree | GNAT | [9] |
| Voronoi Tree | VT | [17] |
| Pivoting M-Tree | PMT | [38] |
| Nearest-neighbor Graphs | NNG | [39] |

Therefore, the M-Tree is the unique exact indexing scheme, which supports dynamic operations efficiently [11]. The M-Tree and its variants have been designed specifically for secondary memory operations and can be balanced to maintain the I/O cost low, compared with the aforementioned exact similarity search indexing schemes. However, due to the problem of Dimensionality Curse [4, 10], M-Trees are transformed to high level trees with many internal nodes, resulting in enormous increase of the I/O cost, unsuitable for performing efficient content-based search and retrieval in databases with billions of records such as the ones in Social Networking Sites.

## *2.2 Approximate Similarity Search*

In the family of approximate methods, a state-of-the-art approach is the Locality Sensitive Hashing (LSH) [20] and its variants, which are used for indexing of high dimensional data for multimedia search and retrieval. The basic idea of LSH is (a) to encode the distances between the multimedia objects into the form of compressed sequences of bits, while using hash functions, and (b) to store the encoding distances into tables, in order to ensure that the probability of collision is much higher for multimedia objects that are close to each other than those that are far apart. Therefore, the LSH-based indexing schemes vary according to the respective hashing function, trying to reduce the search time, while maximizing the retrieval accuracy, by minimizing the approximation error. Thus, the LSH variants are categorized in data dependent [16, 13, 34, 43, 25, 33, 22, 37] and data independent ones [32, **?**, 23, 35, 26], where efficiency improvements of data dependent methods over independent ones have been proved in several studies [23, 44, 35].

Alternative approximate techniques have also been introduced in the literature, such as the Spatial Approximation Tree (SAT) [30], the Approximating Eliminating Search Algorithm (AESA) [42] and the Linear Approximating Eliminating Search Algorithm (LAESA) [28]. The main advantage of these approximate methods is the significant reduction of search time, while their main disadvantage is the low retrieval accuracy. Moreover, such approximate methods require a time consuming preprocessing step of the multimedia content, in order to increase the retrieval accuracy [27]. Additionally, these methods are not able to efficiently support dynamic changes of the insertion and deletion operators, since for each change a full preprocessing step is required [11]. For example, SAT does not support insertions and consequently, the whole indexing structure has to be built from scratch [11]. Despite the fact that AESA and LAESA support dynamic operations, during their search approach all disk pages of the indexing structure have to be read, which results in limited I/O performance and consequently, the search time is highly increased [11]. Therefore, research has been focused on LSH-based approaches, which are able to support both dynamic operations and efficient search time. However, their achieved retrieval accuracy is rather low.

## 3 Use case: Flickr's 1 Million Images

In this Section we present a case study of Flickr's 1 Million image dataset[1] [18, 19]. The photo sharing website Flickr has over 6 billion images and is a representative example of the large scale problem in multimedia indexing, search and retrieval in the large databases of social networks.

In order to build the evaluation dataset, a recent variant of SIFT descriptors [40] was used. Consequently, several collections of descriptor vectors were constructed, by varying the number of dimensions from 64 to 1024, where the resulted datasets are denoted by: SIFT-64dim, SIFT-128dim, SIFT-256dim, SIFT-512dim, SIFT-1024dim, respectively. All collections were indexed by the exact approach of M-Tree and the approximate approach of Locality Sensitive Hashing, since both methods are superior over other indexing schemes in exact and approximate similarity search approach, respectively.

[1] For further details visit ImageCLEF 2011, the "Visual Concept Detection and Annotation" task.

However, in the case of M-Tree, it is infeasible to preprocess the 1M SIFT-1024dim dataset for any parameter combination, like node size, utilization, split strategy, etc. [15]. Therefore, we measured the corresponding results similar to the case of 100K multiplied by 10, assuming that 10 M-Trees are built, by splitting the SIFT-1024dim dataset into equal size datasets of 100K. For the remaining datasets we report directly the performance of M-Tree, by identifying the optimal parameter selection.

Moreover, in the case of LSH, we varied the number of hash tables $L$, to achieve the maximum retrieval accuracy, while preserving the search time below the respective search time of exhaustive search. Therefore, we concluded that in the case of SIFT descriptors, the maximum number of $L$ hash tables equals 2, since further increase results in exceeding the search time of exhaustive search. Additionally, in order to encode the multimedia distances, hash keys of 1024 bits were used, resulting in the maximum retrieval accuracy of LSH.

In order to demonstrate and identify the aforementioned challenges, as presented in Sect. 2, three respective experiments were conducted, concerning: (a) space requirements; (b) search time; and (c) retrieval accuracy.

Firstly, in Table 2 we present the construction requirements for (a) the exact method of the M-Tree family and (b) the approximate method of the LSH family. The column "exhaustive" denotes the case of performing exhaustive search and consequently, an indexing scheme is not required. Therefore, the required disk space is equal to the size of each dataset. Based on the experimental results shown in Table 2, we conclude to: (a) the M-Tree indexing scheme requires a significant amount of space, 10 times greater than the corresponding dataset space, since high level trees are constructed, consisting of many internal nodes and (b) LSH requires an important amount of additional space for storing the constructed hash tables, linked to buckets, in which the IDs of the corresponding multimedia are stored.

