Article

Multiple Feature Hashing Learning for Large-Scale Remote Sensing Image Retrieval

Dongjie Ye [1], Yansheng Li [1],*, Chao Tao [2], Xunwei Xie [1] and Xiang Wang [1]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China;
    yedjchn@outlook.com (D.Y.); xwxie@whu.edu.cn (X.X.); mld@whu.edu.cn (X.W.)
[2] School of Geosciences and Info-Physics, Central South University, Changsha 410083, China;
    kingtaochao@csu.edu.cn
* Correspondence: liyansheng99@gmail.com; Tel.: +86-027-6877-0771

Abstract: Driven by the urgent demand of remote sensing big data management and knowledge discovery, large-scale remote sensing image retrieval (LSRSIR) has attracted more and more attention. As is well known, hashing learning has played an important role in coping with big data mining problems. In the literature, several hashing learning methods have been proposed to address LSRSIR. Until now, existing LSRSIR methods take only one type of feature descriptor as the input of hashing learning methods and ignore the complementary effects of multiple features, which may represent remote sensing images from different aspects. Different from the existing LSRSIR methods, this paper proposes a flexible multiple-feature hashing learning framework for LSRSIR, which takes multiple complementary features as the input and learns the hybrid feature mapping function, which projects multiple features of the remote sensing image to the low-dimensional binary (i.e., compact) feature representation. Furthermore, the compact feature representations can be directly utilized in LSRSIR with the aid of the hamming distance metric. In order to show the superiority of the proposed multiple feature hashing learning method, we compare the proposed approach with the existing methods on two publicly available large-scale remote sensing image datasets. Extensive experiments demonstrate that the proposed approach can significantly outperform the state-of-the-art approaches.

Keywords: multiple feature hashing learning; large-scale remote sensing image retrieval; remote sensing big data management

1. Introduction

Along with the rapid development of remote sensing observation technology, the volume of available remote sensing (RS) images has dramatically increased. We have entered an era of remote sensing big data [1–3]. It is well known that a large amount of actionable information hides in remote sensing big data. Information identification from remote sensing images based on manual labor is time-consuming and even impossible when the volume of remote sensing images is oversized. As one of the most fundamental problems in remote sensing big data mining, large-scale remote sensing image retrieval (LSRSIR) is a potential technique to automatically discover knowledge from remote sensing big data. Benefiting from the efforts from multi-domains, such as the remote sensing community and the computer vision community, large numbers of remote sensing image retrieval methods have been proposed and have achieved a certain degree of success when the volume of the remote sensing image dataset is relatively small. However, they often cannot be accustomed to the large-scale case. There exists an intense contradiction between the volume of remote sensing images and the capacity of existing remote sensing image processing methods. As a whole, LSRSIR is an urgent technique in remote sensing big data mining and deserves much more exploration.

As an effective method to manage a large number of images, content-based image retrieval (CBIR) can retrieve the interesting images according to their visual content. Recently, several kinds of CBIR methods have been utilized to cope with the RS image retrieval problem. As is well known, the CBIR performance largely relies on the capability and effectiveness of the feature representations. To characterize remote sensing images, many low-level features have been presented and evaluated in the remote sensing image retrieval task. More specifically, the proposed low-level features included spectral features [4–6], shape features [7–9], texture features [10–12], local invariant features [13], and so forth. Although low-level features have been employed with a certain degree of success, they have a very limited capability in representing the high-level concepts that are presented by remote sensing images (i.e., the semantic content). In order to mine the high-level concept of these low-level feature descriptors, Zhou et al. [14] exploited the auto-encoder model to encode the low-level features. Furthermore, a few graph-based approaches were utilized to remote sensing image retrieval [15–17], which represent and retrieve images by graph models. For instance, Du et al. [15] exploited the intrinsic structural information of the original data to learn the representation of images by incorporating graph regularization, while Chaudhuri et al. [17] presented an unsupervised graph-theoretic approach for region-based RS image retrieval. However, these proposed approaches still depend on hand-crafted features (e.g., the scale-invariant feature transform descriptor). In order to potentially mine the complete characteristic of the original remote sensing image, we proposed an unsupervised cross-diffusion graph model [18], which can collaboratively fuse multiple features, including the hand-crafted features and the data-driven features via multi-layer feature learning [19]. Although some encouraging progress has been made, developing suitable retrieval methods for LSRSIR remains an ongoing challenge, because existing methods depend highly on the high-dimensional feature descriptor and cannot be scalable to the large-scale remote sensing image retrieval task. Accordingly, the scalability problem and the storage of the image descriptors have become critical bottlenecks in LSRSIR.

Hashing learning is a potential technique to cope with big data retrieval because of its excellent ability in compact feature representation. Hashing learning methods generally construct a set of hash functions to project a high-dimensional feature vector into low-dimensional binary features, while preserving the original similarity structure when the image features are represented by binary hash codes. The binary codes can significantly reduce the amount of memory that is required for storing the content of images, and extremely improve the retrieval efficiency because the calculation of the pairwise distance can be performed efficiently in the low-dimensional binary feature space (i.e., hamming space). Thus, hashing learning is a potential and ideal approach to cope with big data problems. The existing hashing methods can be broadly classified into two categories: data-independent and data-dependent methods. Data-dependent algorithms require training data to learn the hashing mapping function. On the contrary, the hashing mapping function is empirically designed in the data-independent methods. Locality sensitive hashing (LSH) [20] is one of the representative data-independent methods that adopts random projections as hash functions without using any training data, while its practical efficiency is limited since it requires long hash codes to achieve a high retrieval performance. Data-dependent methods can learn compact hash codes effectively and organize massive amounts of data efficiently. It can be further divided into unsupervised hashing and supervised hashing methods. More specifically, unsupervised hashing does not utilize the label information of training examples [21–24]. On the contrary, supervised hashing methods try to incorporate semantic (label) information for hashing function learning [25–30]. Although hashing learning has been successfully applied in the natural image retrieval, few studies have been devoted to hashing learning-based RS image retrieval. Generally speaking, remote sensing images often contain abundant and complex visual contents [31]. As a consequence, the complex surface structures and very large variations of image resolutions pose a significant challenge to hashing learning-based RS image retrieval. Hence, it is of great interest to investigate the retrieval of RS images using hashing learning.

In the literature, several data-dependent hashing learning methods have been proposed to retrieve remote sensing images in the recent years. More specifically, Demir and Bruzzone [32] introduced the kernel-based nonlinear hashing learning methods to the remote sensing community. Afterwards, Li and Ren [33] proposed a novel unsupervised hashing method called partial randomness hashing (PRH), which aims to enable an efficient hashing function construction and learns a transformation weight matrix based on the training remote sensing images in an efficient way. However, most existing hashing models only consider one type of feature descriptor for learning hash functions. With the consideration that RS images are represented by complex image categories and various texture structures, the single feature descriptor is insufficient to provide a complete characterization of the RS image content. More recently, the study in [18] has shown the improvement of retrieval precision by multiple features fusion. As depicted in [18], unsupervised multilayer feature learning is utilized for remote sensing image retrieval. However, this approach used the graph-based cross-diffusion model to measure the similarity between the query image and the test image, which is difficult to extend into the large-scale RS image retrieval case due to the high computational complexity and storage complexity of graphs. Therefore, it is necessary to develop appropriate hashing learning methods for LSRSIR.

