Information Fusion 67 (2021) 94–115

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Image retrieval from remote sensing big data: A survey

Yansheng Li [a], Jiayi Ma [b,*], Yongjun Zhang [a,*]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[b] Electronic Information School, Wuhan University, Wuhan 430072, China

ARTICLE INFO

Keywords:
Remote sensing (rs) big data
Rs image retrieval methods
Rs image retrieval applications
Evaluation datasets and performance discussion
Future research directions

ABSTRACT

The blooming proliferation of aeronautics and astronautics platforms, together with the ever-increasing remote sensing imaging sensors on these platforms, has led to the formation of rapidly-growing earth observation data with the characteristics of large volume, large variety, large velocity, large veracity and large value, which raises awareness about the importance of large-scale image processing, fusion and mining. Unconsciously, we have entered an era of big earth data, also called remote sensing (RS) big data. Although RS big data provides great opportunities for a broad range of applications such as disaster rescue, global security, and so forth, it inevitably poses many additional processing challenges. As one of the most fundamental and important tasks in RS big data mining, image retrieval (i.e., image information mining) from RS big data has attracted continuous research interests in the last several decades. This paper mainly works for systematically reviewing the emerging achievements for image retrieval from RS big data. And then this paper further discusses the RS image retrieval based applications including fusion-oriented RS image processing, geo-localization and disaster rescue. To facilitate the quantitative evaluation of the RS image retrieval technique, this paper gives a list of publicly open datasets and evaluation metrics, and briefly recalls the mainstream methods on two representative benchmarks of RS image retrieval. Considering the latest advances from multiple domains including computer vision, machine learning and knowledge engineering, this paper points out some promising research directions towards RS big data mining. From this survey, engineers from industry may find skills to improve their RS image retrieval systems and researchers from academia may find ideas to conduct some innovative work.

1. Introduction

The collaborative progress of multiple disciplines and fields, including but not limited to the sensor networks, communication technologies and storage technologies, enables the advent of the era of big data [1–4]. Generally speaking, big data has five remarkable characteristics: large volume, large variety, large velocity, large veracity and large value. In this era, the real benefit is not related to the data itself, but associated with software and hardware technologies that are capable of extracting knowledge from heterogeneous big data in a tolerable elapse time [2]. Driven by this urgent demand, more and more researchers from both industry and academia work on big data processing, fusion and mining where fusion and mining are two highly correlated tasks (e.g., good fusion results often benefit mining and many mining skills can also be adopted in building fusion algorithms). As a consequence, there is an increasing rate in the number of big data mining studies. However, big data mining is still a worldwide problem, and deserves much more inter-disciplinary research. According to the difference of data domains, big data can be coarsely divided into: social big data [5], urban big data [6], biology big data [7], climate big data [8,9], geospatial big data [10], and remote sensing (RS) big data [11–13]. Although they have some common characteristics such as the scalable and heterogeneous nature, different kinds of big data also have their special data characteristics spawning many domain-specific big data mining technologies. In particular, RS big data, also called big earth data [14–16], are mainly composed of massive RS images recording the earth surface, and play an important role in many applications such as fusion-oriented RS image processing, geo-localization and navigation, disaster rescue, and so on.

In the early years, RS data is one kind of scarce and expensive strategy resource. Along with the launch of more and more earth observation satellites, especially the openly accessible satellites such as the Landsat series [17] and Sentinel series [18], the access to global RS imagery via online partials such as the Copernicus Sentinel Hub [18], Google Earth Engine (GEE) [19], and EarthServer [20] become convenient and cheap. According to the acquisition report, only the Sentinel

* Corresponding authors.
E-mail addresses: yansheng.li@whu.edu.cn (Y. Li), jyma2010@gmail.com (J. Ma), zhangyj@whu.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.inffus.2020.10.008
Received 19 June 2020; Received in revised form 7 October 2020; Accepted 11 October 2020
Available online 14 October 2020

series can produce an estimated data volume of around 20 TB per day. From this point, we can clearly see that the volume of RS data increases with a high velocity [21]. As summarized in [15,16], in addition to the basic characteristics of big data, RS big data has some particular data characteristics: 1) Non-repeatability. Observations of physical objects and processes are unique in space and time and generally cannot be repeated; 2) Uncertainty. Big data involves different approaches to observation and recording, as well as indirect observation and sampling; 3) Multi-dimensionality. A wide range of data sources and complex analysis methods lead to a wide range of dimensionality. The aforementioned characteristics of RS big data result in a high degree of computational complexity in the RS data analysis. To cope with these computational challenges, kinds of advanced hardware infrastructures [22–26] have been proposed to accelerate the computational process. The study of acceleration infrastructure plays an important role on RS big data mining, but is beyond the scope of this paper. In addition to the acceleration infrastructures, exploiting efficiently scalable algorithms is one head-on direction to cope with the high computational complexity problem. Hence, this paper mainly reviews the intelligent techniques to address RS big data mining from the algorithm perspective.

Image retrieval from RS big data aims at retrieving the interested RS images from the massive RS image repositories with the volume of the Peta Bytes/Zetta Bytes (PB/ZB) scale. Generally, RS image retrieval can prepare the auxiliary data or narrow the search space for lots of RS image processing tasks including RS image matching [27–29], RS image registration [30–32] and RS image fusion [33–37]. In addition, image retrieval from RS big data plays an important role on fusion-oriented image processing [38–40], geo-localization and navigation [41,42], disaster rescue [43,44], meteorological analysis [45], economic assessment [46,47], and ecology prediction [48], and so forth. Because of its wide applications, RS image retrieval attracts tremendous research interest. Fig. 1 illustrates the increasing number of publications that are associated with "RS image retrieval" over the past two decades. To pursue the sustainable development of the RS image retrieval technology, this paper gives a survey around image retrieval from RS big data. More specifically, this review discusses the existing RS image retrieval methods grouped in four categories: 1) Conventional content-based RS image retrieval; 2) Hashing-based RS image retrieval; 3) Cross-modal RS image retrieval; 4) Interactive RS image retrieval and presents several classic applications driven by RS image retrieval. Moreover, to facilitate conducting the quantitative evaluation, the paper summarizes the existing evaluation resources including datasets and metrics. While lots of efforts have been made in RS image retrieval, there are still some remaining problems in existing RS image retrieval methods, i.e., the limited scale of RS image retrieval datasets in terms of the volume of samples, the number of categories and the number of modalities, the high dependency on large-scale supervision data with accurate labels, the lack of human-like reasoning abilities. To address these challenges, this survey points out some promising research directions: 1) Developing larger RS image retrieval datasets; 2) Weakly supervised deep learning for RS image retrieval; 3) Visual reasoning for RS image retrieval. To sum up, this paper aims to give a specific and comprehensive review of methods for image retrieval from RS big data. In addition, it tries to shed light on how to further improve the performance of existing methods and points out some advanced research directions to further lift the mining ability of RS big data.

Due to its wide applications, image retrieval has been exploited for several decades. From the 1990s, the pioneers in the computer vision domain [49,50] have launched a series of special issues to guide the research of content-based image retrieval (CBIR). As a consequence, theories and methods towards CBIR have achieved great development. The details about such development can refer to some specific surveys [51–55]. In contrast to natural images, RS images have a large variation in terms of modality, spectral and resolution. To guide the study of content-based RS image retrieval, many RS-oriented sections [56,57] have been organized. To summarize such progresses, many reviews for content-based RS image retrieval [58–67] have been released. As a whole, these existing reviews mainly discuss the achievements from one or several perspectives, e.g., the review of the interest point descriptors [58], the discussion of similarity measures [59] and the comparable analysis of deep features [64]. In addition, these existing reviews have a very limited analysis of the newly emerging scalable retrieval and cross-modal retrieval techniques, which are highly needed in the era of RS big data. With the aforementioned considerations, this paper aims to give a comprehensive and enlightening review about image retrieval in the context of RS big data. In comparison with previous reviews, this survey has several certain advantages. Firstly, this survey points out the opportunities and challenges in image retrieval from RS big data,

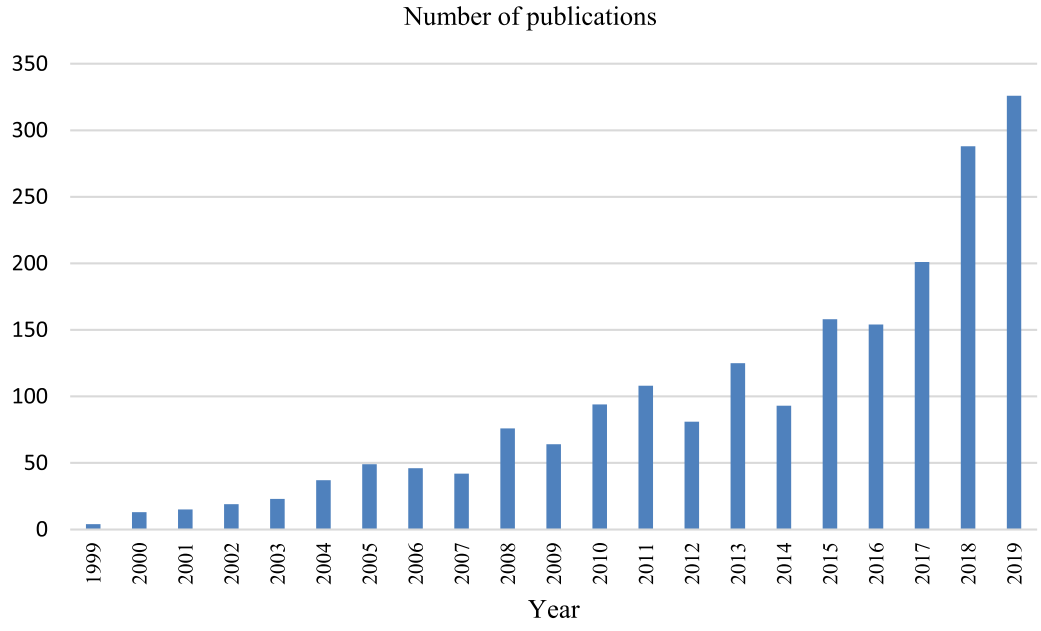

Fig. 1. The increasing number of publications in RS image retrieval from 1999 to 2019. Data is collected by the advanced search of Google Scholar (allintitle: "RS image retrieval" OR "RS image indexing" OR "RS image mining" OR "satellite image retrieval" OR "satellite image indexing" OR "satellite image mining" OR "earth observation image retrieval" OR "earth observation image indexing" OR "earth observation image mining").

guiding the related researches. Secondly, this survey reports nearly all main RS image retrieval methods, points out their RS-domain-specific applications, and discusses their adaptation possibilities and strategies in RS big data mining. Thirdly, this work gives some promising research topics towards intelligently mining RS big data.

This overall structure of this paper is illustrated in Fig. 2. In the following, Section 2 summarizes the primary achievements of RS big data mining and points out the remaining challenges. Section 3 focuses on reviewing the existing methods for image retrieval from RS big data. Section 4 emphasizes several specific applications of RS image retrieval. The evaluation resources and performance discussion are specifically depicted in Section 5. In addition, Section 6 gives some promising research directions of RS big data mining. Finally, Section 7 gives a conclusion of this survey.

2. Image retrieval from remote sensing big data: opportunities and challenges

In this section, we mainly discuss the opportunities and challenges around image retrieval from RS big data, respectively.

2.1. Opportunities in remote sensing big data mining

Generally, the response time is a critical criterion in image retrieval from RS big data. For example, if an emergency such as an earthquake occurred, the RS imagery data of interest should be accurately retrieved from massive distributed databases in a very short time as few seconds can save many lives by timely warnings. Hence, image retrieval from RS big data is a dual data and compute intensive task. As illustrated in Fig. 3 (a), we are witnessing the coming technological leapfrogging in terms of distributed storage database, high performance computing (HPC) and artificial intelligence (AI).

After the RS imagery is transmitted to the ground station, these RS imagery is often stored in a special database system [68,69]. With the rapid growth of RS data, the traditional structured related database systems cannot meet the requirements of managing RS big data. Accordingly, more and more researchers plunge themselves into developing the novel data storage system which can flexibly manage RS big data in the PB/ZB scale [70]. In recent years, the distributed storage system [71] and the resilient distributed storage system [72] have been successively exploited to manage RS big data.

Modern advances in computing hardware have been enabling new opportunities for the manner in which large volumes of distributed imagery data are processed. More specifically, Hadoop [73] has emerged as an early testbed for big data applications due to its excellent large-scale data-handling capability, high fault tolerance, reliability and low cost of operation. In addition, Hadoop has also been successively used to address the large-scale RS image processing tasks such as image segmentation [74]. Based on the basic architecture of Hadoop, MapReduce [75] is exploited to address large-scale image retrieval on massive image databases [76]. To alleviate the heavy usage of disk input/output (I/O) operations in MapReduce when used in conjunction with Hadoop and the extra storage of intermediate results, Apache Spark has been exploited to take advantage of the resilient distributed dataset and shows its effectiveness in the RS image mosaicking task with the frequent data I/O requirement [77]. In the scenario of Apache Spark, partitioning massive amount of data based on the spectral and semantic characteristics for distributed imagery analysis, also benefits improving the computing efficiency [25]. It is worth noting that the emergence of specialized hardware devices such as the graphic processing units (GPUs) [78] dramatically lift the computation efficiency of the advanced artificial intelligence methods. Benefiting from HPC, content-based RS image retrieval [11] is successfully performed to extract variation information from a large-scale image dataset, collected after the terrorist attack to the World Trade Center in New York City.

As one of the most outstanding achievements in the AI domain, deep learning [79–81] has achieved tremendous success in various fields including image classification, speech recognition, natural language processing and so forth. By automatically digesting massive labeled data, deep learning could obtain an excellent hierarchical abstract ability, which benefits narrowing the semantic gap between the low-level raw data and the high-level concept. In the RS scenario, deep

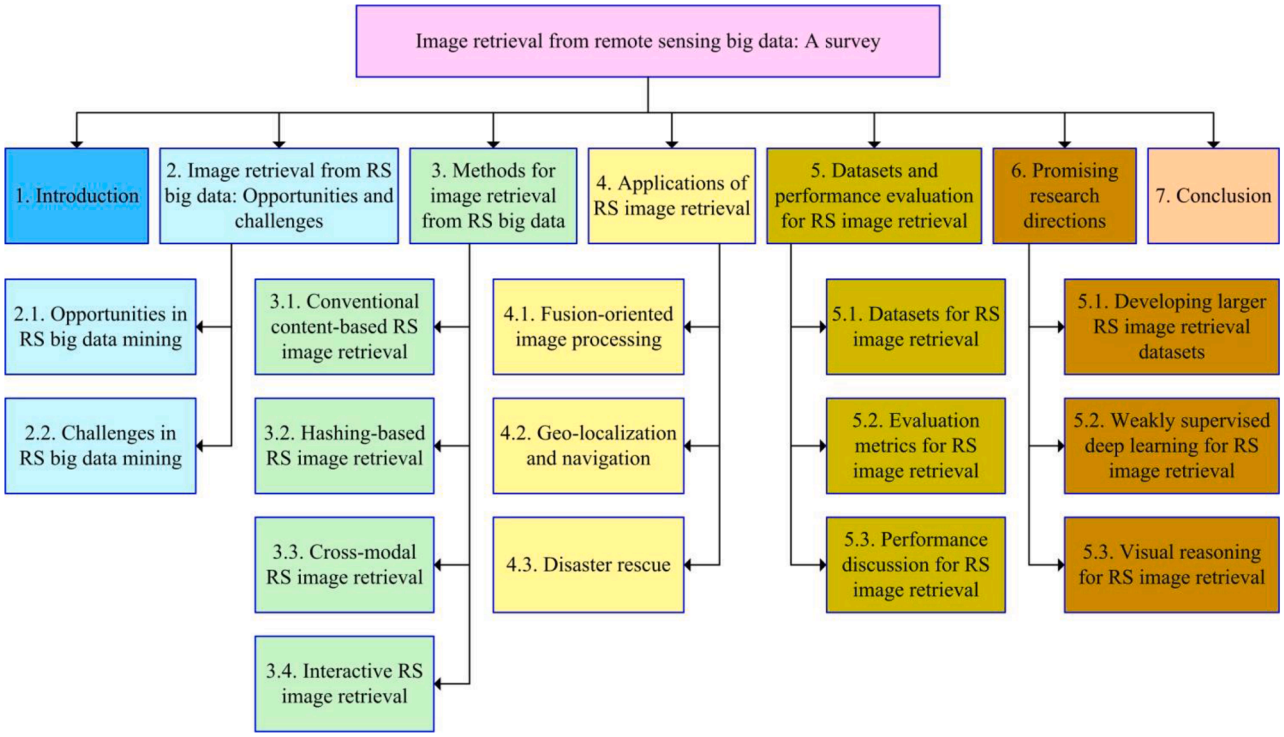

Fig. 2. Structure of this survey.

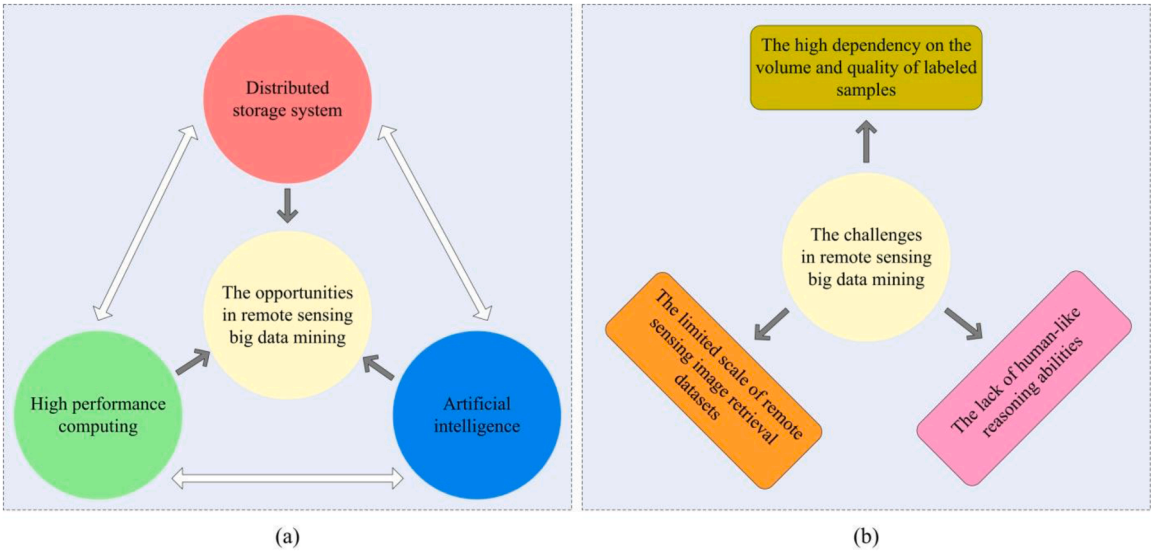

Fig. 3. Visual illustrations about opportunities and challenges in RS big data mining. (a) shows the main opportunities in RS big data mining; (b) lists the main challenges in RS big data mining.

learning has been successfully utilized in geospatial object detection [82–85], RS image fusion [86–90], RS image scene classification [91–95], and so on. It is worth mentioning that deep learning also contributes a lot to improving the RS image retrieval performance in terms of high-compactness visual feature indexing and high-quality visual content representation, which will be specifically discussed in Section 3.

2.2. Challenges in remote sensing big data mining

As reflected in [11], it is still a vital challenge for government institutions to share data unless all participants can achieve material benefits and incentives in data sharing that outweigh the risks. For example, although NASA has been sharing massive RS data under the open government policy, the overwhelming majority of high-quality RS data (i.e., high-resolution RS images) are still unavailable to the public. Therefore, it is necessary to find new ways of collaboration and establish a flexible RS data sharing policy for improved big data access in RS problems and applications.