**Table 2** Disk Space Requirements in GB.

| **SIFT-dataset** | **Exhaustive** | **M-Tree** | **LSH-1L** | **LSH-2L** |
|---|---|---|---|---|
| SIFT-64 dim | 0.238418579 | 2.088517205 | 0.556949615 | 0.871504784 |
| SIFT-128 dim | 0.476837158 | 4.239689925 | 0.795368195 | 1.109923363 |
| SIFT-256 dim | 0.953674316 | 8.097807758 | 1.272205353 | 1.586760521 |
| SIFT-512 dim | 1.907348633 | 15.95268128 | 2.225879669 | 2.540434837 |
| SIFT-1024 dim | 3.814697266 | 30.94820169 | 4.133228302 | 4.44778347 |

Next, we evaluate M-Tree and LSH against the exhaustive search approach, in terms of search CPU time and disk accesses, by varying the size of the SIFT-1024dim dataset. For measuring the search time, CPU and I/O time are separately reported, following either a memory-based or a disk-based approach. Since, the I/O time is completely dependent on (a) the running operating system; (b) memory/disk cache; and (c) hard disk specifications, the corresponding disk accesses (DA) are measured, assuming that a disk page is equal to 4KBytes. Moreover, to produce the

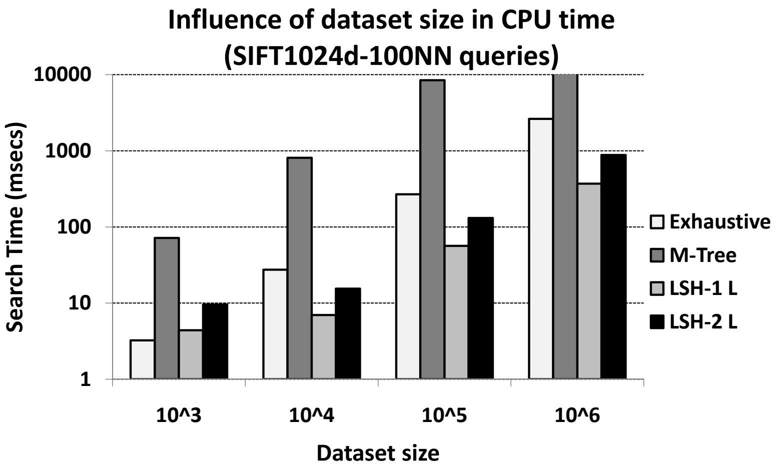


**Fig. 2** Search CPU time versus dataset size for 100-NN queries in SIFT-1024dim dataset.

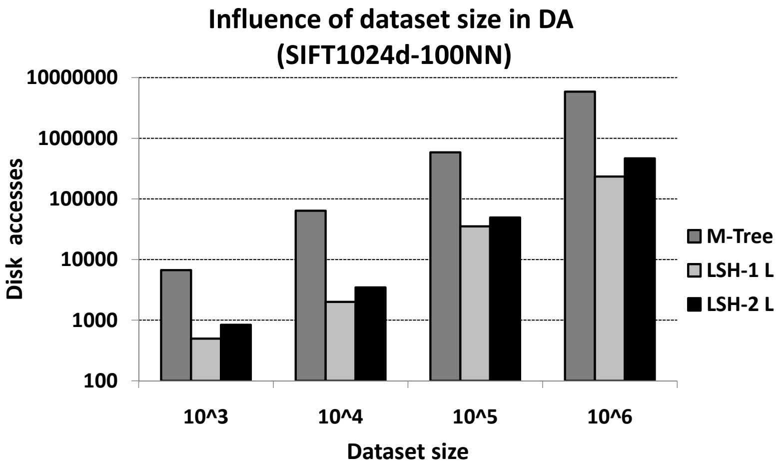


**Fig. 3** Search disk accesses versus dataset size for 100-NN queries in SIFT-1024dim dataset.

different datasets, $10^3, 10^4, 10^5$ images were randomly selected from the initial $10^6$ SIFT-1024dim dataset. In Fig. 2 and 3, the respective results are presented (note that the exhaustive search method is omitted from Fig. 3, since the disk-level implementation is not feasible without using an external indexing scheme). We can make the following observations: (a) in each method, search CPU time and disk accesses are increased with respect to the dataset size; (b) the M-Tree requires significantly high search times and disk accesses, even higher than the respective times of the exhaustive approach, verifying the Dimensionality Curse problem faced by the family of exact similarity search methods; and (c) LSH has high performance, since the respective search time is highly reduced.

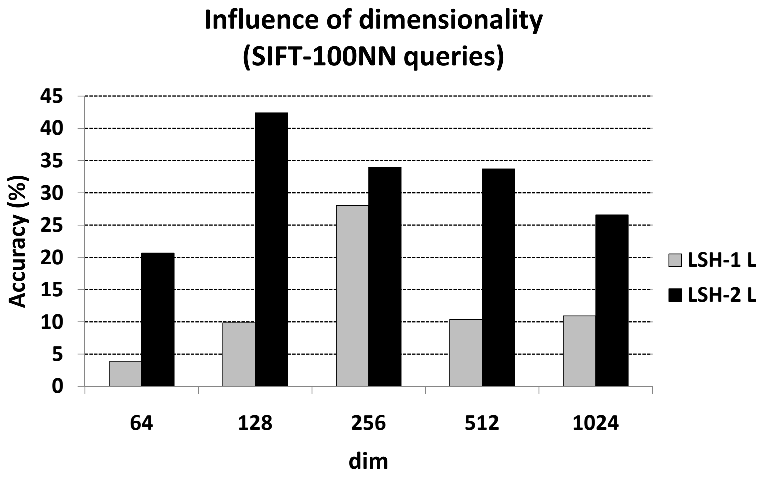


**Fig. 4** LSH' retrieval accuracy versus dimensionality for 100-NN queries in SIFT datasets.

Finally, in Fig. 4, we evaluate the retrieval accuracy of LSH, by performing 1000 top-100 queries (denoted by 100-NN queries) and varying the dimensionality of the SIFT datasets. The retrieval accuracy is measured according to the ratio of the top-$k$ results retrieved by the exhaustive search over the top-$k$ results retrieved by the proposed indexing method. Based on the experimental results in Fig. 4, an important observation is that LSH, despite using the maximum number of hash tables ($L = 2$), achieves retrieval accuracy below 40% in all datasets. The low accuracy of LSH is affected by the poor encoding of the multimedia distances. Additionally, by increasing the number of dimensions, LSH's accuracy is reduced. Note that the comparison with M-Tree is omitted, since M-Tree belongs to the family of exact similarity search and its retrieval accuracy is always equal to 100%.