In this paper, we propose a novel multiple-feature hashing framework for large-scale remote sensing image retrieval (MFH-LSRSIR) to address the LSRSIR problem. Different from the hand-crafted features, which are empirically designed but generally lack a high generalization ability, the proposed approach exploits the data-driven feature by unsupervised multi-layer feature learning [18,19]. The experimental results have shown that the unsupervised features derived from unsupervised feature learning can achieve a higher precision than the conventional features in computer vision [18]. As a first attempt, the proposed approach utilizes the data-driven feature with one layer for calculation efficiency. As for remote sensing image, the unsupervised feature learning approach can automatically extract intrinsic features from RS imagery and can mine the spectral signatures with multiple bands. In addition, multiple features are represented according to different receptive fields. Generally, different features can reflect the different characteristics of one given image and play complementary roles. The features from different sizes of the receptive fields show complementary discrimination abilities. Hence, these features represent the RS image content in various spatial scales. Furthermore, the proposed method takes multiple features as the input and learns the hybrid feature mapping function. The proposed MFH-LSRSIR framework involves two main modules: the feature representation module and the hashing learning module. The feature representation module combines different features with serial fusion to construct multiple representations of images. Based on the hybrid feature vector, the hashing learning module explores the hashing function. The main contributions of this paper are summarized as follows:

- Based on recent supervised hashing learning method, a flexible framework for LSRSIR, named MFH-LSRSIR, is proposed by exploiting data-driven features from multi-spectral bands and investigates hashing learning approach to project the high-dimensional feature to low-dimensional binary feature.
- When considering the characteristics of RS images, the unsupervised feature learning approach with different receptive fields are proposed to generate multiple features of each image, which are further taken as the input of MVH-LSRSIR. The adopted features can make full use of the spectra information and the spatial context. Experimental results show that the multiple features-based method can outperform the single feature-based method.
- The complexity of the adopted hashing learning method is mainly concentrated on the optimization of hash code matrix, which is irrelevant to the feature input. Hence, the advocated hashing learning approach is impactful to implement multiple feature hashing learning.

The other parts of this paper are organized as follows. In Section 2, the proposed MFH-LSRSIR is described in detail. Section 3 presents the experimental results and gives qualitative and quantitative comparisons with existing approaches, and Section 4 provides a summary of this paper.

2. Multiple Feature Hashing Learning for Large-Scale Remote Sensing Image Retrieval

The proposed multiple feature hashing method for large-scale remote sensing image retrieval (MFH-LSRSIR) framework consists of three modules: (1) multiple feature representation; (2) multiple feature hashing learning; and, (3) hamming distance ranking. The flowchart is shown in Figure 1.

As depicted in Figure 1, the proposed MFH-LSRSIR framework contains an offline training stage and an online retrieval stage. In the multiple feature representation step, a series of feature sets are obtained from the training data of image dataset by unsupervised feature learning. Then, the features sets further combine to take multiple complementary features as the input and learn the hybrid feature hashing function. In the hamming distance ranking stage, for a given query image, calculating the distance of a query image corresponding to the hash code with other hash codes in the hamming space is performed. Finally, ranking the similarity between the query image with each image in the dataset based on the hamming distance is conducted to obtain the retrieval result.
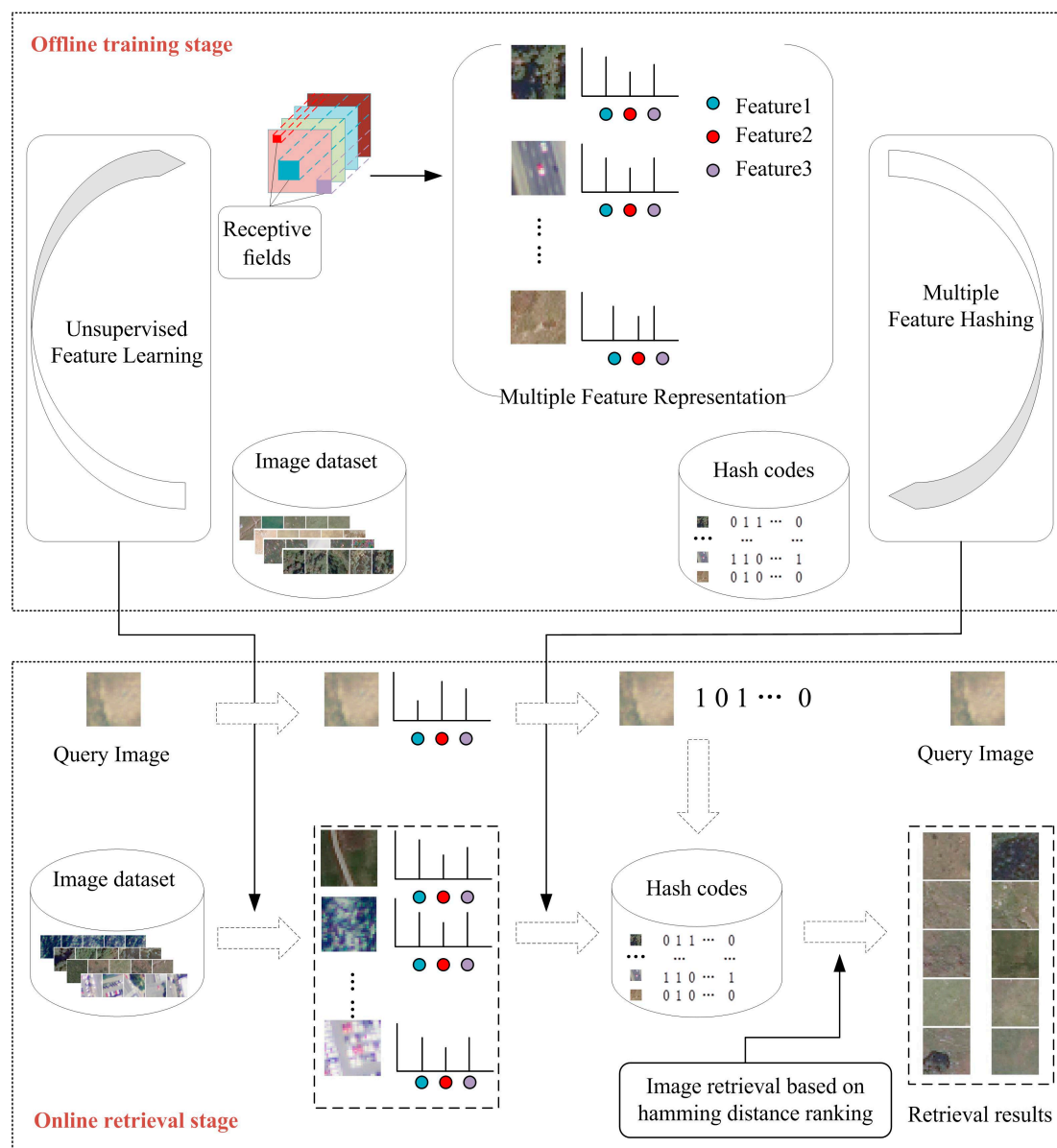
Figure 1. Flowchart of the proposed multiple feature hashing learning method for large-scale remote sensing image retrieval (MFH-LSRSIR).