In addition to the aforementioned RS data access restriction challenge, the state-of-the-art RS image retrieval methods present three main limitations, which are visually shown in Fig. 3(b), from the algorithm perspective. Firstly, the collection cost of RS imagery becomes lower and lower. However, how to annotate large-scale RS data for learning RS image retrieval methods becomes the new challenge. Although kinds of RS image retrieval datasets have been publicly released, the scale of available RS image retrieval datasets in terms of the volume of samples, the number of categories and the number of data modalities is still very limited. Hence, how to collect a large-scale labeled RS image retrieval dataset with fine-grained categories containing multiple classical satellite data types becomes more and more urgent and shows significant meanings in the deep learning era. Secondly, the state-of-the-art RS image retrieval methods often take deep networks as the backbone. As well known, the superior performance of deep networks is conditioned on a large-scale labeled dataset. As a natural transmission, the state-of-the-art performance of RS image retrieval methods also depend on an oversized labeled dataset with accurate labels. In addition, deep networks under weak supervision (e.g., the volume of labeled samples is relatively small or the labels of one oversized dataset are noisy) would significantly degenerate. In reality, the collection of datasets with weak labels in the RS context is relatively simple. Parallel to the research direction towards annotating RS datasets, how to robustly train deep networks under weak supervision would be another promising research direction. Thirdly, although the current content-based image retrieval (CBIR) techniques including hand-crafted features and deep learning-based features are always trying to intrinsically depict the visual content of RS imagery, all of them are still lack of the human-like reasoning abilities that help to accurately perceive the objects in the RS imagery and their spatial topological relationship. As a whole, visual reasoning plays an important role on the advanced RS image retrieval technique and helps to effectively bridge vision and language. To address these limitations in image retrieval from RS big data, we give the promising solutions in Section 6 by examining the most recent progress in the AI domain.

3. Methods for image retrieval from remote sensing big data

Image retrieval from RS big data refers to finding RS images that satisfy an information need from large RS image collections. As depicted in [96], there is an intrinsic difference between CBIR and retrieval by text and metadata. Retrieval methods based on metadata [97,98] rarely examine the visual content of an image itself but rather rely on manually generated tags. In these systems, keywords should be manually generated in advance and are stored as strings to describe the RS images. However, the high complexity of RS images cannot be described easily by keywords; thus, retrieval systems which are based solely on manual annotation often lead to unsatisfactory outcomes. In the contrast, CBIR does not depend on keywords and the desired images can be retrieved automatically based on their visual content similarities to the query image. Because of this reason, the overwhelming majority of existing RS image retrieval methods adopts the CBIR manner.

Briefly, one CBIR system includes three core modules including feature representation, feature indexing and feature similarity measuring. Specifically, feature representation refers to extracting the feature vector from the image to represent its visual content. In addition, feature indexing works for structuring a database (i.e., structuring the feature vectors extracted from images) to lift the search speed. Since response time is one of the key indicators in CBIR systems, the importance of feature indexing becomes incredibly remarkable, especially in a large-scale image database. An efficient database indexing technique can significantly accelerate the retrieval process and reduces memory usage substantially [99]. Conventional methods use a similarity metric to compare the feature vector of the query image to each feature vector in the database. However, whilst comparing the query feature vector to

the entire image dataset might be feasible for a small dataset, this is still an O(N) linear operation where N is the number of images in the dataset. Thus, for large-scale datasets with billions of images, the exhaustive feature search becomes impractical. Furthermore, feature similarity measuring refers to calculating the visual similarities between the query image and images in the dataset by designing appropriate feature distance metrics. As a whole, the feature indexing technique is relatively mature, but feature representation and feature similarity measuring are techniques under developing and probing.

Different from the CBIR technique in the computer vision domain, RS image retrieval needs to further cope with more complex data variation because RS imagery includes so many kinds of data types compared with the single type of natural imagery often with the fixed R-G-B spectral bands. According to the type of RS images, RS image retrieval methods can be coarsely categorized into four main categories including panchromatic/multi-spectral image retrieval methods [100–102], synthetic aperture radar (SAR) image retrieval methods [103–108], hyper-spectral image retrieval methods [109–117], and time-series RS image retrieval methods [118–122]. In the following, we mainly review the common techniques in the existing RS image retrieval methods from four aspects including conventional content-based RS image retrieval methods, hashing-based RS image retrieval methods, cross-modal RS image retrieval methods, and interactive RS image retrieval methods.

3.1. Conventional content-based remote sensing image retrieval

As depicted in Fig. 4, the conventional content-based RS image retrieval framework generally consists of three modules including feature representation, feature indexing, and feature similarity measuring. In the following, we specifically review the progresses around these main modules.

3.1.1. Feature representation

As well known, the CBIR system chiefly cares the visual content of RS imagery, which is often expressed by one or several kinds of feature representations (i.e., feature vectors) extracted from RS imagery based on hand-crafted descriptors or data-driven deep networks. According to the abstract level, the feature representations, which are adopted in RS image retrieval task, can be coarsely divided into three main categories including the low-level feature representations, the middle-level feature representations, and the high-level feature representations.

3.1.1.1. Low-level feature representations. As the primary description of RS imagery, low-level feature representation is designed by domain experts and is often built by mining the spectral, texture, or shape cues of RS imagery.

As well known, RS imagery often has more spectral bands compared with natural images. Spectral is the fundamental unit of RS imagery. Although the spectral feature is one of the simplest feature representations, it encodes the reflectance of the corresponding areas of the Earth surface and depicts the most prominent information of RS imagery [60]. In literature, the spectral feature has been adopted in many RS image retrieval methods [123–126]. However, the spectral feature based RS image retrieval methods often present serious sensitivity to noise and illumination change [64].

As one of the widely adopted hand-crafted features in the computer vision domain, texture is generally understood as repeated structures in the image. To depict the texture information, descriptors, such as gray level co-occurrence matrices (GLCM) [127], wavelet [128], Gabor filters [129], and local binary patterns (LBP) [130,131], have been exploited. Afterwards, based on the characteristic of RS images, kinds of texture descriptors have been modified to address RS image retrieval [132–141]. Recently, Sukhia et al. [142] propose a local ternary pattern (LTP) to depict the visual content of RS images where LTP aims to obtain

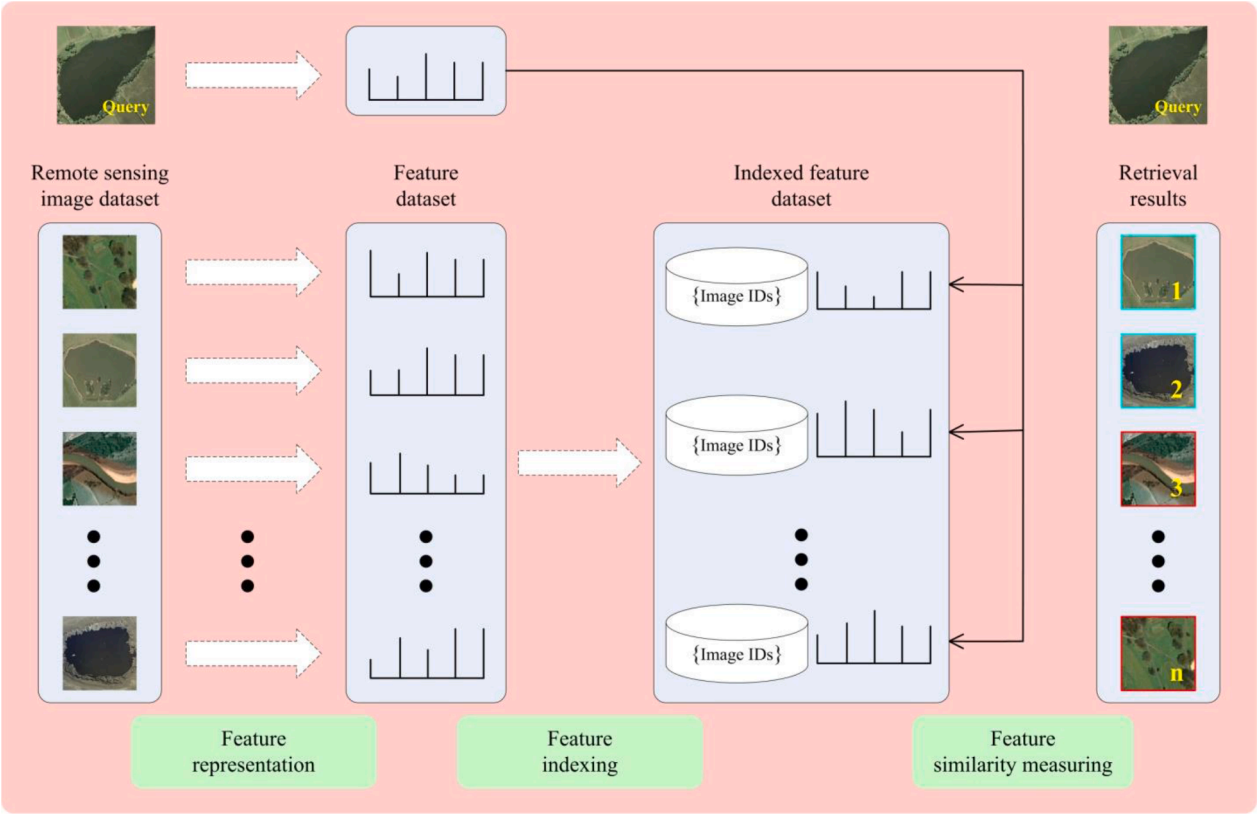

Fig. 4. The workflow of the conventional content-based RS image retrieval technique. In the retrieval results, the samples with the blue rectangles stand for the true positives, and the samples with the red rectangles denote the false positives.

upper and lower texture patches from each down-sampled image and divides them into dense patches to build a final histogram representation. It is worth noting that the wavelet based feature [143] has been proven to be effective in directly representing the visual content of RS imagery in the compressed domain.

Shape is an important recognition cue of geospatial objects on the earth surface in the RS images [45,144–146]. In literature, shape features have been adopted to address infrared image retrieval [144] and optical object retrieval [146]. Shape features generally depict the boundary or outline information of geospatial objects, but have a very limited ability to capture the spatial relationship information. As a special type of shape features, local feature points have also been adopted to address RS image retrieval. For example, salient feature points [147], and scale-invariant feature transform (SIFT) [148] have been proved to be more effective than texture features in RS image retrieval. Structural features derived from shape ensembles and relationships also provide satisfactory performance [149]. In many applications, single type of low-level features lacks enough discrimination. Therefore, researchers combine diverse types of features to improve the retrieval results [150–157]. Generally, different shape features help to make up for each other's defects. As a consequence, the combination of multiple hand-crafted features often presents stronger representation ability and benefit improving RS image retrieval.

3.1.1.2. Middle-level feature representations. In contrast with low-level features, middle-level features embed low-level hand-crafted feature descriptors into representative visual vocabulary space and encode spatial distribution to capture semantic concepts of RS images. Generally, middle-level features are more invariant to appearance difference caused by changes of scale, rotation or illumination. Hence, middle-level features benefit better representing the complex image textures and structures with more compact feature vectors. The general pipeline to extract middle-level features is firstly obtaining the hand-crafted descriptors (e.g., spectral, texture or local invariant features), and then aggregating them into holistic representations using encoding methods (e.g., bag-of-words (BoW) [158], and vector locally aggregated descriptors (VLAD) [159]) and unsupervised learning methods (e.g., auto-encoder [160], and artificial neural network [161]).

Among the encoding methods, BoW is one widely used basic encoding method, which often employs the k-means clustering algorithm to construct visual codebook and calculates the histogram of local feature descriptors based on the visual codebook. It has been utilized in some RS image retrieval research and has achieved desired results. Specifically, [162–164] have shown the effectiveness of encoded features compared with low-level features. It is worth noting that BoW is also helpful in encoding the features of the pre-trained deep networks [165]. In addition, VLAD is an advanced version of BoW, apart from feature distribution, it additionally counts the distance between local features and cluster centers. VLAD is applied to encode local pattern spectra [166] and obtains high-precision retrieval results on high-resolution RS images [167]. In [63], the experimental results demonstrate that BoW behaves better in calculation speed while VLAD behaves better in retrieval accuracy. Afterwards, multi-scale spatial information has also been exploited in feature encoding. For instance, spatial pyramid matching based on sparse codes (ScSPM) [168] can fuse holistic and local features to enhance the discrimination of middle-level features [169]. Recently, VLAD has been utilized to aggregate local deep features to produce a global descriptor for advanced performance of RS image retrieval [170].

To lift the discrimination ability of low-level features, Zhou et al. [160] propose an auto-encoder model to learn sparse feature representation from multiple low-level features for RS image retrieval. To pursue the superior RS image retrieval performance, an ensemble artificial neural network [161] has been proposed to improve the low-level features. By taking the topological structure into consideration, Du et al. [171] propose a local structure learning method, which can map the local features into a manifold space by a Lipschitz smooth function to enhance the representation ability of the features. In addition, the structural feature in the manifold space is further utilized to conduct RS image retrieval.

3.1.1.3. High-level feature representations. The hierarchical architecture of convolutional neural network (CNN) can simulate very complex nonlinear functions and automatically learn hyper-parameters of CNN during the training process. As CNN models are able to capture the essential characteristics of images, the intermediate outputs of CNN models can be taken as high-level features to comprehensively represent the visual content of RS images. Generally speaking, the sufficient update of hyper-parameters of one regular CNN model often depends on millions of accurately labeled samples. Compared with the large-scale labeled natural image dataset (e.g., ImageNet), the volume of labeled samples in the RS domain is relatively small, which goes against the learning of CNN models. Even so, a small amount of RS image retrieval works based on high-level features have been presented up till now. In [172–173], unsupervised deep learning has been proposed to learn deep networks via an unsupervised manner, and also achieved the promising performance on RS image retrieval [174]. Considering that aerial images are relatively similar to natural images, the pre-trained CNN models using ImageNet is transferred to address aerial image representation. More specifically, high-level features of aerial images are represented by existing CNNs from convolutional layers or fully connected layers [175]. In addition, fine-tuning pre-trained CNNs with the target-domain RS image datasets [64] benefits outputting more effective high-level features. In recent years, many specific objective functions have been proposed to train CNNs for RS image retrieval. For example, the center loss function with inter-class dispersion and intra-class compaction is proposed to learn discriminative deep features for RS image retrieval [176]. To capture the spatial detail, the graph convotional network (GCN) with the pairwise similarity constraint [177] is proposed to address RS image retrieval. In addition, the triplet loss [178] has also been adopted in learning discriminative deep features for RS image retrieval. To deal with the limitation of triplet loss, Liu et al. [179] propose a global optimal structured loss, which globally learns an efficient deep embedding space with mined informative sample pairs to force the positive pairs within a limitation and push the negative ones far away from a given boundary. Due to the uneven distribution of sample data in RS image datasets, the pair-based loss currently used in deep metric learning (DML) is not optimal. To improve this problem, Fan et al. [180] propose a distribution consistency loss to make deep networks learn more useful information in a short time. By examining the issues that the existing CNN-based RS image retrieval methods do not deal with large intra-class variations and all similarity learning based RS image retrieval methods consider similarity between two images as a constant, Liu et al. [181] propose a metric learning method with a positive-negative center loss to enable CNNs to cope successfully with within-class variations. As a whole, metric learning has played an important role on learning discriminative deep features for RS image retrieval. The current deep networks still can't accurately encode the object content of RS images with complex and cluttered backgrounds. Many recent works in the computer vision domain [182,183] show that the attention mechanism could effectively guide deep networks to distinguish the important object regions with a high attention bias and ignore the cluttered background regions. Hence, boosting deep networks with the attention mechanism would be a promising solution to cope with the case that objects get buried in the complex and cluttered backgrounds.

In addition to the high-level feature representations based on CNN, human-centered concepts or application-specific knowledge also help to generate the high-level feature representations of RS images. Specifically, the pre-defined concept set [184,185] and category set [186,187]

are taken as semantic bases to encode the semantic features of RS images. Afterwards, the spatial context information [188] is further proposed to improve the concept-driven semantic features. As the formal and explicit specification of a shared conceptualization, ontology shows great potential in knowledge modeling and helps to depict the visual content of RS images in the high abstract level [189,190].

Although high-level features show overwhelming superiority in RS image retrieval compared with low-level and mid-level features, the combination of low-level, middle-level and high-level features still outperforms single type of feature, which reveals the complementary abilities of features from different levels [174]. How to effectively combine hand-crafted and data-driven features would be a promising way to improve RS image retrieval performance and deserves much more exploitation.

3.1.2. Feature indexing

In practical applications, feature indexing [99] is often adopted when searching the oversized RS image archive, but seems to be not so necessary when the RS image dataset is with a relatively small volume. Generally, feature indexing is often coupled with the distributed file system and the MapReduce operation in big data mining [23]. As aforementioned, given one query feature vector, the exhaustive feature searching of N feature vectors (i.e., RS images) in the database would involve of $O(N)$ computations. To accelerate the search process, three kinds of greedy feature indexing methods including tree-based indexing, clustering-based indexing and hashing-based indexing have been adopted in large-scale RS image retrieval.

3.1.2.4. Tree-based feature indexing. The tree-based feature indexing algorithms aim at recursively splitting the feature space into subspaces and forming the subspaces by a tree structure [191]. For instance, a k-dimensional (KD) tree has been introduced in [192], and it is extended to an entropy balanced statistical KD tree in [193]. Three tree-based indexing structures, namely, rectangle-tree, sphere-sphere-tree, and sphere-rectangle-tree, have been compared in [194] to find a suitable and efficient structure for satellite image archives. In the image retrieval stage of tree-based algorithms, the branch-and-bound technique is usually considered to search and retrieve approximate nearest neighbors [195]. Although the searching speed is significantly improved with the partition trees (i.e., the search complexity is $O(\log(N))$), their performance considerably decreases when the dimension of the image features increases [196].

As a whole, the tree-based indexing methods also suffer of memory constraints since tree structures are typically bigger than the original data. Consequently, the use of tree-based indexing strategies is not appropriate for CBIR problems where the RS image descriptors are often high dimensional. Hence, the advanced feature dimension reduction techniques may make the tree-based indexing strategy workable even when the RS image descriptors are high dimensional.

3.1.2.5. Clustering-based feature indexing. Clustering-based feature indexing refers to aggregating feature vectors to clusters (i.e., visual words) [197], where all visual words constitute a visual codebook. Similar to the inverted file skill in text information retrieval, each visual word is followed by a list of image IDs in which the visual word occurs. When searching online, the query feature vector only needs to compare with the visual words instead of comparing all of the feature vectors in the database. Apparently, this clustering-based inverted file structure benefits improving the search efficiency. Afterwards, hierarchical clustering structure [198–200] is proposed to further improve the search efficiency.

The main drawback of such approaches is that the clusters are fixed and not evolving; therefore, adding even a single new image to the database requires the whole procedure, including the clustering to be repeated from scratch. Furthermore, in high dimension, data becomes very sparse and distance measures become increasingly meaningless caused the performance of clustering techniques to be degraded. Along with the great success of deep learning, the usage of deep features may help to alleviate the degradation phenomena.

3.1.2.6. Hashing-based feature indexing. Hashing-based feature indexing and approximate nearest neighbor search techniques have attracted attention in the multimedia communities due to their high time-efficient search capability within huge data archives and high data storage capability [201]. Hashing methods initially embed high-dimensional image features into a low-dimensional Hamming space, where the image features are represented by binary hash codes [201]. The binary codes can significantly reduce the amount of memory required for storing the images' content. The hashing methods initially generate hash functions to be applied to each image in the archive to obtain the binary hash code of the considered image. Then, a hash table is generated, where similar images have the same hash code, being positioned in the same hash bucket. Accordingly, indexing of archive images is achieved. The hash code of the query image is estimated by the use of the same hash functions. Then, the image retrieval can be achieved by using different strategies. In the computer vision literature, there are two criteria commonly used. The first criterion is called "hash lookup" and exploits the hash table to retrieve all the images in the hash buckets that fall within a small Hamming radius of the query image. Searching for ANNs with this criterion is independent from the number of images in the archive and is achieved in a constant time (i.e., O(1)) [202]. The second criterion is called "Hamming ranking" and estimates the Hamming distance between the hash code of the query image and those of all the images in the archive. Then, the images that have the lowest Hamming distance, with respect to the query image, are retrieved. Searching for ANNs with this criterion requires a linear time (i.e., $O(N)$); however, it is very fast in practical applications due to matching only the binary codes [202]. The storage complexity of a hash table can be expressed as $O(NK)$, where K stands for the number of hashing bits (i.e., the length of the hashing feature vector).

Due to the effectiveness of hashing-based feature indexing methods for large-scale RS image retrieval, lots of hashing methods for RS image retrieval have been proposed in recent years. To give a specific discussion, we systematically review the hashing methods in Section 3.2.

3.1.3. Feature similarity measuring

Feature similarity measuring refers to calculating the distance between visual feature vectors, which is the basis of pattern recognition. For it is one of the core issues in CBIR, feature similarity measuring is of great research significance. Generally speaking, similarity metrics can be coarsely divided into three categories including common distance function based similarity metrics, hand-crafted similarity metrics and data-driven metric learning based similarity metrics.