Summarizing our conclusions, in the case study of Flickr's 1 million images, the family of exact similarity search, despite achieving identical retrieval accuracy to exhaustive search, fails to support the high dimensionality and as a consequence, storage space and search time are dramatically increased. Moreover, the approximate search strategy of LSH, achieves to reduce the search time requirements. Nevertheless, there is no analogous progress in terms of retrieval accuracy, since LSH fail to preserve the retrieval accuracy of the exhaustive search.

Clearly, this case study revealed the fragility of multimedia retrieval in large volumes of data at the scale of millions of records. However, the amounts of multimedia objects that should be indexed and retrieved in the large social multimedia databases such as Flickr and youTube are in the scale of dozens of billions making it extremely challenging. Moreover, the retrieved results in the presented multimedia retrieval tasks, did not preserve any of the "subjective" user semantics that were made available through the sharing process in social networking sites.

## 4 The Social Media Era

When social networking sites enabled users to share and publish their content online, multimedia search and retrieval became one of the most important and desired features on the Web. However, the volumes of data shared, introduce new challenges in multimedia retrieval. Some recent YouTube statistics [47] show clearly that the existing multimedia indexing and retrieval methods are not adequate for these volumes of information. According to YouTube, 48 hours of video are uploaded every minute. This is approximately 8 years of content every day or the equivalent of 240,000 full-length films every week. The video uploaded on YouTube per month exceeds the content created by the 3 major US networks in the last 60 years. In August 2011 Flickr reported that it reached more than 6 billion uploaded images and the number continuous to grow steadily [48, 46].

Moreover, the amount and diversity of metadata collected and shared through these enormous social media collections pose more parameters to consider towards efficient indexing and retrieval. In a first evaluation, this extra information increases the complexity of the retrieval tasks dramatically. However, the differentiation of the metadata sources (user tags, sensors' information, social graph relations etc.) construct a rich environment that helps to narrow down these sets to manageable clusters of information.

All these numbers and facts reveal that multimedia usage and applications changed drastically on the social media era, thus revolutionary multimedia indexing, search and retrieval approaches are needed. What we should clearly state here is that not only did the multimedia collection change shape and size but also the users' needs and goals evolved through the available web applications.

In the following subsections, recent works are presented roughly classified based on the usage of social metadata and targeted applications. Sect. 4.1 presents some recent works which aim at involving contextual information to semantically enhance the retrieved results. Sect. 4.2 discusses works on the relevantly new approach of recognizing and indexing events recorded in social multimedia content. Finally, Sect. 4.3 discusses the works conducted on time-related multimedia retrieval from social streams and ephemeral collections.

### *4.1 Context-based Multimedia Indexing and Retrieval*

In an environment as wide and heterogeneous as the World Wide Web is, contextual information is known to be inherently noisy, subjective and ambiguous. However, the information carried out through the context of the various web applications may be overly helpful after applying some filtering and post-processing.

One of the most used and well studied contextual data is the user tags. Users tend to tag content in a very personalized way based on their interests, culture, education, etc., thus the relevance of each tag to the actual content is clearly subjective. In order to build a system able to exploit user tags, so as to enhance multimedia retrieval,

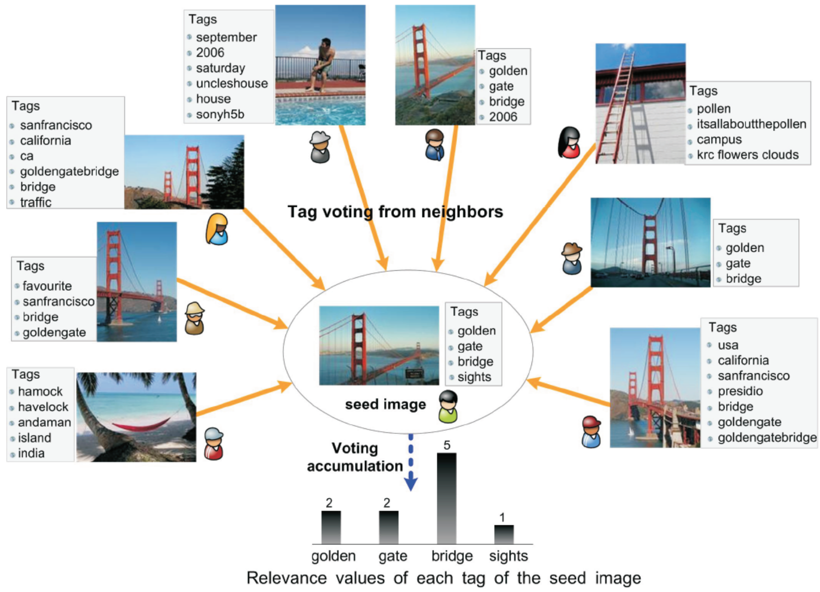


**Fig. 5** Learning tag relevance by neighbor voting. The relevance is computed as the accumulated neighboring votes from visually similar image of the seed image [49].

tags should be found that are relevant to the majority of the users of the system in an objective way [49]. Towards this goal, several methods [50, 51] proposed learning a mapping of low-level visual features to semantic concepts.

Li et al. [49] proposed a technique for learning tag relevance by neighbor voting. The authors rely on the intuition that a relevant tag may be inferred based on the tags of the visual neighbors of that image. The major difference from the related works of [82, 81] is that only common tags between visually similar images are propagated. With this approach no new tags are introduced to the image and thus the technique protects from incorrectly assigning irrelevant tags.

The algorithm reads as following: firstly, top-K visually similar neighbors of the image are found, using common visual features and k-means clustering to divide the dataset into small blocks of clusters. Then, for each neighboring image, only the common tags vote to the examined image tags, i.e. each tag of the examined image accumulates votes from the common tags of the neighbor images (see Fig. 5). Since the approach started with the intuition that common tags from *different* users impose a strong relevance of the tags on the visual content, images in the K-nn set that come from the same user as the examined image are ignored.