2.1. Multiple Feature Representation

Most of the works in the image retrieval literature focus on feature extraction because retrieval performance greatly depends on the power of feature representations. However, most of existing image feature techniques and methods are too limited to represent large-scale RS image features due to the complexity of RS data. For instance, Li and Ren [33] adopted hand-crafted Gist feature extraction to characterize remote sensing images, which led to the loss of spectra information. On the contrary, Han and Chen [34] investigated a hybrid aggregation of multi-spectral analysis approach for remote sensing image feature extraction, but it only takes one scale of spatial information into account. Accordingly, the unsupervised feature learning approach has shown an encouraging performance [18,19]. As is well known, remote sensing images contain rich structure information. In our implementation, the unsupervised feature learning approach is adapted to fully depict the visual content of remote sensing images by employing multispectral signatures with near-infrared band and visible light bands (red, green, blue). Therefore, the advocated unsupervised feature learning method can make full use of the spectra information and the spatial context, simultaneously.

Moreover, in order to describe different scales of spatial context information, we adopt multiple receptive fields to generate the feature set of each image for MFH-LSRSIR. Experimental results show that multiple scales of spatial information can further improve the image retrieval performance.

The RS image can be depicted by a set of features using the different sizes of receptive field. To improve the image retrieval performance, it is desirable to incorporate these heterogeneous feature descriptors into hash function learning, leading to the multiple hashing approach. In this paper, our proposal is to fuse multiple view information with a serial strategy [35].

Suppose that we have n RS images for training data, the feature set contains K types of features for each image. $\mathbf{x}_i = \left\{ f_i^1, f_i^2, \cdots, f_i^k \right\}$ is the feature set of the i-th image data, where $f_i^k \in R^{d(k)}$ denotes the vector of the k-th type of feature and $d(k)$ denotes the dimension of the k-th type of feature. Our method constructs the hybrid feature $\mathbf{x}_i = \left(f_i^1, \cdots, f_i^k \right)^T$ by a serial fusion strategy. The total feature dimension of x_i is defined as $D = [d(1) + d(2) + \cdots + d(k)]$.

As depicted in [19], the more layers that the unsupervised feature learning has, the better the performance of the generated features. However, more layers would remarkably increase the computational complexity. In order to achieve the balance between performance and complexity, this paper only adopts a single-layer network, but extracted the unsupervised features from multi-spectral images, including the near-infrared spectrum and visible spectrum (R, G, B). The significant accuracy gains of the experimental results have shown that the unsupervised feature with the single-layer network has the sufficient ability to characterize remote sensing images.

In addition, the features with different receptive fields of the unsupervised feature extraction network construct different spatial scales of the image. This process is the same as the visual perception ability from simple to complex, which shows complementary discrimination abilities. Hence, this paper attempts to extract many complementary features to depict the remote sensing images. As illustrated in Figure 2, the single-layer unsupervised feature extraction network includes three basic operations:

(1) The convolutional operation: convolutional operation works for feature mapping, which is constrained by the function bases. More specifically, the function bases are generated by unsupervised K-means clustering. Through convolutional operation, we can map the d-channel's RS image to the k-channel's (i.e., the number of cluster) image. Given any w-by-w image patch (i.e., the receptive field), we can thus compute a new k-channel representation of the RS image for that patch.

(2) The local pooling operation: local pooling works to keep the layer invariant to slight translation and rotation and is implemented by the traditional calculation process (i.e., the local maximum).

(3) The global pooling operation: global pooling is implemented by sum-pooling in multiple large windows, which facilitates improving the feature discrimination efficiency. We implement the

global pooling operation by sum-pooling into four equal-sized quadrants, and integrate the multiple sum-pooling results as a feature vector. Therefore, the dimension of the feature vector is 4*k*.
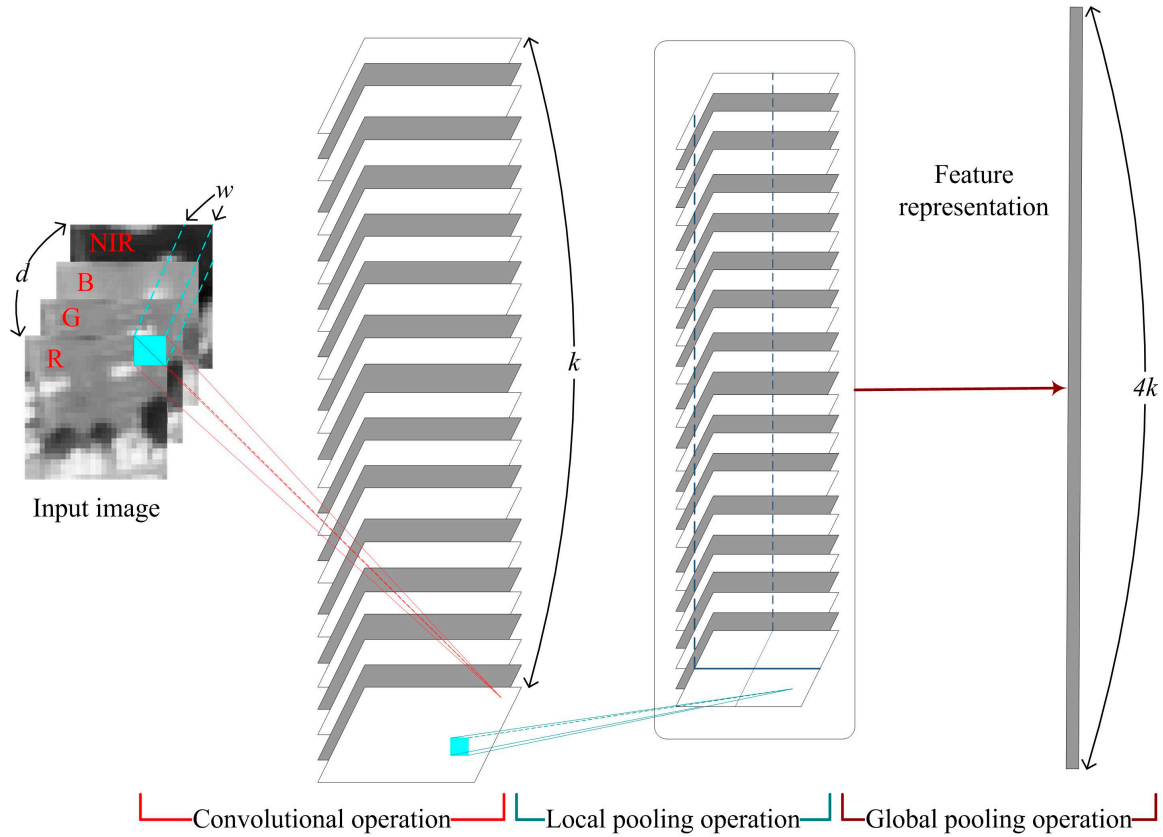
Figure 2. The unsupervised feature extraction network.