3.1.3.7. Common distance function based similarity metrics. Given the same feature representations, different similarity metrics may lead to different ranking results. In [59], eight similarity metrics based on common distance functions have been investigated for RS image retrieval. According to the properties of feature vectors, the similarity metrics can be divided into two groups including general feature vector based similarity metrics and histogram feature vector based similarity metrics.

3.1.3.8. Hand-crafted similarity metrics. In addition to the common distance function based similarity metrics, similarity metrics can also be manually defined based on the specific retrieval task. For example, in [120], an informational similarity metric is introduced for compressed RS data mining. In [115], a specific similarity metric for hyperspectral imagery is proposed. In [116], dictionary-based similarity metrics are proposed to address hyperspectral image retrieval.

3.1.3.9. Data-driven metric learning based similarity metrics. However, manually constructing a similarity metric may be inefficiency and not robust to different data sources; data-driven metric learning can be an ideal alternative. In contrast to hand-crafted similarity metrics, data-driven metric learning is capable of automatically learning distance function for a specific retrieval task according to task requirement [203]. Unsupervised metric learning has been successfully applied to RS retrieval, for example, [204] models RS images with graphs and uses an unsupervised graph-theoretic method to measure the similarity between the query graph and the graphs of images in the archive. In recent years, kinds of similarity measures based loss functions [205–208] have been proposed to train deep networks for RS image retrieval in an end-to-end manner.

In the actual applications, massive RS images are often stored in the distributed file system. To completely address CBIR from RS big data, the whole retrieval from large-scale RS images can be divided into multiple sub-tasks, like the general MapReduce technique in the big data processing domain. In addition, each sub-task resembles the Map operation and the Reduce operation works for fusing the returned results from each sub-task. Here, the fusing operation often needs to re-rank the returned retrieval results based on the aforementioned similarity metrics.

3.2. Hashing-based remote sensing image retrieval

To address large-scale RS image retrieval, feature reduction [209–211] has been adopted as feature reduction benefits not only saving the storage space of feature representation, but also lifting the computational speed of feature comparing. To pursue a thorough reduction, feature hashing [212,213] aims at mapping the high-dimensional feature vector to the low-dimensional binary feature vector. As illustrated in Fig. 5, hashing-based RS image retrieval methods generally include several critical modules: feature extraction, feature hashing and image ranking via the hamming distance which is a distance measure for calculating the similarities between the query binary feature vector and the stored binary feature vectors. It is worth noting that, driven by the end-to-end mechanism of deep learning, feature extraction and feature hashing have been merged into one module in some recent hashing methods. As feature extraction has been introduced in the previous section, this section mainly discusses the feature hashing technique.

It is assumed that the training RS image dataset contains N RS images $\{\mathbf{I}_i\}_{i=1}^{N}$. Based on the feature extraction module, the visual content of these images can be separately represented by feature vectors $\{\mathbf{x}_i\}_{i=1}^{N}$ where $\mathbf{x}_i \in R^D$ and D denotes the dimension of feature vector. The goal of feature hashing is to learn a nonlinear mapping function $f : \mathbf{x} \rightarrow \mathbf{h} \in \{-1, 1\}^K$ which aims at encoding each high-dimensional feature vector $\mathbf{x}$ to the compact K-bit hash code $\mathbf{h} = f(\mathbf{x})$. Existing feature hashing methods can be roughly divided into two categories: unsupervised feature hashing and supervised feature hashing. In the following, Section 3.2.1 focuses on reviewing unsupervised feature hashing for RS image retrieval, and Section 3.2.2 details the supervised feature hashing achievements for RS image retrieval.

3.2.1. Unsupervised feature hashing

The unsupervised feature hashing methods design hash functions using only unlabeled data to generate binary hash codes. The most popular unsupervised feature hashing method is the locality-sensitive hashing (LSH) [214], which constructs the r-th hash bit of a high-dimensional feature vector $\mathbf{x}$ based on the r-th hash function f_r as follows. Let $f_r(\mathbf{x}) = 1$, if $\mathbf{v}_r^T \cdot \mathbf{x} \geq 0$ where $\mathbf{v}_r$ is a random projection vector generated from a multivariate Gaussian with zero mean and an identity covariance matrix of the same dimension as the input $\mathbf{x}$; otherwise, $f_r(\mathbf{x}) = 0$.

In LSH, each projection vector $\mathbf{v}_r$ is randomly initialized. Thus, LSH is not conditioned to any labeled data. Although LSH significantly speeds up the CBIR process, its practical efficiency is still very limited since it requires long hash codes to achieve a high retrieval performance. Afterwards, the LSH has been recently extended to kernel unsupervised LSH (KULSH) in [215], to describe hash functions in the kernel space for

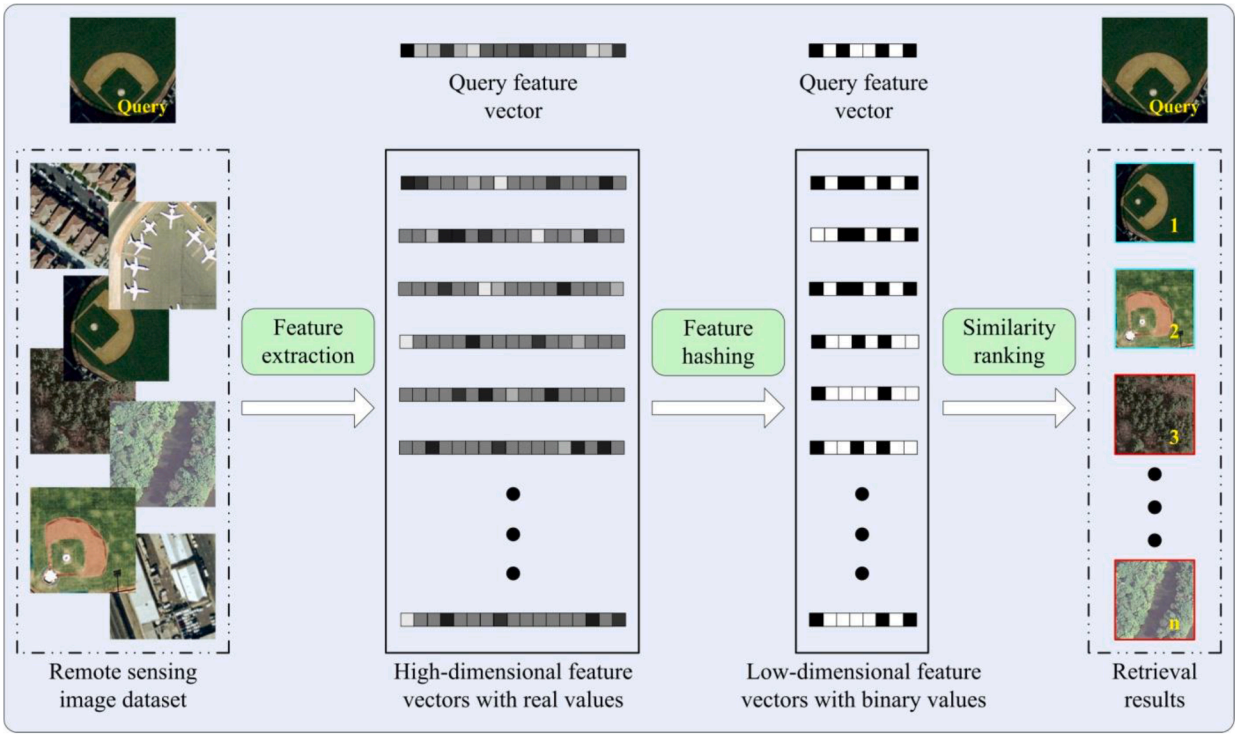

Fig. 5. The workflow of the hashing-based RS image retrieval technique. In the retrieval results, the samples with the blue rectangles stand for the true positives, and the samples with the red rectangles denote the false positives.

nonlinearly separable data. The random Fourier feature method for the shift-invariant kernel-based hashing is presented in [216]. In [217], a graph-based hashing technique is introduced to leverage the low-dimensional manifold structure of data to generate compact hash codes. In [218], binary reconstruction embedding is introduced that minimizes the reconstruction error between the original feature distance and the Hamming distance. In [219], a spherical hashing method that employs the hypersphere-based hashing functions is presented, while spectral hashing that is based on spectral graph partitioning is described in [220]. An inductive manifold hashing framework that provides a connection between manifold learning methods and hash function learning is studied in [221]. In [222], an unsupervised multi-view alignment hashing approach based on regularized kernel nonnegative matrix factorization is introduced. Neighborhood discriminant hashing that learns a discriminant hashing function by exploiting local discriminative information is presented in [223]. Using pseudo labels, an unsupervised deep hashing method [224] is proposed to address scalable image retrieval.

In the RS community, there are few unsupervised hashing strategies available. As an example, the KULSH [215] that defines hash functions for high-dimensional nonlinearly separable RS image descriptors was adopted for RS-based CBIR problems in [225]. The KULSH is defined based on the LSH, formulating the random projections in the kernel space by using a small set of images from the considered archive. Li et al. present in [226] the partial randomness hashing (PRH) method, which uses random projections to produce an initial estimation of the hash codes and then learns a linear model to re-project these codes onto the original feature space. Finally, the transpose of the projection matrix is used to generate the binary codes. Reato et al. [227] propose a multi-code hashing method that initially characterizes the images by using descriptors of primitive sensitive clusters, and then constructs the multi-hash codes from these descriptors using the KLSH. To accurately model the complex semantic content presented in RS images using binary codes, a new probabilistic latent semantic hashing (pLSH) model [228] is proposed to learn the hash codes in an unsupervised manner. To adapt to the continuously emerging RS images, an online batch-based hashing learning approach is introduced in [229].

Note that unsupervised hashing methods are, in general, very fast at generating hash functions. However, the hash functions obtained by using the unsupervised methods might be not discriminative enough for complex RS image retrieval problems.

3.2.2. Supervised feature hashing

In past several years, there are many successful supervised hashing methods that have been developed for fast image retrieval, including binary reconstruction embedding (BRE) [230], minimal loss hashing (MLH) [231], and sparse embedding and least variance encoding (SELVE) [232]. By utilizing the supervised information, RS images from same classes have small feature distances while RS images from different classes have large feature distances in the Hamming space.

In the RS community, Luka et al. [233] proposes a kernelized supervised locality-sensitive hashing (KSLSH) method to address large-scale RS image retrieval. In [234], each RS image is represented by multiple hand-crafted features, which are further mapped to the low-dimensional hash codes via a discrete binary optimization algorithm. In [235], each RS image in the archive is characterized by primitive clusters' descriptors. These descriptors are obtained through an unsupervised approach, which automatically extracts the image regions' descriptors and then associates them with primitive clusters. Furthermore, the primitive clusters' descriptors are transformed into multi-hash codes to represent each RS image. Demir et al. [236] propose a novel class sensitive hashing technique which aims at representing each RS image with multi-hash codes, each of which corresponds to a primitive (i.e., land cover class) present in the RS image. Recently, Kong et al. [237] propose a low-rank hyper-graph hashing (LHH) method to improve the hashing performance where hyper-graphs are able to capture the high-order relationship among data. As a consequence, LHH is suitable to explore the complex structure of RS images.

More advanced, deep hashing based methods which take full advantages of deep networks and hashing learning deliver a better performance for RSIR. In [238], the representational power of the residual net architecture is exploited to establish an end-to-end deep hashing model. The residual hash net is trained subject to a weighted loss strategy that intensifies the cohesiveness of image hash codes within one class. Considering the rotation invariance of the RS target, Zou et al. [239] propose a rotation invariant hashing network that represents an RS image as a binary hash code to accelerate the retrieval process and lift the retrieval accuracy. To mine the pair-wise similarity constraint, Li et al. [240] propose a deep hashing neural network (DHNN) for large-scale RS image retrieval. In such a method, DHNN is optimized by the pair-wise similarity constraint in an end-to-end manner. In [241], a metric and hash-code learning network (MHCLN) was proposed to learn a semantic based metric space, while simultaneously producing binary hash codes for fast and accurate retrieval of RS images in large archives. Song et al. [242] redefine the RS image retrieval problem as visual and semantic retrieval of images. Specifically, a deep hashing CNN is proposed to simultaneously retrieve the similar images and classify their semantic labels in a unified framework. Inspired by generative adversarial networks (GAN), Liu et al. [243] presented a deep supervised hashing model for RS image retrieval. Specifically, a loss function with multiple constraints, including the classification, similarity and bit entropy terms, is proposed to train the generator. In addition, to avoid the case that the leaned hash codes are bit balanced, the unique "true" matrix with the uniform distribution is taken as the input of discriminator. To alleviate the dependency of labeled data, a semi-supervised adversarial hashing method [244] is proposed to address large-scale RS image retrieval. To avoid over-fitting, Roy et al. [245] propose a metric learning-based hashing network, which implicitly re-uses the pre-trained deep CNN without any fine-tune and only focuses on learning the hashing function. Generally, boosting deep hashing with metric learning mainly aims to enlarge the inter-class gap and reduce the intra-class variation. In addition, joint feature hashing learning and attribute prediction [246] also help to alleviate the inter-class confusion and intra-class variation problem. Objectively, the inter-class confusion and intra-class variation is still an open problem and deserves much more exploration.

As well known, traditional deep hashing networks generally tend to be highly expensive in terms of storage space and computing resources and are unsuitable for on-orbit RS image retrieval, which usually operates on resource-limited devices. With this consideration, Li et al. [247] develop a quantized deep learning to hash framework whose weights and activation functions are binarized to low-bit representations, which require comparatively much less storage space and computing resources. In literature, the compact and light-weight deep models [248, 249] have also been exploited in RS image scene classification which is highly related to RS image retrieval. Based on these highly related achievements, RS image retrieval-oriented light-weight deep hashing models can be further improved.

3.3. Cross-modal remote sensing image retrieval

In the RS big data era, we have many different kinds of data, including optical, radar, or laser provided by airplane or satellite or ground sensors. Other kinds of data sources can also be integrated in RS problems. For example, internet textual data (e.g., volunteer geographic information, news, web logs, and so on) [250,251] can be used to help labeling data patterns provided by remote sensors, which involve low or no cost. Also, image data taken by individuals from social networks can be taken into account for assisting in RS data interpretation tasks. Other data formats such as census data, meteorological data, intelligent transportation data, high-fidelity geographical data, healthcare data, and so on, can be of significant help to solve a specific real-world

problem, e.g., monitoring food security. Driven by these advanced RS applications in the big data era, scalable image information mining from heterogeneous generalized RS data (e.g., multi-modal/multi-source RS imagery data and auxiliary data from other domains) is a fundamental task and still deserves a substantial amount of exploration. Until now, lots of cross-modal RS image retrieval methods have been proposed. According to the data types, as illustrated in Fig. 6, cross-modal RS image retrieval includes four main categories: cross-modal retrieval between one kind of RS imagery and another kind of RS imagery (CR-RSI-RSI) [252–255], cross-modal retrieval between RS imagery and sketch (CR-RSI-SKE) [256–261], cross-modal retrieval between RS imagery and text (CR-RSI-TEX) [262], cross-modal retrieval between RS imagery and sound (CR-RSI-SOU) [263–269]. Each kind of cross-modal RS image retrieval is detailed in the following.

CR-RSI-RSI aims to retrieve RS images that have similar contents to the inquiry RS image where the inquiry RS imagery and retrieved RS imagery come from two totally different types of RS data. Apparently, CR-RSI-RSI has to cope with the domain shift problem compared with the traditional content-based image retrieval technique. As the first attempt in this direction, Li et al. [252] collect and release one dual-source RS image dataset (DSRSID), which is composed of two types of RS data (i.e., panchromatic imagery and multi-spectral imagery). To conduct cross-retrieval between panchromatic imagery and multi-spectral imagery, source-invariant deep hashing convolutional neural networks (SIDHCNNs) [252] are proposed to match different kinds of RS imagery in one unified binary feature space. Benefiting from the binary feature representation pursuit, SIDHCNNs are qualified to address the large-scale retrieval case. Instead of measuring in the binary feature space, Chaudhuri et al. [253] recommend to measure in the real-valued feature space, which can achieve an improved retrieval performance. To address the inconsistency between different types of RS data and exploit the intrinsic relation between them, Xiong et al. [254] propose a discriminative distillation network, which aims to enlarge the inter-class variations and reduce the intra-class differences via an alternative learning scheme. From the image generation perspective, the style transfer method via generative adversarial networks (GANs) [255] is proposed to translate one kind of RS imagery to another kind of RS imagery. Furthermore, the inquiry RS imagery can be mapped to the style which is similar to the type of the retrieved RS imagery. Hence, benefiting from this style translation, CR-RSI-RSI is simplified to the traditional content-based image retrieval problem.

To address the unavoidable case that there is no exemplar query RS image available at hand, CR-RSI-SKE aims to retrieve realistic RS images with sketches where the sketch can be directly drawn by users to give an abstract expression of the interested object or scene. In literature, various studies have been conducted to retrieve natural images using sketches [256–259]. Specifically, hand-crafted features such as histogram-of-gradient (HOG) [256] have been adopted to query natural images with sketches. Unsupervised encoding via bag-of-words [257] has also been exploited to address natural image retrieval with sketches. Along with the great success of deep learning, CNN [258] has been modified to address natural image retrieval with sketches and achieves obvious performance improvement compared with hand-crafted features or unsupervised methods. To improve the performance, the

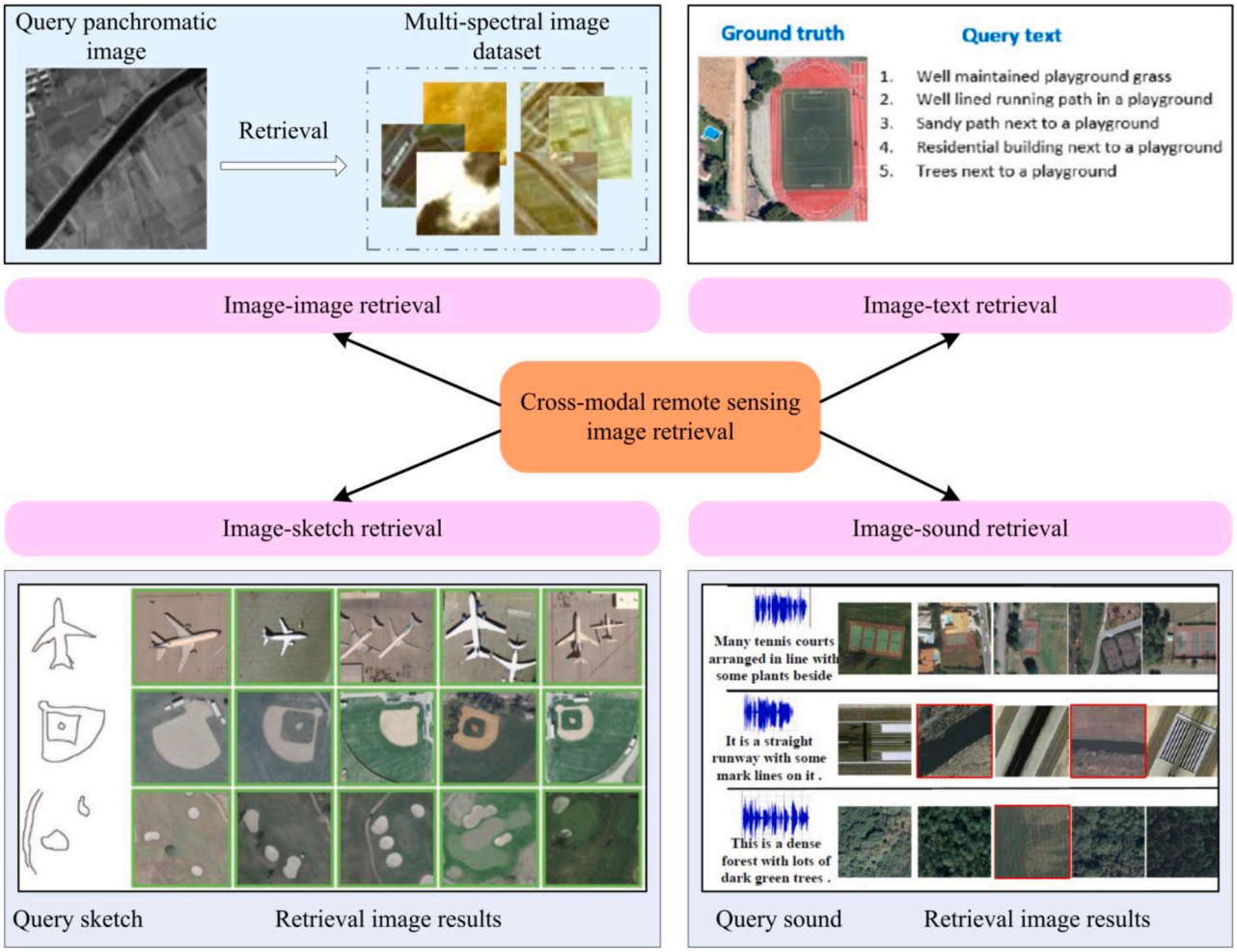

Fig. 6. The main categories of cross-modal remote sensing image retrieval. Some sub-figures are adapted from [252–268].

image-sketch dataset, termed as HUST-SI [259], is proposed to taken as the fuel of deep learning. In a word, it has been demonstrated that cross-domain image-sketch comparison can improve the image retrieval performance when no exemplar query image is available. Although the sketch has been successfully applied in the natural image retrieval, few studies have been devoted to sketch-based RS image retrieval. Owing to the complex surface structures and huge variations of image resolutions, it is very challenging to measure the similarity between such a simple sketch and a fairly complex RS image. The ambiguity inherent in sketches and the gap between aerial images and sketches bring a great difficulty to sketch-based RS image retrieval. The existing methods developed on natural images lose their efficacy when it comes to RS images. To cope with the aforementioned issues, the multi-scale deep cross-domain image representation model [260] has been proposed to conduct sketch-based RS image retrieval and one RS sketch-image database has also been released, which helps a lot to promote the development of the sketch-based RS image retrieval technique. To pursue the domain-invariant representation, one adversarial training strategy [261] is proposed to learn a deep joint embedding space with discriminative losses. In addition, one new sketch-based RS image retrieval dataset and benchmark has been released along with this work.