For evaluating their method, the authors compiled a database of 1 Million images with tags from Flickr and separated a ground truth set for evaluation. Their experiments were evaluated in a tag-based social image retrieval framework where the well known Okapi BM25 ranking function for text retrieval was used [52]. The authors compiled three different experimental set-ups, aiming to identify different aspects of tag-based multimedia retrieval. In the first experiment, a single word was selected as query and various numbers of neighbors were also selected. In the second experiment, the initial queries were expanded with synonyms using WordNet [53] and an online dictionary (http://dico.isc.cnrs.fr/dico/en/search). Finally in the third experiment the impact of database size was examined by dividing the database in 100K parts and increasing the database by incrementally adding the parts to reach the whole 1M images. Both experiments showed clearly the advantage of learning the relevant tags from the visually similar neighbor images, since there was a significant improvement in retrieval accuracy. However, the most interesting result, in terms of scalability of the method, came from the third experiment showing that search performance (in terms of Mean Average Precision) increases as the database size does.

### 4.1.1 Latent Semantic Spaces

A popular approach used in multimedia indexing and retrieval, casted as clustering and classification, is the extraction of the latent semantics of the explored data to reveal hidden relationships, concepts and possible structures that a human mind would easier understand [70, 69]. Revealing the hidden semantics in data is a well studied research field with some interesting statistical tools available [71, 72, 73]. However, the formation of a problem to fit such a tool and the decisions on the design and the social media data to be used, is a very interesting and challenging task.

Bosch et al. [68] performed scene classification using probabilistic Latent Semantic Analysis (pLSA) [72] on visual vocabularies extracted directly from the images. The visual vocabularies were extracted by quantizing content descriptors using k-means and a Bag-of-word model. The results of the classification showed promising performance in categorizing images, however the algorithm does not take advantage of the available knowledge in social media and thus it is not capable to accurately reveal the semantic concepts. On the contrary, Sizov [67] proposed the GeoFolk model for classification of social media documents using only the contextual information of tags, unstructured text, etc. and spatial knowledge (geotags and geo-coordinates). Moreover, in order to reduce problems such as ambiguity of tags or geolocation information and the sparseness of the available metadata, a model that use both meta-information was developed. Yang et al. [69] proposed Heterogeneous Transfer Learning for image classification using both contextual and content information. Their approach was to extract visual words from the images and additional tags from the social web in order to build their annotation-based pLSA implementation, which showed significant improvement compared to k-means clustering and plain pLSA. The authors introduced aPLSA as a combination of two separate pLSA models, one for image to visual features co-occurrence matrix and one for text feature to visual feature co-occurrence matrix, with the same latent variables.

A very interesting study on the combination of Content and Context information to learn a Latent Semantic Space for use in social multimedia retrieval environments

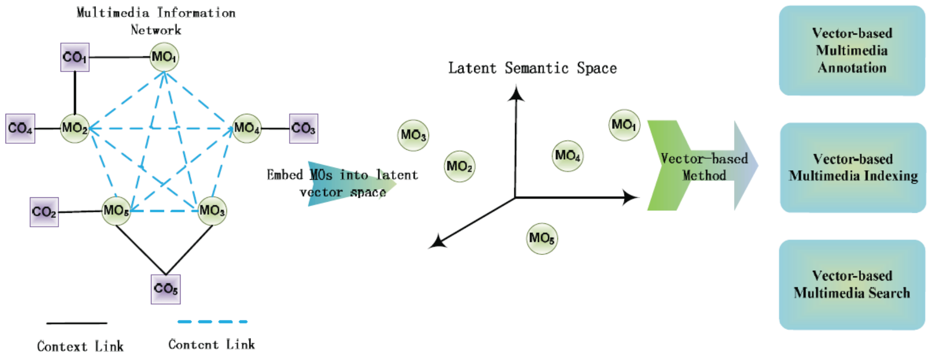


**Fig. 6** Learning Latent Semantic Space from a content and context links.[54]

is the work of Guo-Jun Qi et al. [54]. Most of the works in the field, exploit social metadata (tags, geolocation etc) or content features to learn latent spaces, however they have not addressed problems, directly inherited from the use of content or context data sources. Moreover, the sparsity of the annotated objects (contextual information) is one of the major problems that machine learning algorithms suffer from. In their work, Qi et al. [54] aim to address the metadata sparsity problem. They present the multimedia resources in the form of Multimedia Information Networks with two types of objects - multimedia and context objects linked together (see Fig. 6). While the content similarity links are important for retrieval, the context links are the ones that bring quality to the retrieved results. The proposed algorithm learns a latent space where content and context information is encoded and mapped to it. Then, the multidimensional vectors describing the multimedia objects can be indexed, classified and retrieved with common vector-based methods as the ones described in Sect. 2. The authors make the assumption that similar multimedia objects should be closer to each other in the latent semantic space. This assumption acts as a regularization factor to avoid overfitting problems derived from the sparse context links. To evaluate the algorithm, a database of Flickr images along with their tags was used. A comparison of the different multimedia retrieval schemes was conducted to show the promising results of the approach. The tested retrieval schemes were a) content based multimedia retrieval (CMR); b) context based multimedia retrieval (CxMR); and c) both content and context multimedia retrieval (C2MR). 81 concepts were manually defined in the database and experiments were formulated as multimedia annotation problems. As a performance metric, Average Precision (AP) was selected to measure the retrieval accuracy of each concept. A supervised and an unsupervised method of the model were presented (S-C2MR, U-C2MR) and both had clear improvements against CMR and CxMR. The U-C2MR method improved CMR results by 246.8% and 37.6% CxMR, while S-C2MR improved CMR by 264.2% and CxMR 44.5%.

## 4.2 Event-based Multimedia Indexing and Retrieval

Event detection from web data has attracted a lot of research attention recently [62, 63, 64] due to the immense amount of available information and desire of users to extract/exploit structured information. Moreover, a relatively new approach in detecting, identifing and indexing social events [56, 57, 58, 59], is through the usage of social metadata along with the shared multimedia content. Towards bridging the semantic gap between human perception and plain multimedia analysis, social multimedia researchers developed methods to detect and link events to multimedia objects in order to support a more human-centered retrieval process and new query types. Since humans tend to structure their knowledge and memory based on specific events and experiences, event identification and indexing should become a realistic way to retrieve multimedia content that we perceive as relevant.