2.2. Multiple Feature Hashing Learning Based on Column Sampling

In the hashing learning module, our approach is to learn the hybrid feature mapping function to generate a binary code for a fast search in the hamming space based on column sampling hashing [29]. Specifically, denote $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_n\} \in \mathrm{R}^{n \times D}$ as the whole data matrix. In addition to the feature vectors, we assume the semantic similarity matrix $\mathbf{S} \in (-1, 1)^{n \times n}$ that does not miss label entries, where $S_{ij} = 1$ means that $\mathbf{x}_i$ and $\mathbf{x}_j$ are the similar pairs (with the same label), $S_{ij} = -1$ means that $\mathbf{x}_i$ and $\mathbf{x}_j$ are the dissimilar pairs (with different labels). The goal of hashing is to learn a binary code matrix $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots \mathbf{b}_n\} \in (1, -1)^{n \times r}$ to preserve their similarities in the original space, where $\mathbf{b}_i$ denotes the *r*-bit code for $\mathbf{x}_i$. For one RS image feature vector $\mathbf{x}_i$, we adopt the commonly used hashing function form, which project it from the *D*-dimensional feature space to an *r*-dimensional hamming space by:

$$\mathbf{b}_i = f(\mathbf{x}_i) = sgn\left(\mathbf{W}^T\mathbf{x}_i\right) \qquad (1)$$

where sgn($\cdot$) is the element-wise sign function, which returns 1 if the element is a positive number and the other returns -1. $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r]$ is the projection matrix.

Similarly, the whole data matrix is mapped to the hamming space as follows:

$$\mathbf{B} = f(\mathbf{X}) = sgn\left(\mathbf{W}^T\mathbf{X}\right) \qquad (2)$$

The objective function of the optimization problem can be defined as:

$$\min_{\mathbf{B}\in(1,-1)^{n\times r}}\|r\mathbf{S}-\mathbf{B}\mathbf{B}^T\|_F^2 \tag{3}$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix.

According to Equation (3), the objective function learns binary code matrix **B** based on the semantic similarity matrix **S**, so our method is insensitive to the dimension of the feature vector. Moreover, existing methods attempt to sample only a small subset with m (m < n) points for training and discard the rest of the training points, which leads to unsatisfactory accuracy. Column sampling hashing adopts a strategy that can effectively exploit all of the training data by sampling columns. This is to say, the strategy proposes to sample several columns from **S** in each iteration and several iterations are performed for training. By randomly sampling a set Ω of $\mathcal{N} = \{1, 2, \ldots, n\}$ and then sampling $|\Omega|$ column of **S**, **S** is divided into two kinds of parts in each iteration, one being those indexed by Ω and the other being those indexed by $\Gamma = N - \Omega$. Then, Equation (3) is associated with the sampled columns in each iteration it can be reformulated as follows:

$$\min_{\mathbf{B}^\Omega,\mathbf{B}^\Gamma}\|r\overline{\mathbf{S}}^\Gamma - \mathbf{B}^\Gamma[\mathbf{B}^\Omega]^T\|_F^2 + \|r\overline{\mathbf{S}}^\Omega - \mathbf{B}^\Gamma[\mathbf{B}^\Omega]^T\|_F^2 \tag{4}$$

where $\overline{\mathbf{S}}^\Omega \in \{-1,1\}^{|\Omega|\times|\Omega|}$, $\overline{\mathbf{S}}^\Gamma \in \{-1,1\}^{|\Gamma|\times|\Omega|}$, $\mathbf{B}^\Omega \in \{-1,1\}^{|\Omega|\times r}$, and $\mathbf{B}^\Gamma \in \{-1,1\}^{|\Gamma|\times r}$.

The optimization of Equation (4) involves two alternating steps: (1) updating $\mathbf{B}^\Gamma$ with $\mathbf{B}^\Omega$ fixed; and (2) updating $\mathbf{B}^\Omega$ with $\mathbf{B}^\Gamma$ fixed. This two-step alternating optimization procedure will be repeated several times.

Updating $\mathbf{B}^\Gamma$ with $\mathbf{B}^\Omega$ fixed: By fixing $\mathbf{B}^\Omega$, the objective function of $\mathbf{B}^\Gamma$ is given by:

$$\min_{\mathbf{B}^\Gamma\in\{-1,1\}^{|\Gamma|\times r}} = \|r\overline{\mathbf{S}}^\Gamma - \mathbf{B}^\Gamma[\mathbf{B}^\Omega]^T\|_F^2 \tag{5}$$

Through changing the loss from Frobenius norm to L1 norm, the $\mathbf{B}^\Gamma$ is easily to be computed:

$$\mathbf{B}^\Gamma = sgn\left(\overline{\mathbf{S}}^\Gamma\mathbf{B}^\Omega\right) \tag{6}$$

Updating $\mathbf{B}^\Omega$ with $\mathbf{B}^\Gamma$ fixed: When $\mathbf{B}^\Gamma$ is fixed, the objective function of $\mathbf{B}^\Omega$ is defined as:

$$\min_{\mathbf{B}^\Omega\in\{-1,1\}^{|\Omega|\times r}} = \|r\overline{\mathbf{S}}^\Gamma - \mathbf{B}^\Gamma[\mathbf{B}^\Omega]^T\|_F^2 + \|r\overline{\mathbf{S}}^\Omega - \mathbf{B}^\Omega[\mathbf{B}^\Omega]^T\|_F^2 \tag{7}$$

According to the 2-approximation algorithm, the *k*-th column of $\mathbf{B}^\Omega$ can be acquired in the *t*-th iteration. As a result, we can recover **B** by combining $\mathbf{B}^\Omega$ and $\mathbf{B}^\Gamma$. Please refer to [29] for details.

By choosing linear regularized least-squares classifier, we use linear regression to train **W** over the training set. The optimal **W** can be computed as:

$$\mathbf{W} = \left(\mathbf{X}^T\mathbf{X} + \mathbf{I}\right)^{-1}\mathbf{X}^T\mathbf{B} \tag{8}$$

Therefore, the hashing codes $\mathbf{b}_q$ for a new query image can be computed as follows:

$$\mathbf{b}_q = f(\mathbf{x}_q) = sgn\left(\mathbf{W}^T\mathbf{x}_q\right) \tag{9}$$

As a whole, the main algorithm complexity of our proposed hashing learning method is to optimize B instead of the image features. Thus, when compared with single features, multiple feature hashing learning can not only affect the complexity, but can also flexibly incorporate the merit of multiple features.