CR-RSI-TEX [262] aims to explore the correspondence between RS images and natural language descriptions. In contrast to metadata based RS image retrieval which is obtained by matching keywords, CR-RSI-TEX conducts cross-modal retrieval by deeply bridging the visual content of RS images and the natural language descriptions. Given one query natural language description, CR-RSI-TEX tries to search the RS images, whose visual contents are highly related to the query language description, from the RS image dataset where the RS images don't contain any language tags. To achieve cross-modal retrieval between RS images and texts, Abdullah et al. [262] propose a deep bidirectional triplet network, which is composed of Long Short Term Memory network (LSTM) and pre-trained CNNs. To enable learning of robust embedding, an average fusion strategy is proposed to fuse the features pertaining to the five image sentences. As a whole, this direction is in the beginning stages and it deserves much more exploration with the aid of advanced solutions such as attention mechanism.

The goal of CR-RSI-SOU is to leverage RS images or RS sounds to retrieve relevant RS sounds or RS images. In computer vision, the previous methods [263,264] learn the relationship between sounds and images by using shallow projects. However, these shallow projection-based methods cannot capture complex semantic information of sounds and images. To tackle this issue, some deep image-voice retrieval methods [265,266] are proposed to utilize deep neural networks to capture complex semantic information of sounds and images. To cope with the special complexity of RS images, Guo et al. [267] propose a novel cross-modal RS image-voice retrieval approach, which integrates deep feature learning and multi-modal learning into a unified framework for speech-to-image retrieval. To capture more information of RS data to generate hash codes with low memory and fast retrieval properties, hashing-based cross-modal RS image-sound retrieval methods [268,269] have been exploited.

3.4. Interactive remote sensing image retrieval

Due to the high complexity of RS images, only one query image is not sufficient to definitely depict the user's real query intention that leads to the poor retrieval performance. Given this consideration, relevance feedback (RF) has been adopted to iteratively boost the performance of RS image retrieval by taking the user's feedback into account. The general workflow of the interactive RS image retrieval method with RF can be visually shown in Fig. 7. More specifically, the query user firstly

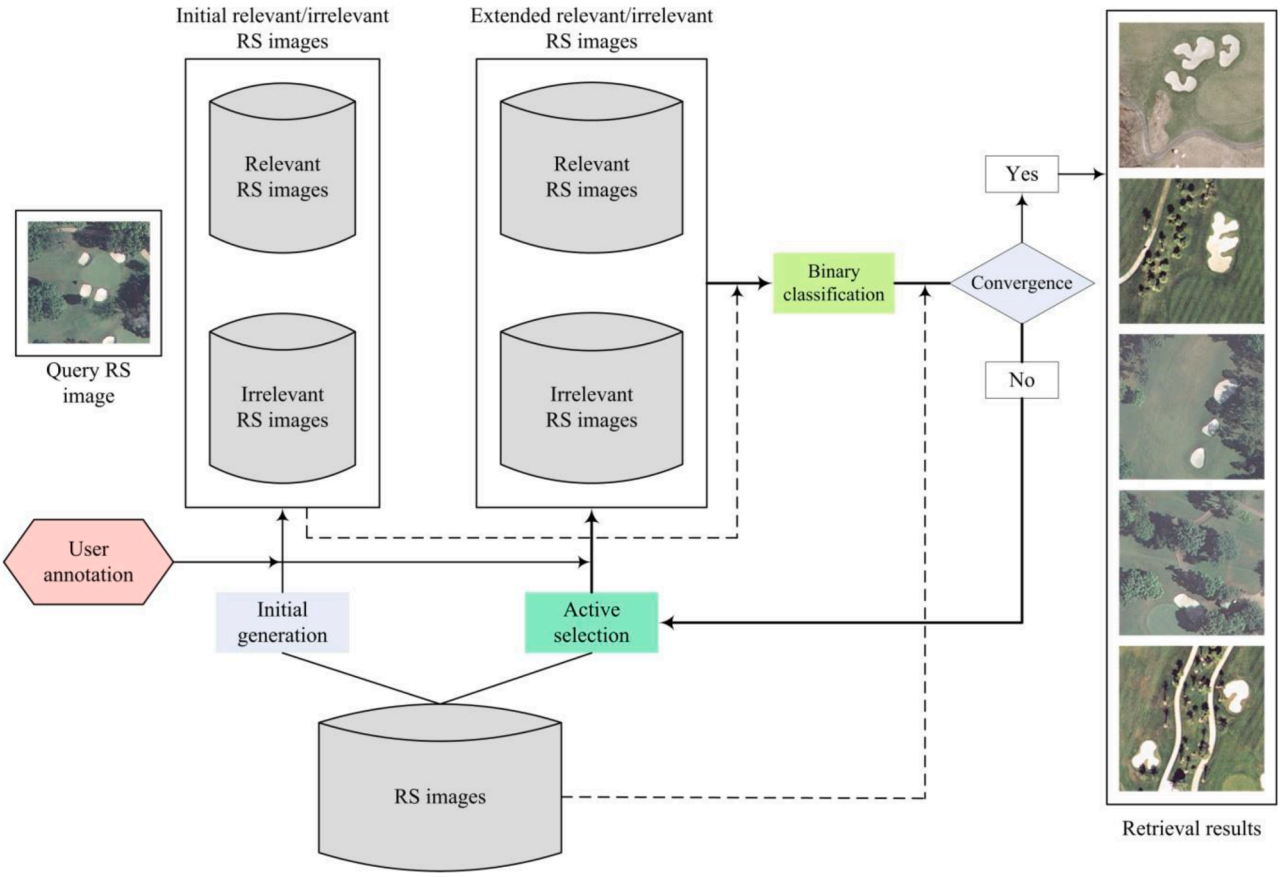

Fig. 7. The flowchart of the interactive remote sensing image retrieval technique.

generates an initial feedback set of relevant/irrelevant RS images based on the query RS image. Based on the relevant/irrelevant RS images, RS image retrieval can be considered as a binary classification problem [270], which can be conducted by kinds of supervised classification methods. In reality, the RF process is repeated several times to extend the set of relevant/irrelevant RS images until that the binary classification algorithm with the extended set converges.

In literature, Bayesian inference-based feedback [271], multiple index weighting-based feedback [272], and binary classification-based feedback [273] have been successively proposed to address interactive optical RS image retrieval. Due to the volume of the iteratively annotated relevant and irrelevant RS images is often small, support vector machine (SVM) has been the mainstream classifier to address this problem [274] because SVM is qualified to output a relatively stable classification result even when the number of labeled samples is very small. In recent years, deep feature has been adopted to improve the binary classification-based interactive RS image retrieval performance [275]. As visually shown in Fig. 7, user annotation is required during the whole feedback process. Obviously, labeling RS images as relevant or irrelevant is time-consuming and thus costly. Accordingly, despite the retrieval success of RF, the conventional RF schemes are not practical and efficient in real applications, especially when large-scale archives of RS images are searched.

A promising approach to reduce the annotation effort in RF is active learning (AL) that aims at finding the most informative RS images in the large-scale RS image archive that, when annotated and included in the set of relevant and irrelevant RS images (i.e., the training set), can significantly improve the retrieval performance as well as lift the interactive effectiveness [276]. One obvious shortcoming of the method [276] is that it does not evaluate the representativeness of RS images in terms of their density in the archive. In fact, RS images that fall into the high-density regions of the image feature (descriptor) space are crucial for CBIR problems particularly when a small number of initially annotated images are available. With this consideration, three criteria including uncertainty, diversity and density have been systematically adopted in the AL phase to effectively and efficiently select the most informative samples [277]. In addition, the importance of normalization in the classification problem with heterogeneous objects is considered to lift the quality of actively selected samples [278]. To ensure the selected RS images are representative and informative enough, multiple different AL algorithms are adopted to conduct different RF processes, and then the contributions of different AL methods are fused using a circular fusion manner [279].

Besides, the RF scheme has also been used to address interactive Synthetic Aperture Radar (SAR) image retrieval [280]. Objectively, it is much harder to represent the visual content of SAR imagery than the optical imagery. To alleviate the negative influence of speckle noise of SAR imagery, superpixel-level texture is adopted to represent the visual content of SAR imagery and one new kernel function is developed to improve the binary classifier in the RF scheme [280]. Recently, RF has been modified to address retrieve the RS change sequence [281]. To cope with the change information retrieval, an improved RF model based on the combination of SVM and genetic algorithm (GA) is proposed where the proposed approach can take consideration of avoiding local maxima in the SVM kernel parameters optimization and the subset feature selection simultaneously by combining GA. As a whole, the RF scheme makes many complex retrieval problems become reality. In addition, experimental results [282] show that the integration of multiple RF methods using reinforcement learning generally outputs better retrieval performance than using only one RF technique.

4. Applications of remote sensing image retrieval

As well known, CBIR was first proposed and used in the computer vision domain. In addition to the classical search engine (e.g., the CBIR function in Google), CBIR also has more applications such as fashion image retrieval on the e-commerce website [283], face verification at the security checkpoints [284], pedestrian re-identification on the video surveillance systems [285]. In the following, we review some distinctive applications based on CBIR in the RS field.

4.1. Retrieval for fusion-oriented image processing

In recent years, RS image retrieval has been successively adopted in multi-source RS image matching, which is a prerequisite of multi-source RS image fusion, and cross-source RS image classification, which is a generalized classification example by fusing multi-source RS images.

As well known, multi-source RS image matching [286–288] is a fundamental task in the RS community. In addition, source-invariant feature descriptors are the key module of this task. Without much expertise or effort in designing descriptors, the aforementioned deep learning-based CR-RSI-RSI methods can automatically learn suitable source-invariant feature descriptors from data, which can be used in multi-source RS image matching. Similar to CR-RSI-RSI methods, Zhu et al. [39] proposed densely-connected CNNs with an augmented cross-entropy loss to match RGB and infrared RS image blocks. As shown in [39], the reported results show improvement on matching rate than the traditional feature matching descriptors such as SIFT, SURF, and so forth.

Due to the diverse distributions of objects and spectral shifts caused by the different acquisition conditions of images, deep networks trained on a certain set of annotated RS images may not be effective when dealing with images acquired by different sensors or from different geolocations. However, retrieval owns a natural source-invariant characteristic to some degree. Tong et al. [40] proposed one retrieval–based cross-source RS image classification method. Specifically, a deep CNN model is first pre-trained with a well-annotated land-use dataset, referred to as the source data. Then, given a target image with no labels, the pre-trained CNN model is utilized to classify the image in a patch-wise manner. The patches with a high classification probability are assigned with pseudo-labels and employed as the queries to retrieve related samples from the source data. The pseudo-labels confirmed with the retrieved results are regarded as supervised information for fine-tuning the pre-trained deep model. Extensive experiments show encouraging results and demonstrate the efficiency of the proposed retrieval-based scheme for learning transferable deep models for RS image classification.

4.2. Retrieval for geo-localization and navigation

Another application of RS image retrieval is CBIR-based localization where the key problem is to find the geo-location of one query image by finding its nearest referenced images. In this application, it is assumed that there exists one large-scale referenced dataset with massive geotagged images in advance. When the query image is captured on the ground surface with the daily life view, this problem is often call geolocalization [289–294] and widely exploited in the computer vision community. By contrast, when the query image is captured on the sky with the aircraft view, this problem is termed as visual navigation [41,42,295] and attracts more attention of the RS community.

In the early stages, hand-crafted features have been adopted to retrieve the nearest referenced images for geo-localization [289,290]. Inspired by the great success, deep feature is utilized to improve the retrieval performance [291] for further lifting the geo-localization accuracy. By considering the 3D scene geometric attributes, the geo-localization accuracy can be further improved based on the strict 3D retrieval model [292] and the flexible 3D retrieval method [293]. As well known, collecting the geo-tagged referenced image dataset is time-consuming and becomes prohibited when the volume of the collected dataset tends to be very large. Given this consideration, Hu et al. [294] exploit the widely available geo-referenced aerial images to replace the ground-based reference dataset and shows the effectiveness

of the cross-view matching network for matching the ground-based query image and the aerial images. Even though GPS is not available, these geo-localization methods can tell one person where she/he locates as long as she/he captures one street view image around her/his location.

By contrast to geo-localization, visual navigation [41,42,295] aims to recover the geographical location of the aerial imaging sensor based on scene matching (i.e., image retrieval) between the captured aerial image (i.e., the query image), and the geo-referenced aerial/satellite images. Considering that the aerial vehicles are often moving at a very high speed, the retrieval process should be done in real time. In addition, visual navigation has to cope with the cross-source retrieval case when the captured imagery and the referenced RS imagery come from different modalities. As mentioned before, deep hashing-based methods could project the image into the low-dimensional binary feature vector, which benefits accelerating the searching process. With this consideration, deep hashing and deep cross-modal hashing would be reasonable ways to address the real-time visual navigation task. Hence, it deserves much exploration about how to improve the visual navigation task with the aid of deep hashing.

4.3. Retrieval for disaster rescue

RS image retrieval plays an important role on disaster rescue. In the following, we discuss two application cases including coastal flood [43] and terrorist attack [114].

Obviously, RS observations comprise a significant portion of the data used by coastal zone monitoring systems. These observation data is particularly valuable because it provides a variety of measurements that are not otherwise available or affordable. However, the use of such valuable information in a rapid assessment scenario is hindered by the fact that it is cumbersome to explore huge RS image databases through manual operations. In a coastal disaster event, it is necessary to obtain information in real time and predictions of water level, storm surge in advance. The dissemination of information that is time critical calls for systems that will facilitate quick assessment of the scenario from multiple perspectives. Hence, the rapid retrieval of the status of different land covers using RS data becomes more and more urgent. To this end, Durbha et al. [43] propose a Rapid Image Information Mining (RIIM) system, which is a region based approach. It localizes interesting zones and extracts characteristic information from them and stores this information in a database for later use during the disaster. This content is then available for a variety of queries based on the image content for searching relevant RS imagery.

In September 2001, the terrorist attacks collapsed the two main towers and other buildings in the World Trade Center (WTC) area in New York City. During the last two weeks of the attack time, a dataset containing 154 high-resolution hyperspectral images with more than 20 TB of data has been gathered by NASA over the WTC area. The hyperspectral imagery has 224 spectral bands. In the retrieval test, the hyperspectral image covering the area, centered at the region where the towers collapsed, is taken as the query RS image. Hence, the main challenge of this retrieval example is how to cope with the voluminous challenge of RS big data. With this consideration, Plaza et al. [104] propose a parallel CBIR system to investigate the parallel properties. The parallel CBIR system successfully retrieved all of the hyperspectral images containing the WTC complex across the dataset. It is worth noting that the retrieval results don't contain any false positives. More specifically, the parallel CBIR technique is implemented on a system composed of 256 dual 2.4-GHz Intel Xeon nodes, each with 1 GB of memory and 80 GB of main memory. Using 256 processors on Thunderhead, the CBIR system can retrieve the most similar hyperspectral images across the full database in only 4 s, resulting in a total processing of approximately 10 s to catalog and fully describe a new entry in the dataset. This application represents a significant improvement over the implementation of the same CBIR process on a single Thunderhead processor, which took over 1 hour of computation for the same operation. Hence, HPC would be a promising solution to address image retrieval from RS big data by enhancing the conventional CBIR techniques.

5. Datasets and performance evaluation for remote sensing image retrieval

In this section, the available datasets, evaluation metrics and performance discussion for RS image retrieval are depicted in detail.

5.1. Datasets for remote sensing image retrieval

In the following, we summarize the existing available RS image retrieval datasets with one single modality in Table 1 and RS datasets with two or more modalities in Table 2. As aforementioned, the RS datasets in Table 1 can support the single-label uni-source RS image retrieval task and multi-label uni-source RS image retrieval task. In addition, RS datasets in Table 2 can be used to evaluate the cross-source RS image retrieval methods.

As shown in Table 1, the most majority of existing RS datasets is constructed based on the optical RS imagery with R-G-B bands, which mainly benefits from the free access characteristic of Google Earth Imagery. Compared with the multi-spectral imagery-driven datasets, the datasets based on SAR or hyper-spectral imagery are relatively scarce. To give a full description of the intrinsic content of RS imagery, multi-label RS datasets (e.g., UCM* [305], AID* [306] and BigEarthNet [307]) have been collected where each RS image scene is annotated with multiple scene-level labels. In addition, some multi-label RS retrieval methods [308,309] have been proposed based on these multi-label RS datasets. As BigEarthNet [307] is automatically labeled with the aid of publicly open land-cover products, the labels of BigEarthNet may contain a certain degree of errors. Hence, researchers should carefully address this case when they try to design RS retrieval methods based on this noisy dataset.

Driven by the urgent requirements for deploying the hybrid RS data, more and more researchers turn to the cross-source RS retrieval task. Specifically, DSRSID [252] and SEN12MS [310] can be adopted to design and evaluate the CR-RSI-RSI methods. Based on Aerial-SI [260] and RSketch [261], the effectiveness of CR-RSI-SKE methods can be verified. TextRS [262] is specifically collected to promote the CR-RSI-TEX technique. In addition, both of CR-RSI-TEX and CR-RSI-SOU methods can be evaluated on UCM-Captions [311], Sydney-Captions [311] and RSICD-Captions [312] as each of these datasets simultaneously contains the RS imagery, the sentence and the sound. As a whole, the volume of SEN12MS is relatively large, but it also suffers from the noisy labels. Except SEN12MS, the volume of other multi-modality RS image retrieval datasets is relatively small.

5.2. Evaluation metrics for remote sensing image retrieval

In the following, we review the widely adopted evaluation metrics for RS image retrieval. Based on the ground-truth (GT) dataset (e.g., the aforementioned datasets in Table 1 and Table 2), given one query, four widely adopted evaluation metrics include Precision-Recall Curve (PRC) [226], Precision@k [226], Mean Average Precision (MAP) [226] and Average Normalized Modified Retrieval Rank (ANMRR) [162].

Given one query with the known label information in advance, Precision@k reflects the consistency rate that k returned results share the same label with the query. In addition, the MAP score can be calculated by:

$$MAP = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{n_i}\sum_{k=1}^{n_i} precision(R_{ik}) \tag{1}$$

where $q_i \in Q$ stands for one query and n_i denotes the number of returned

Table 1
RS image retrieval datasets based on one single data modality.