Users aiming to decide weather they will attend a concert/show, are interested to feel the atmosphere of previous events based on images and videos available in social networks. In [58], Liu et al. use content and metadata information from Flickr, Last.fm, Eventful and Upcoming to identify events (concerts, shows, etc.) in 9 preselected venues, in order to present characteristic content to interested users. Their work is a two step process. In the first step they measure photo sharing activity in known venues, to detect an event occurrence. As a prior knowledge for this measure a bounding box of a venue geo-location is used. To extract this information, the authors use GPS information from Last.fm and Flickr metadata so as to discard any other information that was recorded outside this area.

This approach achieved to reduce the initial collection of images about the venues to only 4604 geo-tagged images from the huge collection of Flickr images. Next, they collected more relevant photos, by querying Flickr with each venue name to finally build a photo collection of approximately 9 thousand images in total. By tracking the number of shared images per day, number of owners and the product of the previous two, the event days were identified with appropriate thresholding. The next step of the process was to use seeds images (representative images) that were explicitly linked to an event (in the events database), to retrieve visually similar images from Flickr. In order to calculate visual similarity, low level visual descriptors such as color moments, gabor texture and edge histograms were used. Note here that the search space was largely reduced due to the first step of keeping only images that were inside the venue bounding box or had venue name as their tag.

In [59], Becker et al. discussed also the problem of event detection and identification by learning similarity metrics and cast it as a clustering problem. The features used for clustering the social media objects to events' clusters, were context features such as tags, descriptions, time/date, location, etc. Moreover, they used an *all-text* feature where all the textual representation of document features are included (title, description, tags, time/date, location). The textual features were used as $tf \cdot idf$ weight vectors, while typical text processing (stop-word elimination, stemming) steps were applied when needed. In order to support a scalable clustering approach, they proposed a single-pass incremental clustering and tested it into two different scoring cases. In the first case, each examined document was compared to every other in the cluster so as to generate an overall score, while in the second approach only a document to cluster centroid comparison was performed. For the selection of similarity metrics ensemble-based similarity and classification based similarity were also examined for the final results. The experimental results showed that *All-text* individual clusters outperformed the other clusters while the similarity based combination outperformed the individual clusters (include *All-text* ). Overall, the classification method showed significant improvement over the typical text-based similarity approaches.

The work of Papadopoulos et al. [60] approached event detection in social multimedia as a graph-based image clustering problem. Their approach combined visual similarity along with tag-based similarity to build two image similarity graphs which are then merged to a unified hybrid image similarity graph. Then, a community detection algorithm was applied to detect clusters of similar images in the hybrid image similarity graph. The authors examined two different cases of image clusters, which are commonly found in social multimedia sharing sites such as Flickr. These are “Landmarks” and “Events”. For classifying the image clusters as Events or Landmarks, four features were used. The first two, introduced in Quack et al. [61], are the duration of the cluster in terms of creation timestamps of the included images and the ratio of owners over the number of the images in the cluster. According to the authors these two features were not adequate to discriminate efficiently “Landmark” clusters from “Event” clusters. Additionally, they also created two tag vectors corresponding to each class and then removed the common tags to build class-specific tag vectors. Finally, the other two features were the counts of tags of the images clusters that belong to the one set of the other. With the classification step, the Event detection task was finished. However, in the case of Landmark detection, another step was also required. Although the method aimed to cluster the image collections in meaningful groups, the authors observed that some of the Landmark clusters referred to the same object. Towards facing this inefficiency, a merging step was applied. By using geolocation information, a new spatial proximity graph was built and the community detection algorithm was used to form new clusters of images. The experimental results compared with k-means clustering of both the visual and the tag features appear to have better geo-spacial focus (which is significant to events and landmarks) and overall higher precision in the subjective evaluation.

### *4.3 Time-related multimedia indexing and search for evolving social multimedia collections*

The speed of multimedia content sharing in the Social web, bloats the Social media databases with enormous volumes of data in very short time windows. Thus, a major difference between the Social web databases and the common multimedia databases, is the fact that they are constantly increasing, populated with fresh content. This feature inspires users to ask for complicated queries that include time/date related information. Moreover, these ever evolving databases store millions of records every day, with ephemeral interest to the users and useless if not consumed in a certain period of time. Thus, queries that filter the retrieved content in specified time windows are also needed. However, the time-evolved social multimedia databases require also new approaches of organising the content in order to enable for efficient search and retrieval.

The study of Lin et al. [74] addressed the social multimedia retrieval problem from exactly this viewpoint. The authors consider Flickr's photo groups as mixtures of themes with similarities in content and context. Their goals were: a) to better organize the content inside each photo group, since the exponential growth of the content make exploration and searching inside a group a challenging task; and b) to reveal the changing interests and trends inside the photo group and reveal the photo genres that a group contains. In order to exploit both content and context information, the authors extracted content features from images, tags, owner information, and post time to build four matrices: a photo-features matrix, a photo-user matrix, a photo-tag matrix and a photo-time matrix. Then, a non-negative joint matrix factorization procedure was applied on these matrices to extract "themes" of photos that change over time. As they clearly stated, their motivation was the development of a method to answer difficult, for pure multimedia retrieval systems, questions. Such questions were : *are there typical patterns in the photo stream? how these patterns evolve over time? how can we extract the patterns and which users and photos follow them?*. As it is clear, such questions are becoming typical and extremely useful in global-scale databases and as such, they add value to the existing social multimedia sharing services.

Another interesting study on multimedia retrieval from the social web is the work of De Silva et al. [75], which proposed interactive spatio-temporal query formulation for quick multimedia retrieval from large multimedia databases. De Silva et al. stated that the presentation of an one-dimensional results list is inadequate for such rich multimedia databases since the query was tested against multiple dimensions (content, tags, time, space, etc.). The proposed interface enables the iterative searching and browsing of content with query refining in any of the available modalities (content, social relations, time, etc.).

Since temporal information is widely available in the shared multimedia content through social media, new approaches emerge exploiting the temporal information to extract usage and sharing patterns or visualize the content and extract valuable information that may be used to cluster multimedia content such as the works of [76, 77].