2.3. Hamming Distance Ranking

The query image (i.e., the $\mathbf{x}_q$) and each image of training set is represented by a binary code through the above steps. We used the hamming distance as the similarity measure to compare two images' degree of similarity and the ranking of the hamming distance can be treated as retrieval result. For binary strings $\mathbf{b}_1$ and $\mathbf{b}_2$, the hamming distance is equal to the number of ones in $\mathbf{b}_1$ XOR $\mathbf{b}_2$.

The specific implementation of the proposed MFH-LSRSIR is summarized in Algorithm 1.

Algorithm 1. MFH-LSRSIR for large-scale RS image retrieval.

Input: the large-scale remote sensing image dataset that contains n images, testing query $\mathbf{x}_q$ and code length r.

1. Calculate the feature set $\mathbf{x}_i = \left\{ f_i^1, f_i^2, \cdots, f_i^k \right\}$ and the whole data matrix $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n\} \in \mathrm{R}^{\mathrm{n} \times \mathrm{D}}$.

2. Construct the multiple feature set $\mathbf{x}_i = \left(f_i^1, \cdots, f_i^k \right)^T$ by serial fusion and get the whole data matrix $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n\} \in \mathrm{R}^{\mathrm{n} \times \mathrm{D}}$.

3. **Repeat:**

    Sample $|\Omega|$ columnto set up $\overline{\mathbf{S}}^{\Omega}$, $\overline{\mathbf{S}}^{\Gamma}$, $\mathbf{B}^{\Omega}$, and $\mathbf{B}^{\Gamma}$.

    Loop until converge or reach maximum iterations:

        Calculate $\mathbf{B}^{\Gamma}$ using Equation (6) by fixing $\mathbf{B}^{\Omega}$.

        Compute $\mathbf{B}^{\Omega}$ by solving the problem (7) when $\mathbf{B}^{\Gamma}$ is fixed.

    Recover $\mathbf{B}$ by combining $\mathbf{B}^{\Omega}$ and $\mathbf{B}^{\Gamma}$.

    Until converge or reach maximum iterations.

4. Compute the optimal parameter $\mathbf{W}$ according to (8), and compute the binary code for the image database and query image by $\mathbf{B} = f(\mathbf{X}) = sgn(\mathbf{W}^T\mathbf{X})$ and $\mathbf{b}_q = f(\mathbf{x}_q) = sgn(\mathbf{W}^T\mathbf{x}_q)$.

5. Get the indexes of the most related images by ranking hamming distance.

Output: Binary code matrix $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots \mathbf{b}_n\} \in (1, -1)^{\mathrm{n} \times \mathrm{r}}$ and binary code of query image $\mathbf{b}_q \in (1, -1)^{\mathrm{r}}$, the most related images.

Finally, we give the computational complexity and running time of key modules. With the consideration that the training stage can be pre-processed in an offline stream, we focus on the complexity analysis of the test stage as it reflects the actual efficiency of the proposed method. For the multiple feature representation stage, the complexity of the feature extraction is $O(whkv)$, where w and h represent the width and height of the image data, respectively, k is the cluster number and $v = w \times w \times d$ stands for the dimension of the image patches with the w-by-w receptive field. The average running time of feature extraction per image is 0.0178 s. For the multiple feature hashing learning stage, the complexity of the hybrid features mapping is $O(lr)$, where l represents the dimension of the hybrid feature, and r is the length of hash code. The average time consumption of hybrid feature mapping per image is 0.000058 s. Similar to other hashing methods, the final hashing codes can be efficiently utilized to retrieve similar remote sensing images. As a whole, the proposed method is not only very effective, but efficient.

3. Experimental Result

In this section, we first introduce two adopted evaluation datasets and criteria in Section 3.1; Section 3.2 analyzes the sensitivity of the key parameter in clustering; Section 3.3 demonstrates the retrieval result on the two datasets, and analyzes the performance with different features of the multiple feature hashing method for a large-scale remote sensing image retrieval (MFH-LSRSIR) framework; Section 3.4 provides a comparison of the results with those of state-of-the-art approaches.

3.1. Evaluation Datasets and Criteria

Two recently released large-scale remote sensing datasets with semantic labels are used to verify the superiority of our proposed method. They are SAT-4 and SAT-6 airborne datasets, which were extracted from the National Agriculture Imagery Program (NAIP) dataset [36]. SAT-4 consists of a

total of 500,000 image patches that are covering four broad land cover classes. These include barren land, trees, grassland, and a class that consists of all land cover classes other than the above three. SAT-6 consists of a total of 405,000 image patches and covering six land-cover classes: barren land, trees, grassland, roads, buildings, and water bodies. The images consist of four bands: red, green, blue, and near-infrared (NIR), and each image patch is size normalized to 28×28 pixels. The sample images from each class in these two datasets are shown in Figures 3 and 4.

For all of the methods, we randomly choose 1000 points as query (test) set with the rest of the data as the training set. The experimental results are reported in terms of mean average precision (MAP) and precision-recall curves to evaluate the retrieval performance in the literature. The MAP score is calculated by:

$$MAP = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{n_i} \sum_{k=1}^{n_i} precision(R_{ik}) \tag{10}$$