Dataset	Data modality	Volume of images	Image size	Annotation information	Task
UCM in [296]	Multi-spectral optical RS imagery with R-G-B bands	2100	256×256	Single-label from 21 categories	Single-label uni-source retrieval
WHU-RS19 in [149]	Multi-spectral optical RS imagery with R-G-B bands	1005	600×600	Single-label from 19 categories	Single-label uni-source retrieval
RSSCN7 in [297]	Multi-spectral optical RS imagery with R-G-B bands	2800	400×400	Single-label from 7 categories	Single-label uni-source retrieval
SIRI-WHU in [298]	Multi-spectral optical RS imagery with R-G-B bands	2400	200×200	Single-label from 12 categories	Single-label uni-source retrieval
AID in [299]	Multi-spectral optical RS imagery with R-G-B bands	10,000	600×600	Single-label from 30 categories	Single-label uni-source retrieval
NWPU-RESISC45 in [300]	Multi-spectral optical RS imagery with R-G-B bands	31,500	256×256	Single-label from 21 categories	Single-label uni-source retrieval
PatternNet in [65]	Multi-spectral optical RS imagery with R-G-B bands	30,400	256×256	Single-label from 38 categories	Single-label uni-source retrieval
RSI-CB128 in [301]	Multi-spectral optical RS imagery with R-G-B bands	36,707	128×128	Single-label from 45 categories	Single-label uni-source retrieval
RSI-CB256 in [301]	Multi-spectral optical RS imagery with R-G-B bands	24,747	256×256	Single-label from 35 categories	Single-label uni-source retrieval
AID++ in [302]	Multi-spectral optical RS imagery with R-G-B bands	Over 400,000	600×600	Single-label from 46 categories	Single-label uni-source retrieval
SAT-4 in [303]	Multi-spectral optical RS imagery with R-G-B-NIR bands	500,000	64×64	Single-label from 4 categories	Single-label uni-source retrieval
SAT-6 in [303]	Multi-spectral optical RS imagery with R-G-B-NIR bands	405,000	64×64	Single-label from 6 categories	Single-label uni-source retrieval
EuroSat in [304]	Multi-spectral optical RS imagery with 13 bands	27,000	64×64	Single-label from 10 categories	Single-label uni-source retrieval
SAR-14 in [280]	SAR imagery with the amplitude band	15,728	256×256	Single-label from 14 categories	Single-label uni-source retrieval
ICONES-HIS in [111]	Hyperspectral imagery with 224 bands	486	300×300	Single-label from 9 categories	Single-label uni-source retrieval
UCM* in [305]	Multi-spectral optical RS imagery with R-G-B bands	2100	256×256	Multi-label from 17 categories	Multi-label uni-source retrieval
AID* in [306]	Multi-spectral optical RS imagery with R-G-B bands	3000	600×600	Multi-label from 17 categories	Multi-label uni-source retrieval
BigEarthNet in [307]	Multi-spectral optical RS imagery with 13 bands	590,326	20×20–120×120	Multi-label from 43 categories	Multi-label uni-source retrieval

Table 2
RS image retrieval datasets based on two or more data modalities.

Dataset	Data modality	Volume of images	Image size	Annotation information	Task
DSRSID in [252]	Modality 1: Multi-spectral optical RS imagery with R-G-B-NIR bands Modality 2: Panchromatic optical RS imagery with one band	60,000	64×64–256×256	Single-label from 6 categories	Cross-source retrieval
SEN12MS in [310]	Modality 1: Multi-spectral optical RS imagery with 13 bands Modality 2: SAR RS imagery with VV-VH bands	180,662	256×256	Multi-label from global land cover maps	Cross-source retrieval
Aerial-SI in [260]	Modality 1: Multi-spectral optical RS imagery with R-G-B bands Modality 2: Each RS imagery with one salient object sketch	3300	600×600	Single-label from 10 categories	Cross-source retrieval
RSketch in [261]	Modality 1: Multi-spectral optical RS imagery with R-G-B bands Modality 2: Each category contains 45 sketches	2000	256×256	Single-label from 20 categories	Cross-source retrieval
TextRS in [262]	Modality 1: Multi-spectral optical RS imagery with R-G-B bands Modality 2: Each RS imagery with 5 sentences	2144	256×256	–	Cross-source retrieval
UCM-Captions in [311]	Modality 1: Multi-spectral optical RS imagery with R-G-B bands Modality 2: Each RS imagery with 5 sentences Modality 3: Each RS imagery with one sound	2100	256×256	Single-label from 21 categories	Cross-source retrieval
Sydney-Captions in [311]	Modality 1: Multi-spectral optical RS imagery with R-G-B bands Modality 2: Each RS imagery with 5 sentences Modality 3: Each RS imagery with one sound	613	500×500	Single-label from 7 categories	Cross-source retrieval
RSICD-Captions in [312]	Modality 1: Multi-spectral optical RS imagery with R-G-B bands Modality 2: Each RS imagery with 5 sentences Modality 3: Each RS imagery with one sound	10,921	224×224	Single-label from 30 categories	Cross-source retrieval

results relevant to q_i in the dataset. Suppose the relevant results are ordered as $\{r_1, r_2, \cdots, r_{n_i}\}$, and then R_{ik}stands for the set of ranked retrieval results from the top result to r_k.

ANMRR considers both the number and order of the ground truth items that appear in the top retrievals. It is assumed that q_i stands for one query with a GT size of $NG(q_i)$. The $Rank(k)$ of the k-th GT item is defined as the position at which it is retrieved. A number $K(q_i) \geq NG(q_i)$ is chosen so that items with a higher rank are given a constant penalty:

$$Rank(k) = \begin{cases} Rank(k), if\, Rank(k) \leq K(q_i) \\ 1.25 * K(q_i), if\, Rank(k) > K(q_i) \end{cases} \tag{2}$$

where $K(q_i)$ is commonly chosen to be $2*NG(q_i)$. The average rank (AVR) for a single query q_i is then computed as:

$$AVR(q_i) = \frac{1}{NG(q_i)} \sum_{k=1}^{NG(k)} Rank(k) \tag{3}$$

To eliminate influences of different $NG(q_i)$, $NMRR(q_i)$ is further calculated by:

$$NMRR(q_i) = \frac{AVR(q_i) - 0.5*(1 + NG(q_i))}{1.25 * K(q_i) - 0.5 * (1 + NG(q_i))} \tag{4}$$

Given a set of queries Q, ANMRR can be computed by Eq. (5) taking the average over Q. Different from the MAP score, ANMRR ranges in value between zero to one with lower values indicating better retrieval performance.

$$NMRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} NMRR(q_i) \tag{5}$$

Besides these evaluation metrics, some improved evaluation measures based on the specific tasks have been also widely used. For example, to fairly evaluate the hashing-based RS image retrieval methods, the aforementioned metrics are often used under the length condition (i.e., the length of hashing features) [240]. When evaluating the interactive RS image retrieval methods, both of the number of feedback operations and the aforementioned metrics are jointly considered.

5.3. Performance discussion of remote sensing image retrieval

In this section, we review two benchmarks about large-scale RS image retrieval including uni-source RS image retrieval and cross-source RS image retrieval, which may guide researchers to directly find the suitable methods to address their tasks.

In the first benchmark, we discuss the performance of uni-source RS image retrieval methods. In this benchmark, we consider the UCM dataset [296] as it has been widely used to evaluate RS image retrieval methods. In order to augment the UCM dataset, the original RS images are rotated by 90˚, 180˚, and 270˚, separately. In this way, the volume of the UCM dataset is increased to 8400. In this evaluation, 7400 images are used as the retrieval database and to train retrieval methods and the remaining 1000 images are used as query data for testing.

In this benchmark, the competitive baselines include the specific hashing methods for RS image retrieval and the general hashing methods for natural image retrieval. Specifically, the hashing methods proposed for RS image retrieval in recent years include PRH [226], KULSH [215], KSLSH [233], DHNN [240], and QDLH [247]. Moreover, we also consider some representative hashing methods used in the computer vision field in the experiments, such as LSH [214], supervised discrete hashing (SDH) [313], column sampling-based discrete supervised hashing (COSDISH) [314], deep supervised hashing (DSH) [315], deep hashing net (DHN) [316], and deep pairwise-supervised hashing (DPSH) [317]. Among the various comparison methods, LSH, KULSH, and PRH are unsupervised hashing methods that do not use label information for hash code generation, and the rest are supervised methods. Moreover, LSH, KULSH, PRH, KSLSH, SDH, and COSDISH are shallow methods and the remaining ones are deep hashing methods based on CNNs. For a fair comparison, we use the FC-7 feature of a pre-trained AlexNet as the input for the shallow methods and the raw RS images as inputs for the deep models. As a whole, the quantitative evaluation results have been summarized in Table 3.

As depicted in Table 3, the supervised methods can significantly outperform the unsupervised methods including LSH, KULSH and PRH. In addition, the supervised methods with a shallow architecture achieve promising results because the pre-trained deep features are taken as the inputs of these methods. Due to lacking the specific consideration of RS image characteristics, the performance of the deep hashing methods in the computer vision domain, including DSH, DHN and DPSH, is still very limited. By contrast, the deep hashing models including DHNN [240] and QDLH [247], specifically designed for RS image retrieval, obviously outperform other methods. Moreover, the weights and activation functions in the QDLH framework are binarized to low-bit representations, which require much less storage and computing resources, which makes QDLH be suitable for on-orbit RS image retrieval.

In the second benchmark, we review the performance of cross-source RS image retrieval methods. Here, we consider the DSRSID dataset [252] as it has been specifically collected to evaluate cross-source RS image retrieval methods. DSRSID is composed of a great quantity of pairs of panchromatic (PAN) and multi-spectral (MUL) images which are acquired by the GF-1 multi-spectral sensor and GF-1 panchromatic sensor, respectively. It includes 80,000 pairs of multi-source images of 8 land-cover categories, including aqua-farm, cloud, forest, high building, low building, farm land, river, and water. Each category contains 10,000 pairs of multi-source images. The size of multi-spectral image is 64×64 with a resolution of 8 m, and the number of spectral channels is 4. While the size of panchromatic image is 256×256 with a spatial resolution of 2 m, the number of spectral channels is 1. In this setting, 75,000 multi-source images are used as the retrieval database and to train retrieval methods and the remaining 5000 images are used as the query data for testing. It is worth noting that the query image and images from the retrieval database come from different sources.

In this benchmark, the evaluation methods include the hashing-based cross-source RS image retrieval methods and some recently proposed cross-source RS image retrieval methods with the high-dimensional retrieval feature. Specifically, the hashing-based cross-source RS image retrieval methods include canonical correlation analysis (CCA) [213], semantic correlation maximum (SCM) [318], deep cross-modal hashing (DCMH) [319] and SIDHCNNs [252]. In addition, two recently proposed cross-source RS image retrieval methods [254, 255] including multiple network-driven variants are also considered. As a whole, both CCA and SCM adopt the hand-crafted features where CCA is trained via an unsupervised manner but SCM is trained by a supervised way. DCMH is first proposed in the computer vision domain and can be trained in an end-to-end manner. Here, we transfer it to the

Table 3
MAP comparison of large-scale RS image retrieval methods with different hashing lengths on UCM.

Methods	Architecture	Supervision	The length of hashing codes		
			32 bits	64 bits	96 bits
LSH in [214]	Shallow	Unsupervised	0.3886	0.5141	0.5540
KULSH in [215]	Shallow	Unsupervised	0.5379	0.6246	0.6566
PRH in [226]	Shallow	Unsupervised	0.5717	0.6561	0.6769
KSLSH in [233]	Shallow	Supervised	0.8874	0.9023	0.9128
SDH in [313]	Shallow	Supervised	0.9119	0.9342	0.9320
COSDISH in [314]	Shallow	Supervised	0.8713	0.8704	0.8776
DSH in [315]	Deep	Supervised	0.6317	0.6750	0.7502
DHN in [316]	Deep	Supervised	0.6707	0.7313	0.7707
DPSH in [317]	Deep	Supervised	0.7478	0.8174	0.8640
DHNN in [240]	Deep	Supervised	0.9396	0.9718	0.9762
QDLH in [247]	Deep	Supervised	0.9681	0.9764	0.9846

cross-source RS image retrieval task. As aforementioned, SIDHCNNs is specifically designed for cross-source RS image retrieval. Besides these hashing-based methods, we also evaluate the style transfer-based CI-GAN+VGG16 and CI-GAN+VGG19 [255], and the discriminative distillation-based Distillation-Res18 and Distillation-Res50 [254] on the DSRSID dataset. Based on two cross-source retrieval modes including PAN → MUL and MUL → PAN, all of the existing methods have been evaluated on the DSRSID dataset and the quantitative evaluation values are summarized in Table 4. It is noted that the evaluation results of CI-GAN+VGG16, CI-GAN+VGG19, Distillation-Res18 and Distillation-Res50 are generated based on their actual retrieval feature dimension as they don't belong to the hashing-based methods.

As depicted in Table 4, the supervised methods overall perform better than the unsupervised method (i.e., CCA), which reflects that the supervision plays a critical role on the cross-source retrieval task. Among the supervised methods, SCM with the shallow architecture lags behind the other methods using the deep architecture. Due to the adoption of more supervised constraints in the objective function, SIDHCNNs obviously outperforms DCMH under different lengths of hashing codes. CI-GAN+VGG16, CI-GAN+VGG19, Distillation-Res18, and Distillation-Res50 slightly outperform the SIDHCNNs, but all of them depend on the high-dimensional retrieval feature, which restricts their applications on the large-scale retrieval task. To address the small-scale cross-source RS image retrieval, CI-GAN+VGG16, CI-GAN+VGG19, Distillation-Res18, and Distillation-Res50 would be good choices. However, to conduct cross-source retrieval from the over-sized RS image dataset, SIDHCNNs is recommended.

6. Promising research directions

In this section, we point out some promising research directions along the RS image retrieval avenue.

6.1. Developing larger remote sensing image retrieval datasets

Objectively speaking, deep learning has been the mainstream technique of the state-of-the-art RS image retrieval methods [66]. As well known, the great success of deep learning highly depends on the quality and volume of annotation data [80]. However, the current RS image retrieval datasets are still very limited in terms of the volume of samples, the number of categories and the number of modalities. As a consequence, it's very hard to train deep networks from scratch. Although transferring the pre-trained deep networks can avoid the lack of annotated data, fully training deep networks using sufficient labeled data is still the optimal solution as, in many cases, the RS data from the target domain may be totally unique. In addition, most of the current RS image retrieval datasets are collected based on the freely Google Earth Imagery, which can't fully reflect the characteristic of satellite imagery. Hence, developing new large-scale RS image retrieval datasets with fine-grained categories becomes more and more urgent. Based on the previous experience, unsupervised aggregation [320] or active learning [321] would be reasonable techniques to accelerate the annotation process while maintaining the high annotation accuracy.

6.2. Weakly supervised deep learning for remote sensing image retrieval

In the RS big data era, we can easily collect a large amount of raw data, but accurately labeling oversized data becomes the real challenge because there exist so many kinds of RS images compared with the fixed R-G-B format of natural images in the computer vision domain. Hence, how to train deep networks with weak supervision (e.g., a relatively limited number of labeled samples or auxiliary data even containing a certain degree of noisy labels) will be promising research directions towards addressing image retrieval from RS big data. In the RS image classification tasks, semi-supervised deep learning [322] has been widely exploited to improve the classification performance by fully leveraging the limited number of labeled samples. In addition, error-tolerant deep learning methods [91,323–325] adopt the error-robust loss function or the error-label correction strategy to robustly learn deep networks from RS image datasets with noisy labels. Although RS image retrieval suffers from the similar data issue that RS image classification also meets, weakly supervised deep learning is seldom exploited in the context of RS image retrieval. Hence, weakly supervised deep learning would be a promising way to improve the performance of RS image retrieval based on the current limited but already available data.

6.3. Visual reasoning for remote sensing image retrieval

With massive training data and powerful computing resources, the key advantage of deep neural networks is the end-to-end design that generalizes to a large spectrum of domains, minimizing the human efforts in domain specific knowledge engineering. However, large gaps between human and machines can be still observed in 'high-level' vision-language tasks. In particular, recent studies in the RS community show that the end-to-end models are easily optimized to conduct cross image-text retrieval [262], image caption [326–328] and visual question answering [329]. However, these works still do not involve in visual reasoning [330–333] which attempts to understand the topological graph from the structured raw image and deductively draw inferences via conceptual rules and statements to proceed from known facts to novel conclusions. As well known, humans are not only capable of

Table 4
MAP comparison of cross-modal RS image retrieval methods on DSRSID.

Methods	Architecture	Supervision	The cross-source retrieval mode	The length of hashing codes	
				16 bits	32 bits
CCA in [213]	Shallow	Unsupervised	PAN → MUL	0.1593	0.1502
			MUL → PAN	0.1594	0.1505
SCM in [318]	Shallow	Supervised	PAN → MUL	0.3472	0.3767
			MUL → PAN	0.3671	0.3871
DCMH in [319]	Deep	Supervised	PAN → MUL	0.8076	0.8509
			MUL → PAN	0.8023	0.8445
SIDHCNNs in [252]	Deep	Supervised	PAN → MUL	0.9552	0.9643
			MUL → PAN	0.9725	0.9789
CI-GAN+VGG16 in [255]	Deep	Supervised	PAN → MUL	0.9766	
			MUL → PAN	0.9683	
CI-GAN+VGG19 in [255]	Deep	Supervised	PAN → MUL	0.9731	
			MUL → PAN	0.9652	
Distillation-Res18 in [254]	Deep	Supervised	PAN → MUL	0.9697	
			MUL → PAN	0.9701	
Distillation-Res50 in [254]	Deep	Supervised	PAN → MUL	0.9798	
			MUL → PAN	0.9811	

learning, but also talented at reasoning. Given one high-level question retrieval, it can be rationally expected that visual reasoning-driven RS image retrieval system can not only accurately return the expected RS images, but also semantically gives the question answer. Hence, visual reason deserves much more exploration in the context of RS image retrieval.

7. Conclusion

As one of the most fundamental and important tasks in RS big data mining, image retrieval (i.e., image information mining) from RS big data has attracted continuous research interests in the last several decades. This paper first discusses the opportunities and challenges of image retrieval from RS big data, then systematically reviews the emerging achievements. Besides the conventional CBIR application, this paper points out several RS-domain-specific applications based on RS image retrieval. To facilitate the quantitative evaluation of the RS image retrieval technique, the paper gives a list of publicly open datasets and evaluation metrics. In addition, it also gives some comments on two representative RS image retrieval benchmarks, which may help researchers intuitively find the mainstream methods over different RS image retrieval tasks. It is worth noting this paper emphasizes some interesting applications driven by RS image retrieval. Finally, it also gives some promising research directions of RS big data mining, which may guide young researchers to find the key unaddressed problems in a short time and attract more scientists in the related domains to collaboratively address these issues.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under grants 41971284 and 61773295; the Hubei Provincial Natural Science Foundation of China under grants 2018CFB501 and 2019CFA037; the China Postdoctoral Science Foundation under grants 2016M590716 and 2017T100581; and the Fundamental Research Funds for the Central Universities under grant 2042020kf0218.