## 5 Conclusions and Future Challenges

Multimedia indexing and retrieval is a challenging task on its own, and thus different solutions have been proposed, trying to address different angles of the problem. Further, Social Multimedia indexing and retrieval in the large databases of the social networks, advanced the challenges to form a new problem that needs special handling. Multimedia indexing, search and retrieval in the large databases of social networks clearly stated its "uniqueness" mainly in two dominant axes: a) as multimedia analysis task with heterogeneous, noisy and ambiguous modalities such as text, images, videos, audio along with tags, free-text, geotags, geo-coordinates, time information, social relations and communities; and b) as a web-scale information retrieval task with all the scalability and performance issues carried along.

The majority of the works presented in this chapter were evaluated using Flickr images. The Flickr image sharing service has a characteristic, socially "sound", design that enables the evolution of the database in terms of time, themes, groups, trending tag annotations and of course users that form communities and groups, follow other users' works and give ratings.

In this chapter, Social Multimedia indexing, search and retrieval techniques and algorithms were presented, aiming to shed light to different viewpoints of the problem. Sect 2 discussed shortly the state-of-the-art multidimensional indexing structures by classifying them in the exact and approximate approaches. Then, in Sect. 3, Social Multimedia content was used to present a case study of indexing 1 Millions images from Flickr photo sharing site. In this case study, indexing was performed based only on the content of the multimedia objects. This approach showed that in the large social multimedia databases of billions of records, these approaches are inefficient in terms of response time and memory/storage needs and/or accuracy of the retrieved results. The major issue though, is that such indexing methods do not consider the available social metadata that enclose "subjective" user semantics. However, these semantics are crucial for improving both the quality of the retrieved results and the performance in qualitative aspects. Sect. 4 presented exactly these aspects of social multimedia retrieval. Moreover, in the context of social media, multimedia retrieval was explored as another means of a higher level query formulation to answer complex questions.

Since the social multimedia databases will continue to grow exponentially to unmanageable volumes, revolutionary approaches in content search and retrieval are necessary. Future challenges to this field include: searching in multimedia streams, classification of ephemeral data for subscription purposes and trending algorithms for identifying popular multimedia content. Moreover, algorithms, techniques and search schemes that enable users to improvise in querying the enormous social multimedia collections are also sought.

## 6 Acknowledgements

This work was partially supported by the EC FP7 funded project CUBRIK, ICT-287704 (www.cubrikproject.eu).

## References


1. R. Baeza-Yates, W. Cunto, U. Manber, and S. Wu, "Proximity matching using fixed-queries trees", in *Proc. 5th Combinatorial Pattern Matching (CPM), LNCS 807*, 1994, pp. 198-212.
2. R. Baeza-Yates and G. Navarro, "Fast approximate string matching in a dictionary", in *Proc. 5th South American Symposium on String Processing and Information Retrieval (SPIRE), IEEE CS Press*, 1998, pp. 14-22.
3. M. Batko, F. Falchi, C. Lucchese, D. Novak, R. Perego, F. Rabitti, J. Sedmidubsky, and P. Zezula, "Building a Web-scale Image Similarity Search System", *Journal of Multimedia Tools and Applications*, vol. 47, no. 3, pp. 599-629, 2010.
4. R. Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton University Press, 1961.
5. J. L. Bentley, "Multidimensional binary search trees in database applications", *IEEE Transactions on Software Engineering*, vol. 5, no. 4, pp. 333-340, 1979.
6. E. Bugnion, S. Fhei, T. Roos, P. Widmayer, and F. Widmer, "A spatial index for approximate multiple string matching". in *Proc. 1st South American Workshop on String Processing (WSP)*, 1993, pp. 43-53.
7. W. A. Burkhard and R. M. Keller, "Some approaches to best-match file searching", *Communications of the ACM*, vol. 16, no. 4, pp. 230-236, 1973.
8. T. Bozkaya and M. Ozsoyoglu, "Distance-based indexing for high-dimensional metric spaces", in *Proc. ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 1997, pp. 357-368.
9. S. Brin, "Near Neighbor Search in Large Metric Spaces", in *Proc. 21st International Conference on Very Large Data Bases (VLDB)*, 1995, pp. 574-584.
10. E. Chavez, J. Marroquin, and G. Navarro, "Overcoming the curse of dimensionality", in *Proc. European Workshop on Content-Based Multimedia Indexing (CBMI)*, 1999, pp. 57-64.
11. E. Chavez, G. Navarro, R. Baeza-Yates and J. L. Marroquin, "Searching in metric spaces", *ACM Computing Surveys*, vol. 33, no. 3, pp. 273-321, 2001.
12. T. Chiueh, "Content-based image indexing", in *Proc. 20th Conference on Very Large Databases (VLDB)*, 1994, pp. 582-593.
13. O. Chum, J. Philbin, and A. Zisserman, "Near duplicate image detection: min-hash and tf-idf weighting", in *Proc. British Machine Vision Conference*, 2008.
14. P. Ciaccia, M. Patella, and P. Zezula, "M-tree: an efficient access method for similarity search in metric spaces", in *Proc. 23rd Conference on Very Large Databases (VLDB)*, 1997, pp. 426-435.
15. P. Ciaccia, M. Patella and P. Zezula, "Bulk Loading the M-tree", in *Proc. 9th Australasian Database Conference (ADC)*, 1998, pp. 15-26.
16. M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions", in *Proc. Symposium on Computational Geometry*, 2004, pp. 253-262.
17. F. Dehne and H. Nolteimer, "Voronoi trees and clustering problems", *Information Systems*, vol. 12, no. 2, pp. 171-175, 1987.
18. Flickr's 1 Million image dataset, "Visual Concept Detection and Annotation", ImageCLEF, 2011.
19. Flickr's website, [Online], Available: http://www.flickr.com/
20. A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing", in *Proc. 25th International Conference on Very Large Data Bases (VLDB)*, 1999, pp. 518-529.
21. A. Guttman, "R-Trees: A Dynamic Index Structure for Spatial Searching", in *Proc. ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 1984, pp. 47-57.
22. H. Jegou, M. Douze and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search", in *Proc. European Conference on Computer Vision, volume I of LNCS, Springer*, 2008, pp. 304-317.
23. H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 33, no. 1, pp. 117-128, 2011.