where $q_i \in Q$ is a query and n_i is the number of points relevant to q_i in the dataset. Suppose that the relevant points are ordered as $\{r_1, \ r_2, \ \cdots, \ r_{n_i}\}$, and then R_{ik} is the set of ranked retrieval results from the top result until getting to r_k [33].
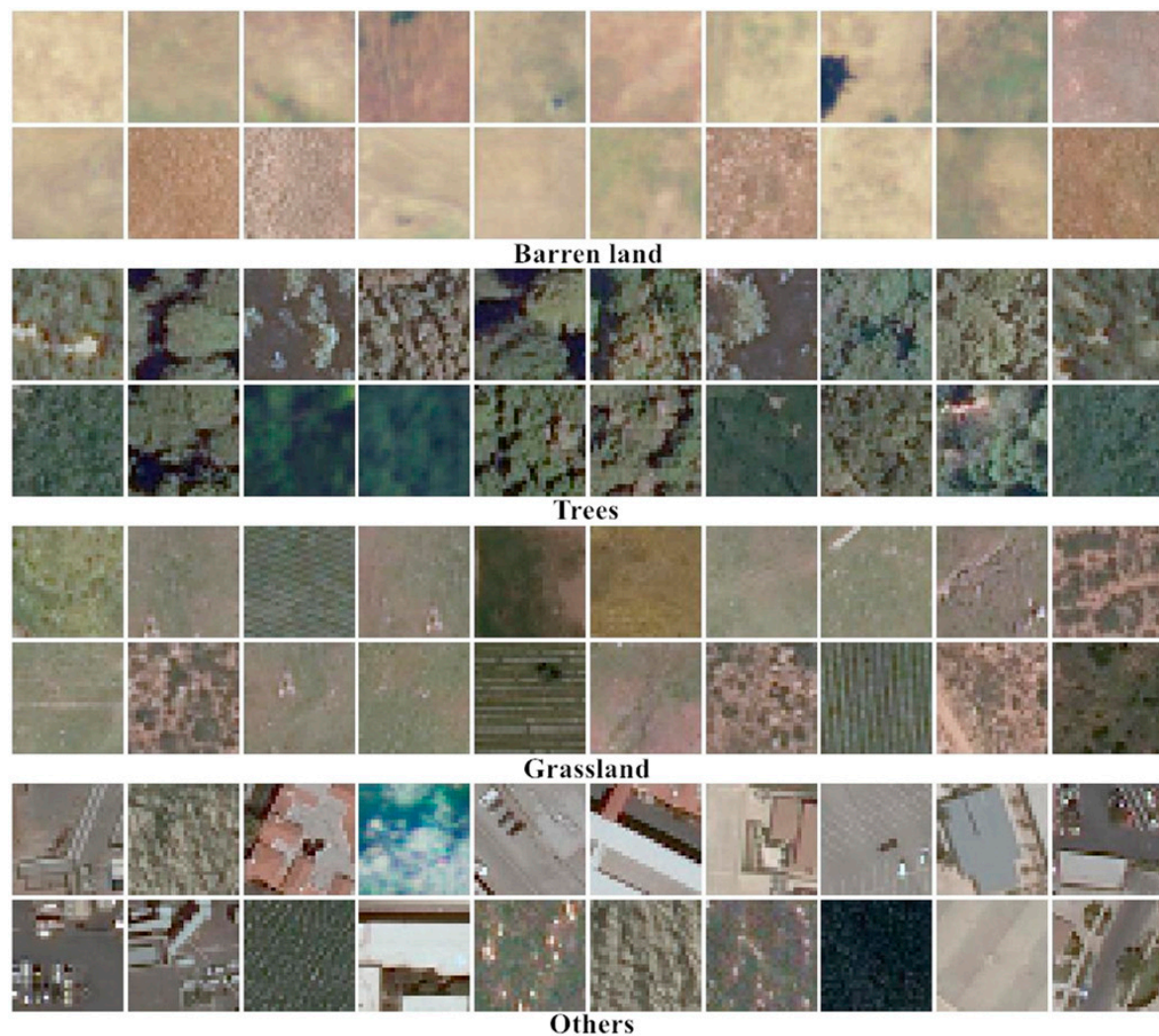
Figure 3. Visual samples of the SAT-4 dataset.
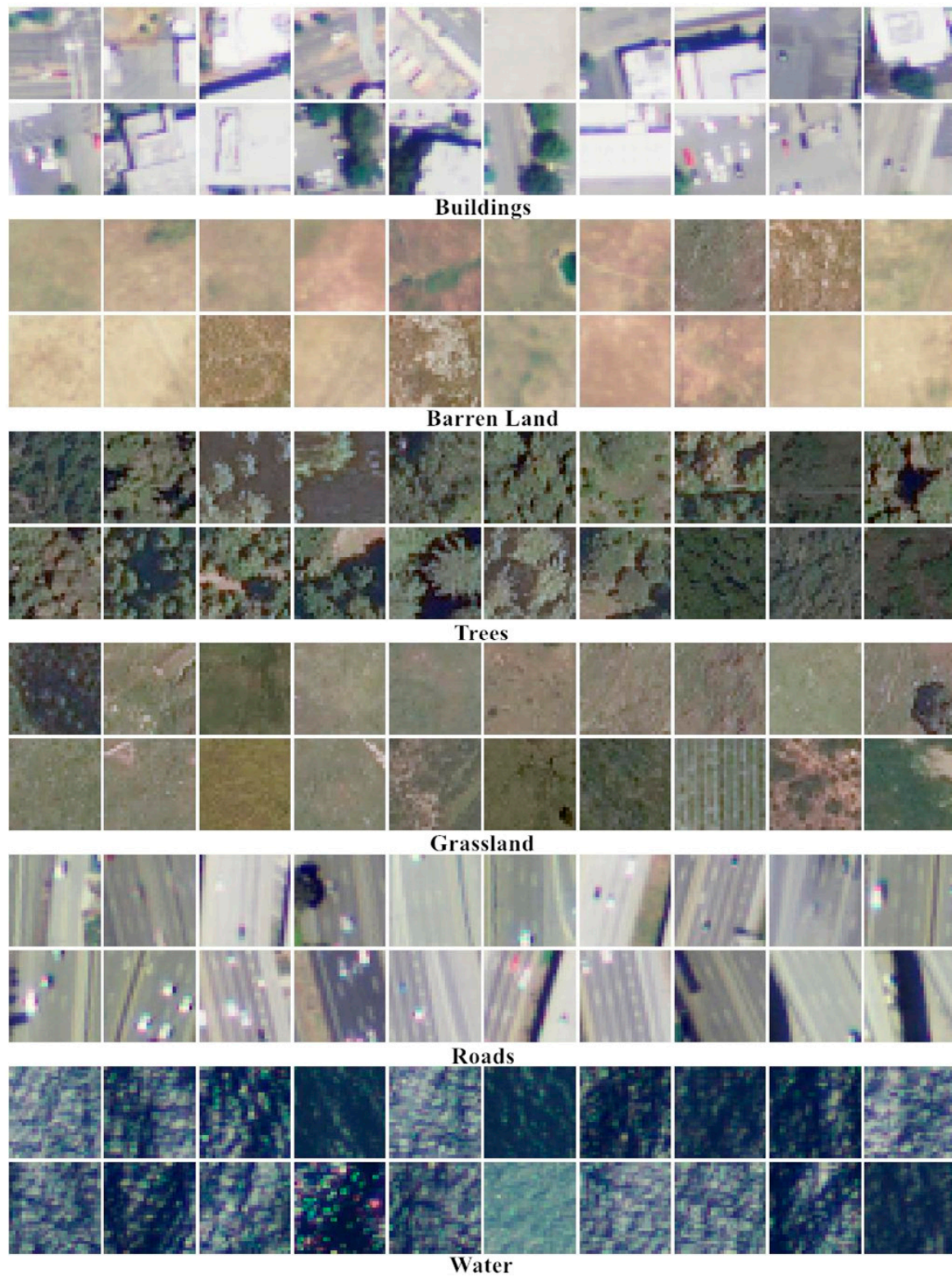
Figure 4. Visual samples of the SAT-6 dataset.

Furthermore, we also take the precision-recall curve as the evaluation indicator. More specifically, precision and recall are defined as below:

$$precision = \frac{tp}{tp + fp} \tag{11}$$

$$recall = \frac{tp}{tp + fn} \tag{12}$$

where tp is the number of similar points, fp is the number of non-similar points, and fn is the number of similar points that are not retrieved [30].