Reference

[1] C. Lynch, Big data: how do your data grow? Nature 455 (7209) (2008) 28–29.
[2] S. Ramirez-Gallego, A. Fernandez, S. Garcia, M. Chen, F. Herrera, Big data: tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce, Inf. Fusion 42 (2018) 51–61.
[3] A. Kleiner, A. Talwalkar, P. Sarkar, M. Jordan, The big data bootstrap, in: Proceedings of ICML, 2012, pp. 1–8.
[4] Q. Zhang, L. Yang, Z. Chen, P. Li, A survey on deep learning for big data, Inf. Fusion 42 (2019) 146–157.
[5] G. Bello-Orgaz, J. Jung, D. Camabo, Social big data: recent achievements and new challenges, Inf. Fusion 28 (2016) 45–59.
[6] J. Liu, et al., Urban big data fusion based on deep learning: an overview, Inf. Fusion 53 (2020) 123–133.
[7] V. Marx, Biology: the big challenges of big data, Nature 498 (7453) (2013) 255–260.
[8] J. Ford, S. Tilleard, L. Berrang-Ford, M. Araos, R. Biesbroek, A. Lesnikowski, G. MacDonald, A. Hsu, C. Chen, L. Bizikova, Opinion: big Data has Big Potential for Applications to Climate Change Adaptation, Proc. Natl Acad. Sci. 113 (39) (2016) 10729–10732.
[9] A. Karpatne, V. Kumar, Big data in climate: opportunities and challenges for machine learning, in: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017.
[10] J. Lee, M. Kang, Geospatial big data, Big Data Res. 2 (2015) 74–81.
[11] M. Chi, A. Plaza, J. Benediktsson, Z. Sun, J. Shen, Y. Zhu, Big data for remote sensing: challenges and opportunities, Proc. IEEE 104 (2016) 2207–2219.
[12] L. Wang, H. Zhong, R. Ranjan, A. Zomaya, P. Liu, Estimating the statistical characteristics of remote sensing big data in the wavelet transform domain, IEEE Trans Emerg Top Comput 2 (3) (2014) 324–337.
[13] P. Liu, L. Di, Q. Du, L. Wang, Remote sensing big data: theory, methods, and applications, Remote Sens (Basel) 10 (2018) 711.
[14] H. Guo, L. Wang, F. Chen, D. Liang, Scientific big data and digital earth, Chin. Sci. Bull. 59 (35) (2014) 5066–5073.
[15] H. Guo, Z. Liu, H. Jiang, C. Wang, J. Liu, D. Liang, Big Earth Data: a new challenge and opportunity for digital earth's development, Int. J. Digit. Earth 10 (1) (2017) 1–12.
[16] M. Sudmanns, et al., Big earth data: disruptive changes in earth observation data management and analysis? Int. J. Digit. Earth (2019) in press.
[17] M. Wulder, et al., Current status of Landsat program, science, and applications, Remote Sens Environ 225 (2019) 127–147.
[18] L. Carrasco, A. O'Neil, R. Morton, C. Rowland, Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with google earth engine, Remote Sens (Basel) 11 (3) (2019) 288.
[19] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, R. Moore, Google earth engine: planetary-scale geospatial analysis for everyone, Remote Sens Environ 202 (2017) 18–27.
[20] P. Baumann, et al., Big data analytics for earth sciences: the earthserver approach, Int. J. Digit. Earth 9 (1) (2016) 3–29.
[21] T. Esch, et al., Exploiting big earth data from space – first experiences with the timescan processing chain, Big Earth Data 123 (12) (2018) 1–20.
[22] C. Lee, et al., Recent developments in high performance computing for remote sensing: a review, IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 4 (3) (2011) 508–527.
[23] L. Mascolo, M. Quartulli, P. Guccione, G. Nico, and I. Olaizola. Distributed mining of large scale remote sensing image archives on public computing infrastructures. arXiv. 2015, arXiv:1501.05286.
[24] Y. Ma, H. Wu, L. Wang, B. Huang, R. Ranjan, A. Zomaya, W. Jie, Remote sensing big data computing: challenges and opportunities, Future Generation Computer Systems 51 (2015) 47–60.
[25] D. Lunga, J. Gerrand, H. Yang, C. Layton, and R. Stewart. Apache spark accelerated deep learning inference for large scale satellite image analytics. arXiv. 2019, arXiv:1908.04383.
[26] H. Xia, W. Huang, N. Li, J. Zhou, D. Zhang, PARSUC: a parallel subsampling-based method for clustering remote sensing big data, Sensors 19 (2019) 3438.
[27] A. Sedaghat, M. Mokhtarzade, H. Ebadi, Uniform robust scale-invariant feature matching for optical remote sensing images, IEEE Trans Geosci. Remote Sens 49 (11) (2011) 4516–4527.
[28] Jiayi Ma, et al., Robust feature matching for remote sensing image registration via locally linear transforming, IEEE Trans. Geosci. Remote Sens 53 (12) (2015) 6469–6481.
[29] Jiayi Ma, J. Zhao, J. Jiang, H. Zhou, X. Guo, Locality preserving matching, Int J Comput Vis 127 (5) (2019) 512–531.
[30] Y. Ye, J. Shan, L. Bruzzone, L. Shen, Robust registration of multimodal remote sensing images based on structural similarity, IEEE Trans Geosci Remote Sens. 55 (5) (2017) 2941–2958.
[31] Jiayi Ma, et al., Guided locality preserving feature matching for remote sensing image registration, IEEE Trans Geosci. Remote Sens 56 (8) (2018) 4435–4447.
[32] W. Ma, J. Zhang, Y. Wu, L. Jiao, H. Zhu, W. Zhao, A novel two-step registration method for remote sensing images based on deep and local features, IEEE Trans. Geosci. Remote Sens 57 (7) (2019) 4834–4843.
[33] Jiayi Ma, Wei Yu, Chen Chen, Pengwei Liang, Xiaojie Guo, Junjun Jiang, Pan-GAN: an unsupervised pan-sharpening method for remote sensing image fusion, Inf. Fusion 62 (2020) 110–120.
[34] M. Schmitt, L. Hughes, X. Zhu, The SEN1-2 dataset for deep learning in SAR-optical data fusion, in: Proceedings of ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2018, pp. 141–146.
[35] G. Simone, A. Farina, F. Morabito, S. Serpico, L. Bruzzone, Image fusion techniques for remote sensing applications, Inf. Fusion 3 (1) (2002) 3–15.
[36] H. Ghassemian, A review of remote sensing image fusion methods, Inf. Fusion 32 (2016) 75–89.
[37] L. Deng, M. Feng, X. Tai, The fusion of panchromatic and multispectral remote sensing images via tensor-based sparse modeling and hyper-Laplacian prior, Inf. Fusion 52 (2019) 76–89.
[38] P. Du, S. Liu, J. Xia, Y. Zhao, Information fusion techniques for change detection from multi-temporal remote sensing images, Inf. Fusion 14 (2013) 19–27.
[39] R. Zhu, D. Yu, S. Ji, M. Lu, Matching RGB and Infrared remote sensing images with densely-connected convolutional neural networks, Remote Sens (Basel) 11 (2019) 2836.
[40] X. Tong, G. Xia, Q. Lu, H. Shen, S. Li, S. You, L. Zhang, Land-cover classification with high-resolution RS images using transferable deep models, Remote Sens Environ (2019) in press.
[41] Q. Yu, et al., Full-parameter vision navigation based on scene matching for aircrafts, Sci. China Inf. Sci. 57 (5) (2014) 1–10.
[42] C. Ivancsits, M. Lee, Visual navigation system for small unmanned aerial vehicles, Sens. Rev. 33 (2013) 267–291.
[43] S. Durbha, R. King, V. Shah, N. Younan, Image information mining for coastal disaster management, in: Proceedings of IGARSS, 2007, pp. 342–345.
[44] G. Panteras, G. Cervone, Enhancing the temporal resolution of satellite-based flood extent generation using crowdsourced data for disaster monitoring, Int J Remote Sens 39 (5) (2016) 1459–1474.
[45] F. Dell'Acqua, P. Gamba, Query-by-shape in meteorological image archives using the point diffusion technique, IEEE Trans Geosci Remote Sens 39 (9) (2001) 1834–1843.
[46] S. Rivest, Y. Bedard, M. Proulx, M. Nadeau, F. Hubert, J. Pastor, SOLAP technology: merging business intelligence with geospatial technology for

interactive spatio-temporal exploration and analysis of data, ISPRS J. Photogramm. Remote Sens. 60 (1) (2005) 17–33.
[47] J. Leeuw, A. Vrieling, A. Shee, C. Atzberger, K. Hadgu, C. Biradar, H. Keah, C. Turvey, The potential and uptake of remote sensing in insurance: a review, Remote Sens (Basel) 6 (11) (2014) 10888–10912.
[48] O. Reichman, M. Jones, M. Schildhauer, Challenges and opportunities of open data in ecology, Science 331 (6018) (2011) 703–705.
[49] V. Gudivada, V. Raghavan, Content-based image retrieval systems – guest editors' introduction, IEEE Comput 28 (9) (1995) 18–22.
[50] A. Smeulders, T. Huang, T. Gevers, Special issue on content-based image retrieval, Int J Comput Vis 56 (1) (2004) 5–6.
[51] A. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Trans Pattern Anal Mach Intell 22 (12) (2000) 1349–1380.
[52] D. Zhang, G. Lu, Content-based shape retrieval using different shape descriptors: a comparative study, in: Proceedings of ICME, 2001, pp. 1139–1142.
[53] N. Sebe, M. Lew, X. Zhou, T. Huang, E. Bakker, The state of the art in image and video retrieval, in: Proceedings of International Conference on Image Video Retrieval, 2003, pp. 1–8.
[54] M. Lew, N. Sebe, C. Djeraba, R. Jain, Content-based multimedia information retrieval: state of the art and challenges, ACM Trans. Multime. Comput Commun Appl 2 (1) (2006) 1–19.
[55] R. Datta, D. Joshi, J. Li, J. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Comput Surv 40 (2) (2008) 1–60.
[56] M. Datcu, S. D'Elia, R. King, L. Bruzzone, Introduction to the special section on image information mining for Earth observation data, IEEE Trans. Geosci. Remote Sens 45 (4) (2007) 795–798.
[57] M. Datcu, R. King, S. D'Elia, Introduction to the special issue on image information mining: pursuing automation of geospatial intelligence for environment and security, IEEE Geosci. Remote Sens Lett 7 (1) (2010) 3–6.
[58] S. Newsam, Y. Yang, Comparing global and interest point descriptors for similarity retrieval in remote sensed imagery, in: Proceedings of ACM-GIS, 2007.
[59] Q. Bao, P. Guo, Comparative studies on similarity measures for remote sensing image retrieval, in: : Proceedings of the IEEE International Conference on Systems, Man & Cybernetics, 2004.
[60] P. Du, Y. Chen, H. Tang, T. Fang, Study on content-based remote sensing image retrieval, in: Proceedings of IGARSS, 2008, pp. 707–710.
[61] Y. Li, S. Yang, T. Liu, X. Dong, Comparative assessment of semantic-sensitive satellite image retrieval: simple and context-sensitive Bayesian networks, Int. J. Geogr. Inf Sci 26 (2) (2012) 247–263.
[62] M. Quartulli, I. Olaizola, A review of EO image information mining, ISPRS J Photogramm Remote Sens 75 (2013) 11–28.
[63] S. Ozkan, T. Ates, E. Tola, M. Soysal, E. Esen, Performance analysis of state-of-the-art representation methods for geographical image retrieval and categorization, IEEE Geosci. Remote Sens Lett 11 (11) (2014) 1996–2000.
[64] G. Xia, X. Tong, F. Hu, Y. Zhong, M. Datcu, and L. Zhang. Exploiting deep features for remote sensing image retrieval: a systematic investigation. arXiv. 2017, arXiv: 1707.07321.
[65] W. Zhou, S. Newsam, C. Li, Z. Shao, PatternNet: a benchmark dataset for performance evaluation of remote sensing image retrieval, ISPRS J. Photogramm Remote Sens. 145 (2018) 197–209.
[66] Y. Gu, Y. Wang, Yansheng Li, A survey on deep learning-driven remote sensing image scene understanding: scene classification, scene retrieval and scene-guided object detection, Appl. Sci. 9 (10) (2019) 2110.
[67] S. Sudha, S. Aji, A review on recent advances in remote sensing image retrieval techniques, J Indian Soc. Remote Sens (2019).
[68] K. Seidel, R. Mastropietro, M. Datcu, New architectures for remote sensing image archives, in: Proceedings of IGARSS, 1997.
[69] C. Chang, B. Moon, A. Acharya, C. Shock, A. Sussman, J. Saltz, Titan: a high-performance remote sensing databases, in: Proceedings of the 13th International Conference on Data Engineering, 1997, pp. 375–384.
[70] J. Hurwitz, et al., Big Data For Dummies, 1st ed., Wiley, New York, NY, USA, 2013.
[71] R. Rajak, D. Raveendran, M. Bh, S. Medasani, High resolution satellite image processing using hadoop framework, in: InL Proceedings of IEEE International Conference on Cloud Computing in Emerging Markets, 2015, pp. 16–21.
[72] W. Huang, L. Meng, D. Zhang, W. Zhang, In-memory parallel processing of massive remotely sensed data using an apache spark on hadoop yarn model, IEEE J. Sel. Topics in Appl. Earth Obs. Remote Sens. 10 (1) (2017) 3–19.
[73] "Apache hadoop, " http://hadoop.apache.org/. Available.
[74] J. Li, L. Zhu, F. Cao, Remote sensing image segmentation based on Hadoop cloud platform, in: Proceedings of SPIE International Conference on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology, 2018.
[75] J. Dean, S. Ghemawat, Mapreduce: a flexible data processing tool, Commun. ACM 53 (2010) 72–77.
[76] H. Shi, G. Hu, J. Cao, M. Wang, Y. Tian, A new approach for large-scale scene image retrieval based on improved parallel k-means algorithm in mapreduce environment, Math. Probl. Eng (2016).
[77] W. Jing, S. Huo, Q. Miao, X. Chen, A model of parallel mosaicking for massive remote sensing images based on spark, IEEE Access 5 (2017) 18229–18237.
[78] E. Lindholm, J. Nickolls, S. Oberman, J. Montrym, NVIDIA Tesla: a unified graphics and computing architecture, IEEE Micro 28 (2008) 39–55.
[79] G. Hinton, R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[80] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[81] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, Deep learning and process understanding for data-driven earth system science, Nature 566 (7743) (2019) 195–264.
[82] Yansheng Li, Y. Zhang, X. Huang, A. Yuille, Deep networks under scene-level supervision for multi-class geospatial object detection from remote sensing images, ISPRS J Photogramm. Remote Sens. 146 (2018) 182–196.
[83] C. Tao, L. Mi, Yansheng Li, J. Qi, Y. Xiao, J. Zhang, Scene context-driven vehicle detection in high-resolution aerial images, IEEE Trans. Geosci. Remote Sens 57 (10) (2019) 7339–7351.
[84] Y. Tan, S. Xiong, Yansheng Li, Automatic extraction of built-up areas from panchromatic and multispectral remote sensing images using double-stream deep convolutional neural networks, IEEE J. Sel. Topics in Appl. Earth Obs. Remote Sensing 11 (11) (2018) 3988–4004.
[85] Yansheng Li, et al., Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning, Remote Sens Environ 250 (2020).
[86] X. Liu, Q. Liu, Y. Wang, Remote sensing image fusion based on two-stream fusion network, Inf. Fusion 55 (2020) 1–15.
[87] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Inf Fusion 36 (2017) 191–207.
[88] Y. Liu, X. Chen, Z. Wang, Z. Wang, R. Ward, X. Wang, Deep learning for pixel-level image fusion: recent advances and future prospects, Inf Fusion 42 (2018) 158–173.
[89] Jiayi Ma, et al., FusionGAN: a generative adversarial network for infrared and visible image fusion, Inf Fusion 48 (2019) 11–26.
[90] Jiayi Ma, et al., Infrared and visible image fusion via detail preserving adversarial learning, Inf Fusion 54 (2020) 85–98.
[91] Yansheng Li, Y. Zhang, Z. Zhu, Error-tolerant deep learning for remote sensing image scene classification, IEEE Trans Cybern (2020) in press.
[92] Z. Gong, P. Zhong, Y. Yu, W. Hu, Diversity-promoting deep structural metric learning for remote sensing scene classification, IEEE Trans Geoscie. Remote Sens 56 (1) (2018) 371–390.
[93] N. He, L. Fang, S. Li, A. Plaza, J. Plaza, Remote sensing scene classification using multilayer stacked covariance pooling, IEEE Trans. Geosci. Remote Sens 56 (12) (2018) 6899–6910.
[94] W. Han, R. Feng, L. Wang, L. Gao, Adaptive spatial-scale-aware deep convolutional neural network for high-resolution remote sensing imagery scene classification, in: Proceedings of IGARSS, 2018.
[95] J. Xie, N. He, L. Fang, A. Plaza, Scale-free convolutional neural network for remote sensing scene classification, IEEE Trans. Geosci Remote Sens. 57 (9) (2019) 6916–6928.
[96] R. Larson, Introduction to information retrieval, J. Am. Soc. Inf. Sci. 61 (2010) 852–853.
[97] M. Wolfmuller, D. Dietrich, E. Sireteanu, S. Kiemle, E. Mikusch, M. Bottcher, Data flow and workflow organization-The data management for the TerraSAR-X payload ground segment, IEEE Trans Geosci. Remote Sens 47 (1) (2009) 44–50.
[98] X. Wang, N. Chen, Z. Chen, X. Yang, J. Li, Earth observation metadata ontology model for spatiotemporal-spectral semantic-enhanced satellite observation discovery: a case study of soil moisture monitoring, GIsci Remote Sens 53 (1) (2016) 22–44.
[99] P. Angelov, P. Sadeghi-Tehran, A nested hierarchy of dynamically evolving clouds for big data structuring and searching, Procedia Comput Sci 53 (2015) 1–8.
[100] S. Newsam, L. Wang, S. Bhagavathy, B. Manjunath, Using texture to analyze and manage large collections of remote sensed image and video data, Appl Opt 43 (2004) 210–217.
[101] B. Luo, S. Jiang, L. Zhang, Indexing of remote sensing images with different resolutions by multiple features, IEEE J Sel Topics in Appl Earth Obs. Remote Sens 6 (4) (2013) 1899–1912.
[102] G. Healy, A. Jain, Retrieval multispectral satellite images using physics-based invariant representations, IEEE Trans Pattern Anal Artif Intell 18 (1996) 34–47.
[103] L. Jiao, X. Tang, B. Hou, S. Wang, SAR images retrieval based on semantic classification and region-based similarity measure for earth observation, IEEE J Sel Topics in Appl Earth Obs Remote Sens 8 (2015) 3876–3891.
[104] X. Tang, L. Jiao, W. Emery, SAR image retrieval based on fuzzy similarity and relevance feedback, IEEE J Sel Topics in Appl Earth Obs Remote Sens 10 (2017) 1824–1842.
[105] X. Tang, L. Jiao, Fusion similarity-based reranking for SAR image retrieval, IEEE Geosci. Remote Sens Lett 55 (2017) 5798–5817.
[106] B. Hou, X. Tang, L. Jiao, S. Wang, SAR image retrieval based on Gaussian mixture model classification, in: Proceedings of APSAR, 2009, pp. 796–799.
[107] X. Tang, L. Jiao, W. Emery, F. Liu, D. Zhang, Two-stage reranking for remote sensing image retrieval, IEEE Trans Geosci Remote Sens 55 (2017) 5798–5817.
[108] F. Ye, W. Luo, M. Dong, H. He, W. Min, SAR image retrieval based on unsupervised domain adaptation and clustering, IEEE Geosci Remote Sens Lett. (2019).
[109] F. Omruuzun, B. Demir, L. Bruzzone, Y. Cetin, Content Based Hyperspectral Image Retrieval Using Bag Of Endmembers Image Descriptors, in: Proceedings of WHISPERS, 2016.
[110] J. Zhang, W. Geng, X. Liang, J. Li, L. Zhuo, Q. Zhou, Hyperspectral remote sensing image retrieval system using spectral and texture features, Appl Opt 56 (16) (2017) 4785–4796.
[111] O. Ben-Ahmed, T. Urruty, N. Richard, C. Fernandez-Maloigne, Toward content-based hyperspectral remote sensing image retrieval (CB-HRSIR): a preliminary study based on spectral sensitivity functions, Remote Sens (Basel) 11 (5) (2019) 600.