24. A. Joly, O. Buisson, “Random Maximum Margin Hashing”, in *Proc. 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 873-880.
25. A. Joly, C. Frelicot, and O. Buisson, “Feature statistical retrieval applied to content-based copy identification”, in *Proc. International Conference on Image Processing*, 2004, pp. 681-684.
26. B. Kulis and K. Grauman, “Kernelized locality-sensitive hashing for scalable image search”, in *Proc. 12th IEEE International Conference on Computer Vision (ICCV)*, 2009, pp. 2130-2137.
27. Q. Lv, W. Josephson, Z. Wang, M. Charikar, and K. Li, “Multi-Probe LSH: Efficient Indexing for High-Dimensional Similarity Search”, in *Proc. 33rd International Conference on Very Large Data Bases (VLDB)*, 2007, pp. 950-961.
28. L. Mico, J. Oncina, and E. Vidal, “A new version of the nearest-neighbor approximating and eliminating search (AESA) with linear preprocessing-time and memory requirements”, *Pattern Recognition Letters*, vol. 15, pp. 9-17, 1994.
29. M-Tree Web Site, “The M-Tree Project”, [Online], Available: http://www-db.deis.unibo.it/Mtree, 2008.
30. G. Navarro, “Searching in metric spaces by spatial approximation”, *The VLDB Journal*, vol. 11, no. 1, pp. 28-46, 2002.
31. H. Nolteimer, K. Verbarg, and C. Zirkelbach, “Monotonous Bisector Trees: a tool for efficient partitioning of complex schemes of geometric objects”, in *Data Structures and Efficient Algorithms, LNCS 594, Springer-Verlag*, pp. 186-203, 1992.
32. L. Pauleve, H. J’egou, and L. Amsaleg, “Locality sensitive hashing: A comparison of hash function types and querying mechanisms”. *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1348-1358, 2010.
33. S. Poullot, O. Buisson, and M. Crucianu, “Z-grid-based probabilistic retrieval for scaling up content-based copy detection”, in *Proc. 6th ACM International Conference on Image and video retrieval (CIVR)*, 2007, pp. 348-355.
34. M. Raginsky and S. Lazebnik, “Locality-sensitive binary codes from shift-invariant kernels”, in *Proc. ACM NIPS*, 2009, pp. 1509-1517.
35. R. Salakhutdinov, A. Mnih, and G. Hinton, “Restricted boltzmann machines for collaborative filtering”, in *Proc. 24th ACM International Conference on Machine learning*, 2007, pp. 791-798.
36. H. Samet, “The quadtree and related hierarchical data structures”, *ACM Computing Surveys (CSUR)*, vol. 16, no. 2, pp. 187-260, 1984.
37. G. Shakhnarovich, T. Darrell, and P. Indyk, *Nearest Neighbor Methods in Learning and Vision: Theory and Practice*, MIT Press, 2006.
38. T. Skopal, “Pivoting M-tree: A Metric Access Method for Efficient Similarity Search”, in *Proc. Annual International Workshop on Databases, Texts, Specifications and Objects (DATESO)*, 2004, pp. 27-37.
39. T. Skopal and D. Hoksza, “Improving the performance of M-tree family by nearest-neighbor graphs”, in *Proc. 11th East European Conference on Advances in Databases and information Systems (ADBIS)*, 2007, pp. 172-188.
40. J.R.R. Uijlings, A.W.M. Smeulders, and R.J.H. Scha, ”Real-Time Visual Concept Classification”, *IEEE Transactions on Multimedia*, vol. 12, no. 7, pp. 665-681, 2010.
41. David G. Lowe, ”Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, 60, 2 (2004), pp. 91-110
42. E. Vidal, “An algorithm for finding nearest neighbors in (approximately) constant average time.”, *Pattern Recognition Letters*, vol. 4, pp. 145-157, 1986.
43. R. Weber, H. J. Schek, and S. Blott, “A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces”, in *Proc. 24th International Conference on Very Large Data Bases (VLDB)*, 1998, pp. 194-205.
44. Y. Weiss, A. Torralba, and R. Fergus, “Spectral hashing”, in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2008, pp. 1753-1760.
45. P. N. Yianilos, “Excluded middle vantage point forests for nearest neighbor search”, *NEC Research Institute, Princeton University*, Technical Report, 1998.