In the following, all experiments are conducted on Cloud Virtual Machine with an Intel E5-2667 Broadwell (v4) 3.2 GHz CPU and 32 GB RAM. We evaluate the results of the SAT-4 dataset and SAT-6 dataset, respectively.

3.2. Sensitivity Analysis of the Key Parameter

In the process of multiple feature representation based on unsupervised feature learning, a lot of parameters are involved. The selection of the number of clusters is critical for the whole method. Furthermore, our experiments considered cluster numbers with 64, 128, 256, and 512 to analyze the effect on the MAP scores. In this comparison, we consider one feature when the receptive field is set to 2 and the length of hash codes is fixed to 32.

Figure 5 clearly shows effect of the number of clusters. The MAP score lifts along with the increase of the cluster number on SAT-4 and SAT-6 datasets. When compared with the cluster number 64 and 128, $k = 256$ can obviously lift the MAP score. Although $k = 512$ can also improve the MAP score, the increasing amplitude of accuracy has been relatively small. Based on the aforementioned computational complexity introduction discussion in Section 2.3, the computational complexity is linearly correlated with the number of clusters. With the overall consideration of the retrieval accuracy and computational efficiency, the cluster k is set to 256.
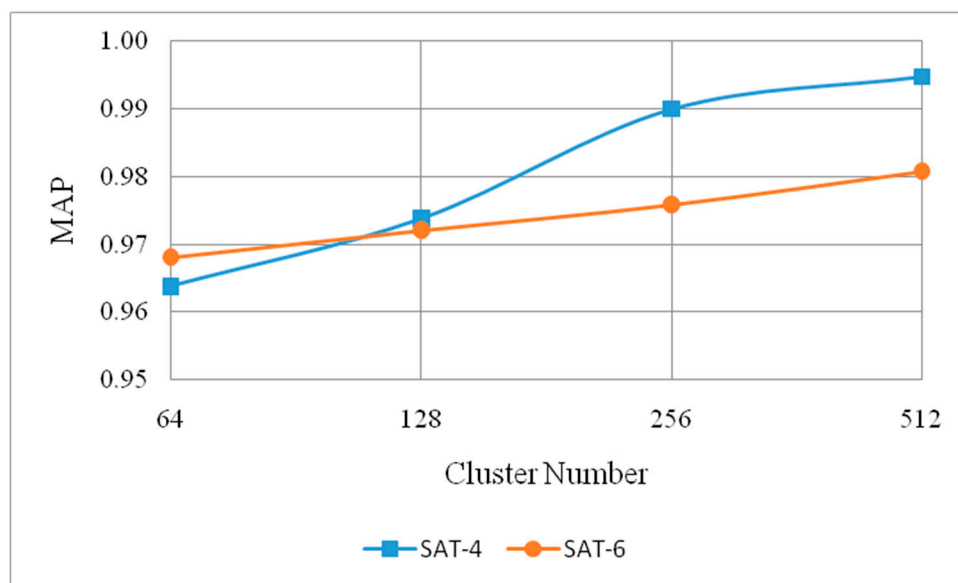
Figure 5. MAP score under different cluster numbers on SAT-4 and SAT-6 datasets.

3.3. Superiority of Multiple Feature Hashing Learning

The number of cluster k is set to 256 and the dimension of the final feature vector is 1024. Given one input remote sensing image, we make full use of ample band information (i.e., the near-infrared band and visible light bands) to obtain different types of features via the different sizes of the receptive field. In Table 1, the three different types of features represented by the receptive field sizes w with 2, 4, 6 are abbreviated as UF1, UF2, and UF3, respectively.

Table 1. Feature set for remote sensing images.

Feature Type	Feature Dimension	Receptive Field
UF1	1024	2
UF2	1024	4
UF3	1024	6

In the framework of our proposed MFH-LSRSIR, different type of feature introduced in Section 2.2 and multiple features are tested for demonstrating the complementary characteristics of the introduced features. MAP is one of the most comprehensive criteria to evaluate the retrieval performance in the literature. Different types of features' MAP scores with diverse hash bits for MFH-LSRSIR method on the SAT-4 dataset are shown in Table 2.

Table 2. MAP scores of the SAT-4 dataset.

Method	8 Bits	16 Bits	32 Bits	64 Bits
MFH-LSRSIR (UF1)	0.9801	0.9899	0.9900	0.9902
MFH-LSRSIR (UF2)	0.9731	0.9874	0.9844	0.9856
MFH-LSRSIR (UF3)	0.9656	0.9659	0.9681	0.9733
MFH-LSRSIR (UF1 + UF2 + UF3)	0.9914	0.9952	0.9958	0.9952

As demonstrated in Table 2, generally, our MFH-LSRSIR with different unsupervised feature based on various receptive fields achieve higher MAP scores. The best performance on each hash code length is achieved by the multiple feature that combination of UF1, UF2, and UF3. It is reasonable that different features can be the different characteristics of one given image, and the multiple feature plays complementary roles of various features to improve the retrieval accuracy. The best result is obtained by multiple feature on 64 bits while the worst MAP is 96.56% by UF3 on 8 bits. In addition, it can also be observed that the hash code length also has effect on the MAP scores, longer length of hash code achieves the higher retrieval accuracy in most case. Among these unsupervised features, UF1 that extracted by two receptive fields can achieve the best performance than any single feature. This is because, in the case of small image size (28 × 28 pixels), two receptive fields could better learn the detailed features of the image.

The MAP score results of the SAT-6 dataset are depicted in Table 3. Similar to the result on the SAT-4 dataset, the multiple feature also achieves the best performance on the SAT-6 dataset, the MAPs are 98.66%, 98.37%, 98.78%, and 99.00% when the hash code length is 8, 16, 32, and 64 bits, respectively. We can see that the improvement is most visible on the shorter hash code length.

Based on the intuitive results, the multiple features achieve the best remote sensing image retrieval performance.

Table 3. MAP scores of the SAT-6 dataset.

Method	8 Bits	16 Bits	32 Bits	64 Bits
MFH-LSRSIR (UF1)	0.9769	0.9744	0.9759	0.9752
MFH-LSRSIR (UF2)	0.9410	0.9779	0.9793	0.9817
MFH-LSRSIR (UF3)	0.9569	0.9832	0.9837	0.9850
MFH-LSRSIR (UF1 + UF2 +UF3)	0.9866	0.9837	0.9878	0.9900

3.4. Comparison with the State-of-the-Art Approaches

In order to validate the effectiveness of our presented method, the proposed method is compared with recent unsupervised hashing approach partial randomness hashing (PRH) [33] and some representative supervised hashing methods, including supervised hashing with kernels (KSH) [26], supervised discrete hashing (SDH) [28], and column sampling-based discrete hashing (COSDISH) [29]. We implement the PRH method by ourselves and the other approaches are implemented by the public source code provided by the corresponding authors. All of the other parameters are tuned to the best performance. These methods all extract a 512-dimension Gist feature vector for each image. For KSH and SDH, 1000 randomly selected anchor points are used. For KSH, we cannot use the entire training set for training due to the high time complexity; thus, we randomly sample 5000 training data of KSH.