[112] J. Zhang, C. Lu, X. Liang, L. Zhuo, Q. Tian, Hyperspectral image secure retrieval based on encrypted deep spectral-spatial features, J Appl Remote Sens 13 (1) (2019), 018501.
[113] I. Alber, Z. Xiong, N. Yeager, M. Farber, W. Pottenger, Fast retrieval of multi- and hyper-spectral images using relevance feedback, in: Proceedings of IGARSS, 2001.
[114] A. Plaza, J. Plaza, A. Paz, Parallel heterogeneous CBIR system for efficient hyperspectral image retrieval using spectral mixture analysis, Concurr Comput.: Pract. Exp. 22 (9) (2010) 1138–1159.
[115] M. Grana, M. Veganzones, An endmember-based distance for content based hyperspectral image retrieval, Pattern Recognit 45 (9) (2012) 3472–3498.
[116] M. Veganzones, M. Datcu, M. Grana, Dictionary based hyperspectral image retrieval, in: Proceedings of ICPRAM, 2012, pp. 426–432.
[117] J. Zhang, L. Chen, L. Zhuo, X. Liang, J. Li, An efficient hyperspectral image retrieval method: deep spectral-spatial feature extraction with DCGAN and dimensionality reduction using t-SNE-based NM hashing, Remote Sens (Basel) 10 (2) (2018) 271.
[118] F. Bovolo, B. Demir, L. Bruzzone, A Cluster-Based Approach to Content Based Time Series Retrieval, in: Proceedings of IGARSS, 2015.
[119] A. Julea, N. Meger, P. Bolon, C. Rigotti, M. Doin, C. Lasserre, E. Trouve, V. Lazarescu, Unsupervised spatiotemporal mining of satellite image time series using grouped frequent sequential patterns, IEEE Trans Geosci Remtoe Sens 49 (4) (2011) 1417–1430.
[120] L. Gueguen, M. Datcu, A similarity metric for retrieval of compressed objects: application for mining satellite image time series, IEEE Trans Knowl Data Eng 20 (4) (2008) 562–575.
[121] L. Gueguen, M. Datcu, Image time-series data mining based on the information-bottleneck principle, IEEE Trans Geosci Remote Sens 45 (4) (2007) 827–838.
[122] F. Petitjean, J. Inglada, P. Gancarski, Satellite image time series analysis under time warping, IEEE Trans Geosci Remote Sen 50 (8) (2012) 3081–3095.
[123] T. Bretschneider, R. Cavet, O. Kao, Retrieval of remotely sensed imagery using spectral information content, in: Proceedings of IGARSS, 2002, pp. 2253–2255.
[124] T. Bretschneider, O. Kao, A retrieval system for remotely sensed imagery, in: Proceedings of International Conference on Imaging Science, Systems, and Technology, 2002.
[125] A. Vellaikal, C. Kuo, S. Dao, Content-based retrieval of remote sensed images using vector quantization, in: Proceedings of SPIE Visual Information Processing, 1995, pp. 178–189.
[126] G. Healey, A. Jain, Retrieving multispectral satellite images using physics-based invariant representations, IEEE Trans Pattern Anal Mach Intell 18 (8) (1995) 842–848.
[127] R.M. Haralick, K.S. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Systems, Man, and Cybernetics 3 (6) (1973) 610–621.
[128] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693.
[129] J.G. Daugman, Complete discrete 2-d gabor transforms by neural networks for image analysis and compression, IEEE Trans. Acoustics, Speech, and Signal Processing 36 (7) (1988) 1169–1179.
[130] M. Pietikainen, T. Ojala, Z. Xu, Rotation-invariant texture classification using feature distributions, Pattern Recognit 33 (1) (2000) 43–52.
[131] I. Tekeste, B. Demir, Advanced Local Binary Patterns for Remote Sensing Image Retrieval, in: Proceedings of IGARSS, 2018.
[132] B. Luo, J.F. Aujol, Y. Gousseau, S. Ladjal, Indexing of satellite images with different resolutions by wavelet features, IEEE Trans Image Proc. 17 (2008) 1465–1472.
[133] Y. Hongyu, L. Bicheng, C. Wen, Remote sensing imagery retrieval based-on gabor texture feature classification, in: Proceedings of ICSP, 2004.
[134] S. Newsam, L. Wang, S. Bhagavathy, B. Manjunath, Using texture to analyze and manage large collections of remote sensed image and video data, IEEE Trans Pattern Anal Mach Intell 18 (8) (1996) 837–842.
[135] V. Shah, S. Durbha, N. Younan, R. King, Coalescing ICA and wavelets coefficients for image information mining in Earth observation data archives, in: Proceedings of IGARSS, 2006, pp. 9–12.
[136] V. Shah, N. Younan, S. Durbha, R. King, Wavelet features for information mining in remote sensing archives, in: Proceedings of IGARSS, 2005, pp. 5630–5633.
[137] V. Shah, N. Younan, S. Durbha, R. King, A systematic approach to wavelet-decomposition-level selection for image information mining from geospatial data archives, IEEE Trans. Geosci. Remote Sens 45 (4) (2007) 875–878.
[138] V. Shah, S. Durbha, N. Younan, R. King, A wavelet-based approach for knowledge mining in earth observation data archives, in: Proceedings of ESA-EUSC, 2005.
[139] Z. Shao, W. Zhou, L. Zhang, J. Hou, Improved color texture descriptors for remote sensing image retrieval, J Appl Remote Sens 8 (1) (2014), 083584.
[140] S. Bouteldja, A. Kourgli, Multiscale texture features for the retrieval of high resolution satellite images, in: Proceedings of IWSSIP, 2015.
[141] P. Maheshwary, N. Srivastava, Prototype system for retrieval of remote sensing images based on color moment and gray level co-occurrence matrix, Int. J Comput Sci Issues 3 (2009) 20–23.
[142] K. Sukhia, M. Riaz, A. Ghafoor, S. Ali, Content-based remote sensing image retrieval using multi-scale local ternary pattern, Digit Signal Process 104 (2020), 102765.
[143] Byju, B. Demir, L. Bruzzone, A progressive content-based image retrieval in JPEG 2000 compressed remote sensing archives, IEEE Trans Geosci.Remote Sens (2020).
[144] A. Ma, I. Sethi, Local shape association based retrieval of infrared satellite images, in: Proceedings of ISM, 2005.
[145] P. Agouris, J. Carswell, A. Stefanidis, An environment for content-based image retrieval from large spatial databases, ISPRS J Photogramm Remote Sens 54 (4) (1999) 263–272.
[146] G. Scott, M. Klaric, C. Davis, C. Shyu, Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases, IEEE Trans Geosci Remote Sens 49 (2011) 1603–1616.
[147] X. Wang, Z. Shao, X. Zhou, J. Liu, A novel remote sensing image retrieval method based on visual salient point features, Sensor Review 34 (4) (2014) 349–359.
[148] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int J Comput Vis 60 (2) (2004) 91–110.
[149] G. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, H. Maitre, Structural high-resolution satellite image indexing, in: Proceedings of ISPRS TC VII Symposium – 100 Years ISPRS, 2010, pp. 298–303.
[150] K. Tobin, B. Bhaduri, E. Bright, A. Cheriyadat, T. Karnowaski, P. Palathingal, T. Potok, J. Price, Large-scale geospatial indexing for image-based retrieval and analysis, In: ISVC (2005).
[151] G. Marchisio, J. Cornelison, Content-based search and clustering of remote sensing imagery, in: Proceedings of IGARSS, 1999.
[152] K. Koperski, G. Marchisio, Multi-level indexing and GIS enhanced learning for satellite imageries, in: Proceedings of the International Workshop on Multimedia Data Mining, 2000.
[153] J. Li, R. Narayanan, Integrated information mining and image retrieval in remote sensing, IEEE Trans Geosci Remote Sens 42 (3) (2004) 673–685.
[154] Y. Li, T. Bretschneider, Semantic-sensitive satellite image retrieval, IEEE Trans Geosci Remote Sens 45 (4) (2007) 853–860.
[155] A. Samal, S. Bhatia, P. Vadlamani, D. Marx, Searching satellite imagery with integrated measures, Pattern Recognit 42 (11) (2009) 2502–2513.
[156] H. Sebai, A. Kourgli, A. Serir, Dual-tree complex wavelet transform applied on color descriptors for remote-sensed images retrieval, J Appl Remote Sens 9 (1) (2015), 095994.
[157] M. Wang, T. Song, Remote sensing image retrieval by scene semantic matching, IEEE Trans Geosci Remote Sens 51 (2013) 2874–2886.
[158] J. Sivic, A. Zisserman, Video google: a text retrieval approach to object matching in videos, in: Proceedings of ICCV, 2003.
[159] H. Jegou, M. Douze, C. Schmid, P. Perez, Aggregating local descriptors into a compact image representation, in: Proceedings of CVPR, 2010.
[160] W. Zhou, Z. Shao, C. Diao, Q. Cheng, High-resolution remote-sensing imagery retrieval using sparse features by auto-encoder, Remote Sens Lett. 6 (2015) 775–783.
[161] C. Ma, F. Chen, J. Yang, J. Liu, W. Xia, X. Li, A remote-sensing image-retrieval model based on an ensemble neural networks, Big Earth Data 2 (4) (2018) 351–367.
[162] Y. Yang, S. Newsam, Geographic image retrieval using local invariant features, IEEE Trans Geosci. Remote Sens 51 (2) (2012) 818–832.
[163] E. Aptoula, Remote sensing image retrieval with global morphological texture descriptors, IEEE Trans Geosci Remote Sens. 52 (2014) 3023–3034.
[164] J. Yang, J. Liu, Q. Dai, An improved bag-of-words framework for remote sensing image retrieval in large-scale image databases, Int J Digit Earth 8 (4) (2015) 273–292.
[165] X. Tang, X. Zhang, F. Liu, L. Jiao, Unsupervised deep feature learning for remote sensing image retrieval, Remote Sens (Basel) 10 (8) (2018) 1243.
[166] P. Maragos, Pattern spectrum and multiscale shape representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 701–716.
[167] P. Bosilj, E. Aptoula, S. Lefevre, E. Kijak, Retrieval of remote sensing images with pattern spectra descriptors, ISPRS Int J Geoinf 5 (12) (2016) 228.
[168] J. Yang, K. Yu, Y. Gong, T.S. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: Proceedings of CVPR, 2009.
[169] Y. Wang, L. Zhang, X. Tong, L. Zhang, Z. Zhang, H. Liu, A three-layered graph-based learning approach for remote sensing image retrieval, IEEE Trans Geosci Remote Sens 54 (2016) 6020–6034.
[170] R. Imbriaco, C. Sebastian, E. Bondarev, P. deWith, Aggregated deep local features for remote sensing image retrieval, Remote Sens (Basel) 11 (5) (2019) 493.
[171] Z. Du, X. Li, X. Lu, Local structure learning in high resolution remote sensing image retrieval, Neurocomputing 207 (2016) 813–822.
[172] Yansheng Li, C. Tao, Y. Tan, K. Shang, J. Tian, Unsupervised multilayer feature learning for satellite image scene classification, IEEE Geosci Remote Sens Lett. 13 (2) (2016) 157–161.
[173] Yansheng Li, X. Huang, H. Liu, Unsupervised deep feature learning for urban village detection from high-resolution remote sensing images, Photogramm. Eng Remote Sens 83 (8) (2017) 567–579.
[174] Yansheng Li, Y. Zhang, C. Tao, H. Zhu, Content-based high-resolution remote sensing image retrieval via unsupervised feature learning and collaborative affinity metric fusion, Remote Sens (Basel) 8 (9) (2016) 709.
[175] P. Napoletano, Visual descriptors for content-based retrieval of remote sensing images, Int J Remote Sens 39 (2018) 1343–1376.
[176] W. Xiong, Y. Lv, Y. Cui, X. Zhang, X. Gu, A discriminative feature learning approach for remote sensing image retrieval, Remote Sens (Basel) 11 (3) (2019) 281.
[177] U. Chaudhuri, B. Banerjee, A. Bhattacharya, Siamese graph convolutional network for content based remote sensing image retrieval, Comput Vis. Image Underst 184 (2019) 22–30.
[178] R. Cao, Q. Zhang, J. Zhu, Q. Li, Q. Li, B. Liu, G. Qiu, Enhancing remote sensing image retrieval using a triplet deep metric learning network, Int J Remote Sens (2019).
[179] P. Liu, G. Gou, X. Shan, D. Tao, Q. Zhou, Global optimal structured embedding learning for remote sensing image retrieval, Sensors 20 (1) (2020) 291.

[180] L. Fan, H. Zhao, H. Zhao, Distribution consistency loss for large-scale remote sensing image retrieval, Remote Sens (Basel) 12 (1) (2020) 175.
[181] Y. Liu, Z. Han, C. Chen, L. Ding, Y. Liu, Eagle-eyed multitask CNNs for aerial image retrieval and scene classification, IEEE Trans Geosci Remote Sens (2020).
[182] J. Fu, et al., Dual attention network for scene segmentation, in: Proceedings of CVPR, 2019, pp. 3146–3154.
[183] Z. Huang, et al., CCNet: criss-Cross attention for semantic segmentation, IEEE Trans Pattern Anal Mach Intell (2020).
[184] M. Datcu, K. Seidel, Human-centered concepts for exploration and understanding of earth observation images, IEEE Trans Geosci Remote Sens. 43 (3) (2005) 601–609.
[185] Y. Liu, D. Zhang, G. Lu, Region-based image retrieval with high-level semantics using decision tree learning, Pattern Recognit 41 (2008) 2554–2570.
[186] M. Datcu, et al., Information mining in remote sensing image archives: system concepts, IEEE Trans. Geosci. Remote Sens. 41 (12) (2003) 2923–2936.
[187] I. Munoz, M. Datcu, System design considerations for image information mining in large archives, IEEE Geosci Remote Sens Lett. 7 (1) (2010) 13–17.
[188] M. Wang, Q. Wan, L. Gu, T. Song, Remote-sensing image retrieval by combining image visual and semantic features, Int J Remote Sens 34 (12) (2013) 4200–4223.
[189] S. Durbha, R. King, Semantics-enabled framework for knowledge discovery from earth observation data archives, IEEE Trans. Geosci Remote Sens. 43 (11) (2005) 2563–2572.
[190] N. Ruan, N. Huang, W. Hong, Semantic-based image retrieval in remote sensing archive: an ontology approach, in: Proceedings of IGARSS, 2006.
[191] K. Tobin, B. Bhaduri, E. Bright, A. Cheriyadat, T. Karnowski, P. Palathingal, T. Potok, J. Price, Automated feature generation in large-scale geospatial libraries for content-based indexing, Photogramm. Eng. Remote Sens. 72 (5) (2006) 531–540.
[192] S. Kalluri, et al., Hierarchical data archiving and processing system to generate custom tailored products from AVHRR data, in: Proceedings of IGARSS, 1999, pp. 2374–2376.
[193] C. Shyu, M. Klaric, G.J. Scott, A. Barb, C. Davis, K. Palaniappan, GeoIRIS: geospatial Information Retrieval and Indexing System—Content mining, semantics modeling, and complex queries, IEEE Trans. Geosci. Remote Sens. 45 (2007) 839–852.
[194] C. Ghirardini, S.A. Chun, V. Atluri, I. Kamel, N.R. Adam, A study on the indexing of satellite images at NASA regional application center, database and expert systems applications, in: Proceedings of 12th Int. Workshop Database Expert Syst. Appl., 2001, pp. 859–864.
[195] J. Bentley, Multidimensional binary search trees used for associative searching, Commun. ACM. 18 (9) (1975) 509–517.
[196] M. Muja, D. Lowe, Fast approximate nearest neighbors with automatic algorithm configuration, in: Proceedings of VISAPP Int. Conf. Comput. Vis. Theory Appl., 2009, pp. 331–340.
[197] H. Jegou, M. Douze, C. Schmid, Improving bag-of-features for large scale image search, Int J Comput Vis (2010).
[198] F. Jiang, H. Hu, J. Zheng, B. Li, A hierarchal BoW for image retrieval by enhancing feature salience, Neurocomputing 175 (2016) 146–154.
[199] D. Dister, H. Stewenius, Scalable recognition with a vocabulary tree, in: Proceedings of IEEE CVPR, 2006.
[200] P. Sadeghi-Tehran, P. Angelov, N. Virlet, M. Hawkesford, Scalable database indexing and fast image retrieval based on deep learning and hierarchically nested structure applied to remote sensing and plant biology, J Imaging 5 (2019) 33.
[201] X. Zhu, L. Zhang, Z. Huang, A Sparse embedding and least variance encoding approach to hashing, IEEE Trans. Image Process. 23 (9) (2014) 3737–3750.
[202] J. Wang, S. Kumar, S.-.F. Chang, Semi-supervised hashing for large-scale search, IEEE Trans. Pattern Anal. Mach. Intell. 34 (12) (2012) 2393–2406.
[203] J. Lee, R. Jin, A. Jain, Rank-based distance metric learning: an application to image retrieval, in: Proceedings of IEEE CVPR, 2008.
[204] B. Chaudhuri, B. Demir, L. Bruzzone, S. Chaudhuri, Region-Based Retrieval of Remote Sensing Images using an Unsupervised Graph-Theoretic Approach, IEEE Geosci. Remote Sens Lett. 13 (7) (2016) 987–991.
[205] Y. Liu, L. Ding, C. Chen, Y. Liu, Similarity-based unsupervised deep transfer learning for remote sensing image retrieval, IEEE Trans. Geosci. Remote Sens. (2020).
[206] Y. Cao, et al., DML-GANR: deep metric learning with generative adversarial network regularization for high spatial resolution remote sensing image retrieval, IEEE Trans. Geosci. Remote Sens. (2020) in press.
[207] Y. Liu, et al., High-resolution remote sensing image retrieval based on classification-similarity networks and double fusion, IEEE J Sel Topics in Appl Earth Obs. Remote Sens. 13 (2020) 1119–1133.
[208] Y. Liu, et al., Remote-sensing image retrieval with tree-triplet-classification networks, Neurocomputing 405 (2020) 48–61.
[209] W. Zhou, S. Newsam, C. Li, Z. Shao, Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval, Remote Sens (Basel) 9 (2017) 489–508.
[210] T. Mukhtar, N. Khurshid, M. Taj, Dimensionality reduction using discriminative autoencoders for remote sensing image retrieval, in: Proceedings of International Conference on Image Analysis and Processing, 2019.
[211] Y. Wang, S. Ji, M. Lu, Y. Zhang, Attention boosted bilinear pooling for remote sensing image retrieval, Int J Remote Sens 41 (7) (2019) 2704–2724.
[212] J. Wang, W. Liu, S. Kumar, S. Chang, Learning to hash for indexing big data-A survey, Proc. IEEE 104 (1) (2016) 34–57.
[213] Y. Gong, S. Lazebnik, A. Gordo, F. Perronnin, Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval, IEEE Trans Pattern Anal Mach Intell 35 (12) (2013) 2916–2929.
[214] M. Slaney, M. Casey, Locality-sensitive hashing for finding nearest neighbors, IEEE Signal Process Mag 25 (2) (2008) 128–131.
[215] B. Kulis, K. Grauman, Kernelized locality-sensitive hashing, IEEE Trans. Pattern Anal. Mach. Intelligence 34 (6) (2012) 1092–1104.
[216] M. Raginsky, S. Lazebnik, Locality-sensitive binary codes from shift-invariant kernels, in: Proceedings of Int. Conf. Comput. Vis, 2009, pp. 1509–1517.
[217] W. Liu, J. Wang, S. Kumar, S.-.F. Chang, Hashing with graphs, in: Proceedings of Int. Conf. Mach. Learn, 2011, pp. 1–8.
[218] B. Kulis, T. Darrell, Learning to hash with binary reconstructive embeddings, in: Proceedings of Adv. Neural Inf. Process. Syst, 2009, pp. 1042–1050.
[219] J.-.P. Heo, Y. Lee, J. He, S.-.F. Chang, S.-.E. Yoon, Spherical hashing, in: Proceedings of Conf. Comput. Vis. Pattern Recog, 2012, pp. 2957–2964.
[220] Y. Weiss, A. Torralba, R. Fergus, Spectral hashing, in: Proceedings of NIPS, 2008, pp. 1753–1760.
[221] F. Shen, et al., Hashing on nonlinear manifolds, IEEE Trans. Image Process. 24 (6) (2015) 1839–1851.
[222] L. Liu, M. Yu, L. Shao, Multiview alignment hashing for efficient image search, IEEE Trans. Image Process. 24 (3) (2015) 956–966.
[223] J. Tang, Z. Li, M. Wang, R. Zhao, Neighborhood discriminant hashing for large-scale image retrieval, IEEE Trans. Image Process.. 24 (9) (2015) 2827–2840.
[224] H. Zhang, L. Liu, Y. Long, L. Shao, Unsupervised deep hashing with pseudo labels for scalable image retrieval, IEEE Trans Image Processing 27 (4) (2018) 1626–1638.
[225] B. Demir, L. Bruzzone, Hashing-based scalable remote sensing image search and retrieval in large archieves, IEEE Trans Geosci. Remote Sens 54 (2) (2016) 892–904.
[226] P. Li, P. Ren, Partial randomness hashing for large-scale remote sensing image retrieval, IEEE Geosci. Remote Sens Lett 14 (3) (2017) 464–468.
[227] T. Reato, B. Demir, L. Bruzzone, An unsupervised multicode hashing method for accurate and scalable remote sensing image retrieval, IEEE Geosci. Remote Sens. Lett. 16 (2) (2019) 276–280.
[228] R. Fernandez-Beltran, B. Demir, F. Pla, A. Plaza, Unsupervised remote sensing image retrieval using probabilistic latent semantic hashing, IEEE Geosci. Remote Sens Lett. (2020) in press.
[229] P. Li, X. Zhang, X. Zhu, P. Ren, Online hashing for scalable remote sensing image retrieval, Remote Sens (Basel) 10 (5) (2018) 709.
[230] B. Kulis, T. Darrell, Learning to hash with binary reconstructive embeddings, in: Proceedings of Adv. Neural Inf. Process. Syst, 2009, pp. 1042–1050.
[231] M. Norouzi, D.M. Blei, Minimal loss hashing for compact binary codes, in: Proceedings of Int. Conf. Mach. Learn, 2011, pp. 353–360.
[232] X. Zhu, L. Zhang, Z. Huang, A sparse embedding and least variance encoding approach to hashing, IEEE Trans. Image Process. 23 (9) (2014) 3737–3750.
[233] N. Luka, B. Zalik, S. Cui, M. Datcu, GPU-based kernelized locality-sensitive hashing for satellite image retrieval, in: Proceedings of IGARSS, 2015, pp. 1468–1471.
[234] D. Ye, Yansheng Li, C. Tao, X. Xie, X. Wang, Multiple feature hashing learning for large-scale remote sensing image retrieval, ISPRS Int J Geoinf 6 (11) (2017) 364.
[235] T. Reato, B. Demir, L. Bruzzone, Primitive Cluster Sensitive Hashing for Scalable Content-Based Image Retrieval in Remote Sensing Archives, in: Proceedings of IGARSS, 2017.
[236] T. Reato, B. Demir, L. Bruzzone, A Novel Class Sensitive Hashing Technique for Large-Scale Content-Based Remote Sensing Image Retrieval, in: Proceedings of SPIE Image and Signal Processing for Remote Sensing, 2017.
[237] J. Kong, Q. Sun, M. Mukherjee, J. Lloret, Low-rank hypergraph hashing for large-scale remote sensing image retrieval, Remote Sens (Basel) 12 (2020) 1164.
[238] L. Han, P. Li, X. Bai, C. Grecos, X. Zhang, P. Ren, Cohesion intensive deep hashing for remote sensing image retrieval, Remote Sens (Basel) 12 (1) (2020) 101.
[239] C. Zou, S. Wan, P. Jin, X. Li, A novel rotation invariance hashing network for fast remote sensing image retrieval, in: Proceedings of ICDIP, 2018.
[240] Yansheng Li, Y. Zhang, X. Huang, H. Zhu, J. Ma, Large-scale remote sensing image retrieval by deep hashing neural networks, IEEE Trans. Geosci. Remote Sens 56 (2) (2018) 950–965.
[241] S. Roy, E. Sangineto, B. Demir, N. Sebe, Deep metric and hash-code learning for content-based retrieval of remote sensing images, in: Proceedings of IGARSS, 2018, pp. 4539–4542.
[242] W. Song, S. Li, and J. Benediktsson. Deep hashing learning for visual and semantic retrieval of remote sensing images. arXiv, 2019, arXiv:1909.04614v1.
[243] C. Liu, J. Ma, X. Tang, X. Zhang, L. Jiao, Adversarial hash-code learning for remote sensing image retrieval, in: Proceedings of IGARSS, 2019, pp. 4324–4327.
[244] X. Tang, C. Liu, J. Ma, X. Zhang, F. Liu, L. Jiao, Large-scale remote sensing image retrieval based on semi-supervised adversarial hashing, Remote Sens (Basel) 11 (17) (2019) 2055.
[245] S. Roy, E. Sangineto, B. Demir, N. Sebe, Metric-learning-based deep hashing network for content-based retrieval of remote sensing images, IEEE Geosci. Remote Sens Lett. (2020) in press.
[246] Y. Li, et al., Two birds, one stone: jointly learning binary code for large-scale face image retrieval and attributes prediction, in: Proceedings of ICCV, 2015, pp. 3819–3827.
[247] P. Li, L. Han, X. Tao, X. Zhang, C. Grecos, A. Plaza, P. Ren, Hashing nets for hashing: a quantized deep learning to hash framework for remote sensing image retrieval, IEEE Transactions on Geoscience and Remote Sensing (2020).
[248] G. Chen, et al., Training small networks for scene classification of remote sensing images via knowledge distillation, Remote Sens (Basel) 10 (2018) 5.