46. Flickr website. http://www.flickr.com/ . Cited 20 Feb 2012
47. YouTube statistics. http://www.youtube.com/t/press_statistics. Cited 20 Feb 2012
48. Flickr record in Wikipedia. http://en.wikipedia.org/wiki/Flickr#cite_note-3 Cited 20 Feb 2012
49. Xirong Li, Cees G.M. Snoek, and Marcel Worring. 2008. "Learning tag relevance by neighbor voting for social image retrieval". *In Proceedings of the 1st ACM international conference on Multimedia information retrieval (MIR '08)*. ACM, New York, NY, USA, 180-187.
50. J. Li and J. Z. Wang. Real-time computerized annotation of pictures.*"TPAMI"*, 30(6):985-1002,2008.
51. K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan. Matching words and pictures. *JMLR*, pages 1107-1135, 2003.
52. K. Sparck Jones, S. Walker, and S. E. Robertson. 2000. A probabilistic model of information retrieval: development and comparative experiments. Inf. Process. Manage. 36, 6 (November 2000), 779-808.
53. Christiane Fellbaum (1998, ed.) WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.
54. Guo-Jun Qi, Charu Aggarwal, Qi Tian, Heng Ji, Thomas Huang, "Exploring Context and Content Links in Social Media: A Latent Space Method," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99, no. PrePrints, , 2011
55. Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. 2000. Content-Based Image Retrieval at the End of the Early Years. IEEE Trans. Pattern Anal. Mach. Intell. 22, 12 (December 2000), 1349-1380. DOI=10.1109/34.895972 http://dx.doi.org/10.1109/34.895972
56. Westermann, U., Jain, R., , "Toward a Common Event Model for Multimedia Applications",*Multimedia, IEEE* , vol.14, no.1, pp.19-29, Jan.-March 2007.
57. R. Troncy, B. Malocha, and A. T. S. Fialho. 2010. Linking events with media. *In Proceedings of the 6th International Conference on Semantic Systems (I-SEMANTICS '10),* Adrian Paschke, Nicola Henze, and Tassilo Pellegrini (Eds.). ACM, New York, NY, USA, , Article 42 , 4 pages. DOI=10.1145/1839707.1839759 http://doi.acm.org/10.1145/1839707.1839759
58. X. Liu, R. Troncy, and B. Huet. 2011. "Using social media to identify events." In Proceedings of the 3rd ACM SIGMM international workshop on Social media (WSM '11). ACM, New York, NY, USA, 3-8. DOI=10.1145/2072609.2072613 http://doi.acm.org/10.1145/2072609.2072613
59. Hila Becker, Mor Naaman, and Luis Gravano. 2010. Learning similarity metrics for event identification in social media. In Proceedings of the third ACM international conference on Web search and data mining (WSDM '10). ACM, New York, NY, USA, 291-300. DOI=10.1145/1718487.1718524 http://doi.acm.org/10.1145/1718487.1718524
60. Papadopoulos, S.; Zigkolis, C.; Kompatsiaris, Y.; Vakali, A.; , "Cluster-Based Landmark and Event Detection for Tagged Photo Collections," Multimedia, IEEE , vol.18, no.1, pp.52-63, Jan. 2011
61. Till Quack, Bastian Leibe, and Luc Van Gool. 2008. World-scale mining of objects and events from community photo collections. In Proceedings of the 2008 international conference on Content-based image and video retrieval (CIVR '08). ACM, New York, NY, USA, 47-56. DOI=10.1145/1386352.1386363 http://doi.acm.org/10.1145/1386352.1386363
62. Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In Proceedings of the 19th international conference on World wide web (WWW '10). ACM, New York, NY, USA, 851-860. DOI=10.1145/1772690.1772777 http://doi.acm.org/10.1145/1772690.1772777
63. Ana-Maria Popescu and Marco Pennacchiotti. 2010. Detecting controversial events from twitter. In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM '10). ACM, New York, NY, USA, 1873-1876. DOI=10.1145/1871437.1871751 http://doi.acm.org/10.1145/1871437.1871751
64. Sayyadi H, Hurst M, Maykov A (2009), Event detection and tracking in social streams. In: Proceedings of international AAAI conference on weblogs and social media. AAAI press
65. Yonghong Tian; Srivastava, J.; Tiejun Huang; Contractor, N.; , *Social Multimedia Computing,* Computer , vol.43, no.8, pp.27-36, Aug. 2010

66. M. Naaman, Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications *Multimedia Tools and Applications*, Springer, Computer Science,vol. 56, no.1, pp. 9-34
67. Sergej Sizov. 2010. GeoFolk: latent spatial semantics in web 2.0 social media. In Proceedings of the third ACM international conference on Web search and data mining (WSDM '10). ACM, New York, NY, USA, 281-290. DOI=10.1145/1718487.1718522 http://doi.acm.org/10.1145/1718487.1718522
68. Bosch, Anna; Zisserman, Andrew; Muoz, Xavier. Scene Classification Via pLSA.Computer Vision ECCV 2006. Lecture Notes in Computer Science. 2006 Springer. pp.517-530, vol. 3954.
69. Qiang Yang, Yuqiang Chen, Gui-Rong Xue, Wenyuan Dai, and Yong Yu. 2009. Heterogeneous transfer learning for image clustering via the social web. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1 (ACL '09), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, 1-9.
70. Sivic, J. Russell, B. C.; Efros, A. A.; Zisserman, A. and Freeman, W. T. (2005), Discovering Object Categories in Image Collections, in "Proceedings of the International Conference on Computer Vision"
71. Hofmann, T.: Probabilistic latent semantic indexing. ACM SIGIR, (1998)
72. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. Machine Learning 41 (2001) 177196
73. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (March 2003), 993-1022.
74. Yu-Ru Lin; Sundaram, H.; De Choudhury, M.; Kelliher, A.; , "Temporal patterns in social media streams: Theme discovery and evolution using joint analysis of content and context," Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on , vol., no., pp.1456-1459, June 28 2009-July 3 2009
75. de Silva, G.C.; Aizawa, K.; Arase, Y.; Xing Xie; , "Interactive social, spatial and temporal querying for multimedia retrieval," Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on , vol., no., pp.7-12, 13-15 June 2011
76. Adrian Popescu, Gregory Grefenstette, and Pierre-Alain Mollic. 2009. Mining tourist information from user-supplied collections. In Proceedings of the 18th ACM conference on Information and knowledge management (CIKM '09). ACM, New York, NY, USA, 1713-1716.
77. Gennady Andrienko, Natalia Andrienko, Peter Bak, Slava Kisilevich, and Daniel Keim. 2009. Analysis of community-contributed space-and time-referenced data (example of Panoramio photos). In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '09). ACM, New York, NY, USA, 540-541.
78. Dalal, N.; Triggs, B.; , "Histograms of oriented gradients for human detection," Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on , vol.1, no., pp.886-893 vol. 1, 25-25 June 2005
79. Shechtman, E.; Irani, M.; , "Matching Local Self-Similarities across Images and Videos," Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on , vol., no., pp.1-8, 17-22 June 2007
80. Savvas A. Chatzichristofis and Yiannis S. Boutalis. 2008. CEDD: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In Proceedings of the 6th international conference on Computer vision systems (ICVS'08), Antonios Gasteratos, Markus Vincze, and John K. Tsotsos (Eds.). Springer-Verlag, Berlin, Heidelberg, 312-322.
81. Xin-Jing Wang; Lei Zhang; Xirong Li; Wei-Ying Ma; Annotating Images by Mining Image Search Results. *TPAMI*.2008, pp.1919-1932.
82. Torralba, A.; Fergus, R.; Freeman, W.T.; 80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition. *TPAMI*. 2008. pp. 1958-1970.