[249] B. Zhang, Y. Zhang, S. Wang, A lightweight and discriminative model for remote sensing scene classification with multidilation pooling module, IEEE J Sel Topics in Appl Earth Obs Remote Sens 12 (2019) 2636–2653.
[250] H. Li, C. Tao, Z. Wu, J. Chen, J. Gong, M. Deng, RSI-CB: a large-scale remote sensing image classification benchmark via crowdsource data, arXiv (2017) arXiv: 1705.10450.
[251] D. Hou, Z. Miao, H. Xing, H. Wu, V-RSIR: an open access web-based image annotation tool for remote sensing image retrieval, IEEE Access 7 (2019) 83852–83862.
[252] Yansheng Li, Y. Zhang, X. Huang, J. Ma, Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval, IEEE Trans Geosci Remote Sens 56 (11) (2018) 6521–6536.
[253] U. Chaudhuri, B. Banerjee, A. Bhattacharya, M. Datcu, CMIR-NET: a deep learning based model for cross-modal retrieval in remote sensing, Pattern Recognit Lett 131 (2020) 456–462.
[254] W. Xiong, Z. Xiong, Y. Cui, Y. Lv, A discriminative distillation network for cross-source remote sensing image retrieval, IEEE J. Sel. Topics in Appl Earth Obs. Remote Sens (2020) in press.
[255] W. Xiong, Y. Lv, X. Zhang, Y. Cui, Learning to translate for cross-source remote sensing image retrieval, IEEE Trans Geosci Remote Sensi (2020) in press.
[256] M. Eitz, K. Hildebrand, T. Boubekeur, M. Alexa, Sketch-based image retrieval: benchmark and bag-of-features descriptors, IEEE Trans Visualization Comput. Gr. 17 (11) (2011) 1624–1636.
[257] R. Hu, J. Collomosse, A performance evaluation of gradient field hog descriptor for sketch based image retrieval, Computer Vision and Image Underst. 117 (7) (2013) 790–806.
[258] Y. Qi, Y. Song, H. Zhang, J. Liu, Sketch-based image retrieval via Siamese convolutional neural network, in: Proceedings of ICIP, 2016, pp. 2460–2464.
[259] X. Wang, X. Duan, X. Bai, Deep sketch feature for cross-domain image retrieval, Neurocomputing 207 (2016) 387–397.
[260] T. Jiang, G. Xia, Q. Lu, W. Shen, Retrieving aerial scene images with learned deep image-sketch features, J Comput Sci Technol 32 (4) (2017) 726–737.
[261] F. Xu, W. Yang, T. Jiang, S. Lin, H. Luo, G. Xia, Mental Retrieval of Remote Sensing Images via Adversarial Sketch-Image Feature Learning, IEEE Trans. Geosci. Remote Sens. (2020).
[262] T. Abdullah, Y. Bazi, M. Rahhal, M. Mekhalfi, L. Rangarajan, M. Zuair, TextRS: deep bidirectional triplet network for matching text to remote sensing images, Remote Sens (Basel) 12 (2020) 405.
[263] D. Li, N. Dimitrova, M. Li, I. Sethi, Multimedia content processing through cross-modal association, in: Proceedings of the Eleventh ACM International Conference on Multimedia, 2003, pp. 604–611.
[264] H. Zhang, Y. Zhuang, F. Wu, Cross-modal correlation learning for clustering on image-audio dataset, in: Proceedings of the 15th ACM international conference on Multimedia, 2007, pp. 273–276.
[265] A. Torfi, S. Iranmanesh, N. Nasrabadi, J. Dawson, 3D Convolutional Neural Networks for Cross Audio-Visual Matching Recognition, IEEE Access 5 (2017) 22081–22091.
[266] A. Nagrani, S. Albanie, A. Zisserman, Seeing Voices and Hearing Faces: cross-modal biometric matching, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[267] M. Guo, Y. Yuan, X. Lu, Deep cross-modal retrieval for remote sensing image and audio, in: Proceedings of the 10th Workshop on Pattern Recognition in Remote Sensing, 2018, pp. 1–7.
[268] Y. Chen, X. Lu, A deep hashing technique for remote sensing image-sound retrieval, Remote Sens (Basel) 12 (1) (2020) 84.
[269] Y. Chen, X. Lu, S. Wang, Deep cross-modal image-voice retrieval in remote sensing, IEEE Trans. Geosci. Remote Sens. (2020).
[270] M. Costache, H. Maitre, M. Datcu, Categorization based relevance feedback search engine for earth observation images repositories, in: Proceedings of IGARSS, 2006, pp. 13–16.
[271] M. Schroder, H. Rehrauer, K. Seidel, M. Datcu, Interactive learning and probabilistic retrieval in remote sensing image archives, IEEE Trans Geosci. Remote Sens. 38 (5) (2000) 2288–2298.
[272] M. Klaric, G. Scott, C. Shyu, Mining visual associations from user feedback for weighting multiple indexes in geospatial image retrieval, in: Proceedings of IGARSS, 2006.
[273] Y. Li, T. Bretschneider, Remote sensing image retrieval using a context-sensitive Bayesian network with relevance feedback, in: Proceedings of IGARSS, 2006.
[274] C. Ma, Q. Dai, J. Liu, S. Liu, J. Yang, An improved svm model for relevance feedback in remote sensing image retrieval, Int J Digit. Earth 7 (9) (2014) 725–745.
[275] Y. Boualleg, M. Farah, Enhanced interactive remote sensing image retrieval with scene classification convolutional neural networks model, in: Proceedings of IGARSS, 2018.
[276] M. Ferecatu, N. Boujemaa, Interactive remote-sensing image retrieval using active relevance feedback, IEEE Trans Geosci. Remote Sens 45 (4) (2007) 818–826.
[277] B. Demir, L. Bruzzone, A novel active learning method in relevance feedback for content-based remote sensing image retrieval, IEEE Trans Geosc. Remote Sens. 53 (2015) 2323–2334.
[278] A. Griver, A. Radoi, C. Vaduva, M. Datcu, An active learning approach to the query by example retrieval in remote sensing images, in: Proceedings of International Conference on Communications, 2016, pp. 377–380.
[279] X. Tang, X. Zhang, F. Liu, L. Jiao, Circular Relevance Feedback for Remote Sensing Image Retrieval, in: Proceedings of IGARSS, 2018.
[280] X. Tang, L. Jiao, W. Emery, SAR image content retrieval based on fuzzy similarity and relevance feedback, IEEE J Sel Topics in Appl Earth Obs Remote Sens. 5 (10) (2017) 1824–1842.
[281] C. Ma, F. Chen, J. Liu, J. Duan, An Improved SVM+GA Relevance Feedback Model in the Remote Sensing Image Change Information Retrieval, in: Proceedings of IGARSS, 2018.
[282] P. Yin, et al., Integrating relevance feedback techniques for image retrieval using reinforcement learning, IEEE Trans Pattern Anal Mach Intell 27 (2005) 1536–1551.
[283] C. Corbiere, H. Ben-Younes, A. Rame, C. Ollion, Leveraging weakly annotated data for fashion image retrieval and label prediction, in: Proceedings of ICCV, 2017.
[284] C. Huang, S. Zhu, and K. Yu. Large scale strongly supervised ensemble metric learning, with applications to face verification and retrieval. arXiv. 2012.
[285] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. Hoi. Deep learning for person re-identification: a survey and outlook. arXiv. 2020.
[286] Y. Liu, F. Mo, P. Tao, Matching multi-source optical satellite imagery exploiting a multi-stage approach, Remote Sens (Basel) 9 (2017) 1249.
[287] M. Chen, A. Habib, H. He, Q. Zhu, W. Zhang, Robust feature matching method for SAR and optical images by using Gaussian gamma-shaped bi-windows-based descriptor and geometric constraint, Remote Sens (Basel) 9 (2017) 882.
[288] J. Li, C. Li, T. Yang, Z. Lu, Cross-domain co-occurring feature for visible-infrared image matching, IEEE Access 6 (2018) 17681–17698.
[289] J. Hays, A. Efros, IM2GPS: estimating geographic information from a single image, in: Proceedings of CVPR, 2008.
[290] A. Zamir, M. Shah, Image geo-localization based on multiple nearest neighbor feature matching using generalized graphs, IEEE Trans Pattern Anal Mach Intell 36 (8) (2014) 1546–1558.
[291] T. Weyand, I. Kostrikov, J. Philbin, PlaNet-Photo geolocation with convolutional neural networks, in: Proceedings of ECCV, 2016.
[292] G. Lu, Y. Yan, L. Ren, J. Song, N. Sebe, C. Kambhamettu, Localize me anywhere, anytime: a multi-task point-retrieval approach, in: Proceedings of ICCV, 2015.
[293] Y. Song, X. Chen, X. Wang, Y. Zhang, J. Li, 6-DOF image localization from massive geo-tagged reference images, IEEE Trans Multimedia 18 (8) (2016) 1542–1554.
[294] S. Hu, M. Feng, R. Nguyen, G. Lee, CVM-Net: cross-view matching network for image-based ground-to-aerial geo-localization, in: Proceedings of CVPR, 2018.
[295] F. Andert, S. Krause, Optical aircraft navigation with multi-sensor SLAM and infinite depth features, in: Proceedings of Int. Conf. Unmanned Air cr. Syst, 2017, pp. 1030–1036.
[296] Y. Yang, S. Newsam, Bag-of-visual-words and spatial extensions for land-use classification, in: Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2010, pp. 270–279.
[297] Qin Zou, Lihao Ni, Tong Zhang, Qian Wang, Deep learning based feature selection for remote sensing scene classification, IEEE Geosci. Remote Sens Lett 12 (11) (2015) 2321–2325.
[298] B. Zhao, Y. Zhong, G. Xia, L. Zhang, Dirichlet-Derived Multiple Topic Scene Classification Model Fusing Heterogeneous Features for High Spatial Resolution Remote Sensing Imagery, IEEE Trans. Geosci. Remote Sens 54 (4) (2016) 2108–2123.
[299] G. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, a benchmark dataset for performance evaluation of aerial scene classification, IEEE Trans. Geosci. Remote Sens. 55 (2017) 3965–3981.
[300] G. Cheng, J. Han, X. Lu, Remote sensing image scene classification: benchmark and state of the art, Proc IEEE 105 (2017) 1865–1883.
[301] H. Li, and et al. RSI-CB: a large scale remote sensing image classification benchmark via crowdsource data[J]. arXiv. 2017, arXiv:1705.10450.
[302] J. Pu, G. Xia, H. Fan, Q. Lu, L. Zhang, AID++: an updated version of AID on scene classification, in: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 4721–4724.
[303] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, R. Nemani, DeepSat: a learning framework for satellite imagery, in: Proceedings of ACM SIGSPATIAL, 2015.
[304] P. Helber, B. Bischke, A. Dengel, D. Borth, Introducing EuroSAT: a novel dataset and deep learning benchmark for land use and land cover classification, in: Proceedings of IGARSS, 2018, pp. 204–207.
[305] Z. Shao, K. Yang, X. Zhou, A benchmark dataset for performance evaluation of multi-label remote sensing image retrieval, Remote Sens (Basel) 10 (6) (2018) 964.
[306] Y. Hua, L. Mou, X. Zhu, Recurrently Exploring Class-wise Attention in A Hybrid Convolutional and Bidirectional LSTM Network for Multi-label Aerial Image Classification, ISPRS J. Photogramm Remote Sens 149 (2019) 188–199.
[307] G. Sumbul, M. Charfuelan, B. Demir, V. Markl, BigEarthNet: a Large-Scale Benchmark Archive for Remote Sensing Image Understanding, in: Proceedings of IGARSS, 2019.
[308] O. Dai, B. Demir, B. Sankur, L. Bruzzone, A Novel System for Content based Retrieval of Single and Multi-Label High Dimensional Remote Sensing Images, IEEE J Sel Topics in Appl Earth Obs Remote Sens 11 (7) (2018) 2473–2490.
[309] B. Chaudhuri, B. Demir, S. Chaudhuri, L. Bruzzone, Multi-label Remote Sensing Image Retrieval using a Semi-Supervised Graph-Theoretic Method, IEEE Trans. Geosci Remote Sens 56 (2) (2018) 1144–1158.
[310] M. Schmitt, L. Hughes, C. Qiu, X. Zhu, SEN12MS-A curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion, in: Proceedings of ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2019.

[311] B. Du, X. Li, D. Tao, X. Lu, Deep semantic understanding of high resolution remote sensing image, in: Proceedings of International Conference on Computer, Information and Telecommunication Systems, 2016.
[312] X. Lu, B. Wang, X. Zheng, X. Li, Exploring Models and Data for Remote Sensing Image Caption Generation, IEEE Trans. Geosci. Remote Sens. 56 (4) (2018) 2183–2195.
[313] F. Shen, C. Shen, W. Liu, H.T. Shen, Supervised discrete hashing, in: Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit, 2015, pp. 37–45.
[314] W.C. Kang, W.J. Li, Z.H. Zhou, Column sampling based discrete supervised hashing, in: Proceedings of AAAI Conf. Artif. Intell, 2016, pp. 1230–1236.
[315] H. Liu, R. Wang, S. Shan, X. Chen, Deep supervised hashing for fast image retrieval, in: Proceedings of Comput. Vis. Pattern Recognit, 2016, pp. 2064–2072.
[316] H. Zhu, M. Long, J. Wang, Y. Cao, Deep hashing network for efficient similarity retrieval, in: Proceedings of AAAI Conf. Artif. Intell., 2016, pp. 2415–2421.
[317] W. Li, S. Wang, W.C. Kang, Feature learning based deep supervised hashing with pairwise labels, in: Proceedings of Int. Joint Conf. Artif. Intell., 2016, pp. 1711–1717.
[318] D. Zhang, W. Li, Large-scale supervised multimodal hashing with semantic correlation maximization, in: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014, pp. 2177–2183.
[319] Q. Jiang, W. Li, Deep cross-modal hashing, in: Proceedings of IEEE Conf. Comput. Vis. Pattern Recognit, 2017, pp. 3270–3278.
[320] Yansheng Li, D. Ye, Greedy Annotation of Remote Sensing Image Scenes Based on Automatic Aggregation via Hierarchical Similarity Diffusion, IEEE Access 6 (2018) 57376–57388.
[321] G. Xia, Z. Wang, C. Xiong, L. Zhang, Accurate Annotation of Remote Sensing Images via Active Spectral Clustering with Little Expert Knowledge, Remote Sens (Basel) 7 (11) (2015) 15014–15045.
[322] W. Yang, et al., Learning high-level features for satellite image classification with limited labeled samples, IEEE Trans. Geosci. Remote Sens. 53 (2015) 4472–4482.
[323] X. Kang, P. Duan, X. Xiang, S. Li, J. Benediktsson, Detection and correction of mislabeled training samples for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 56 (2018) 5673–5686.
[324] B. Tu, X. Zhang, X. Kang, G. Zhang, S. Li, Density peak-based noisy label detection for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 57 (2019) 1573–1584.
[325] Yansheng Li, Y. Zhang, Z. Zhu, Learning deep networks under noisy labels for remote sensing image scene classification, in: Proceedings of IGARSS, 2019.
[326] Z. Shi, Z. Zou, Can a Machine Generate Humanlike Language Descriptions for a Remote Sensing Image? IEEE Trans. Geosci. Remote Sens. 55 (6) (2017) 3623–3634.
[327] Z. Zhang, W. Diao, W. Zhang, M. Yan, X. Gao, X. Sun, LAM: remote sensing image captioning with label-attention mechanism, Remote Sens (Basel) 11 (20) (2019) 2349.
[328] X. Zhang, Q. Wang, S. Chen, X. Li, Multi-scale cropping mechanism for remote sensing image captioning, in: Proceedings of IGARSS, 2019.
[329] S. Lobry, D. Marcos, J. Murray, and D. Tuia. RSVQA: visual question answering for remote sensing data. arXiv. 2020, arXiv:2003.07333.
[330] J. Shi, H. Zhang, J. Li, Explainable and explicit visual reasoning over scene graphs, in: Proceedings of IEEE CVPR, 2019.
[331] K. Tang, H. Zhang, B. Wu, W. Luo, W. Liu, Learning to compose dynamic tree structures for visual contexts, in: Proceedings of IEEE CVPR, 2019.
[332] D. Hudson, C. Manning, Learning by abstraction: the neural state machine, in: Proceedigns of NeurIPS, 2019.
[333] R. Krishna, and et al. Visual genome: connecting language and vision using crowsourced dense image annotations. arXiv. 2016, arXiv:1602.07332.