3.4.1. Comparison on the SAT-4 Dataset

Table 4 presents the performance comparison with other studies on the SAT-4 datasets. It shows that our approach has surprisingly high-rate MAP score than the compared methods. For example, the proposed method obtains MAP score of 99.14% with 8 bit hash codes, whereas the MAP of up to 99.52% is achieved with hash code lengths of 64 bits, which always outperformed the second best by about 30% MAP rates. It is clear that, on this large dataset, the proposed method significantly outperforms all of the other approaches by even larger gaps.

This makes sense because of two main reasons. First, the description ability of the hand-crafted features may become limited, or even impoverished, for remote sensing images with complex scenes. By learning features from images instead of relying on manually designed features, we can obtain more discriminative features that are better suited to the problem at hand. Moreover, the unsupervised features can be extracted from multi-spectral imagery, including the near-infrared spectrum and visible spectrum (R, G, B). This enables satisfactorily describing the remotely-sensed image. Second, the hybrid feature is learned by our supervised hashing method, which has the capability to increase discrimination among hash codes and to satisfy the semantic similarity between the images.

Table 4. Comparison MAP scores for different methods with varied hash bits in the SAT-4 dataset.

Method	8 Bits	16 Bits	32 Bits	64 Bits
KSH in [26]	0.5249	0.5379	0.5490	0.5522
SDH in [28]	0.6419	0.6411	0.6347	0.6217
PRH in [33]	0.3975	0.3877	0.4139	0.4188
COSDISH in [29]	0.6449	0.6733	0.6511	0.6809
Ours MFH-LSRSIR	0.9914	0.9952	0.9958	0.9952

The precision-recall curves, which reflect the overall image retrieval performance of different hashing methods, are shown in Figure 6. As illustrated in Figure 6, the precision of the proposed method always retained a high value of close to 1 even with the increase in recall rate. It is interesting to note that the precision is improved, in particular, for larger recalls. This trend is particularly pronounced with the longer bit lengths. This is reasonable since our proposed method is based on the column sampling method, which can exploit all of the available data for training. When the recall rate is high, the correct relevant results have been returned more in the retrieval results owing to the powerful hybrid feature representations of image that are composed of multiple scales.
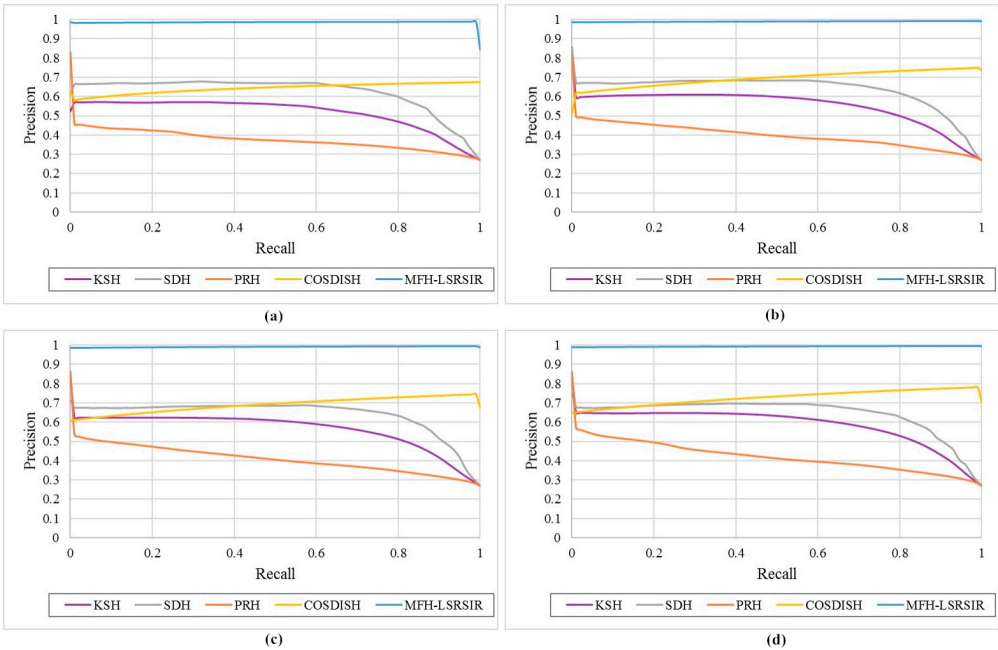
Figure 6. Compared precision-recall curves between different hashing methods with varied hash code lengths on the SAT-4 dataset: (**a**) 8 bits; (**b**) 16 bits; (**c**) 32 bits; and, (**d**) 64 bits.

Figure 7 provides a visual comparison of the different methods with 64 bits, where the retrieval result is set to sample every 500 images. In Figure 7, our MFH-LSRSIR can precisely output the right image scenes based on the query. Therefore, the results show that our results were more accurate.

Figure 7. Visualized retrieval results of the SAT-4 dataset by different hashing methods with 64 bit hash codes. The red rectangles indicate incorrect retrieval results. The abscissa represents the sequence number of the image.

3.4.2. Comparison to the SAT-6 Dataset

Table 5 compares the proposed approach with the published results of the SAT-6 dataset. It exhibits that our approach achieves the best performance on each hash code length. Our method achieved 99.00% on MAP when the code length was 64 bits, while the second highest MAP was 78.53%. It can be observed that the use of multiple features for hashing learning indeed have intensive characterization of remote sensing image to increase the retrieval precision.

Figure 8 displays the precision-recall curves of the SAT-6 dataset. We can observe that our MFH-LSRSIR method still consistently outperforms the alternatives. In fact, the MAP score is the area under the precision-recall curve; thus, these results in Figure 8 are consistent with the trends that we observe in the above experiments, which validate the superiority of our MFH-LSRSIR method.
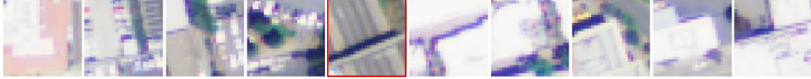
Table 5. Comparison MAP scores for different methods with varied hash bits in the SAT-6 dataset.

Method	8 Bits	16 Bits	32 Bits	64 Bits
KSH in [26]	0.6020	0.6191	0.6411	0.6351
SDH in [28]	0.6310	0.6429	0.6449	0.6617
PRH in [33]	0.4796	0.5044	0.5041	0.5146
COSDISH in [29]	0.7438	0.7641	0.7600	0.7853
Ours MFH-LSRSIR	0.9866	0.9837	0.9878	0.9900

Figure 9 presents select retrieval results corresponding to different hashing methods on the SAT-6 dataset. When compared with existing alternative hashing methods, the proposed MFH-LSRSIR method achieves the state-of-the-art performance under various evaluation criteria. Experimental results show that our approach is well adapted to remote sensing images, and it has strong advantages in dealing with large-scale remote sensing image retrieval problems.
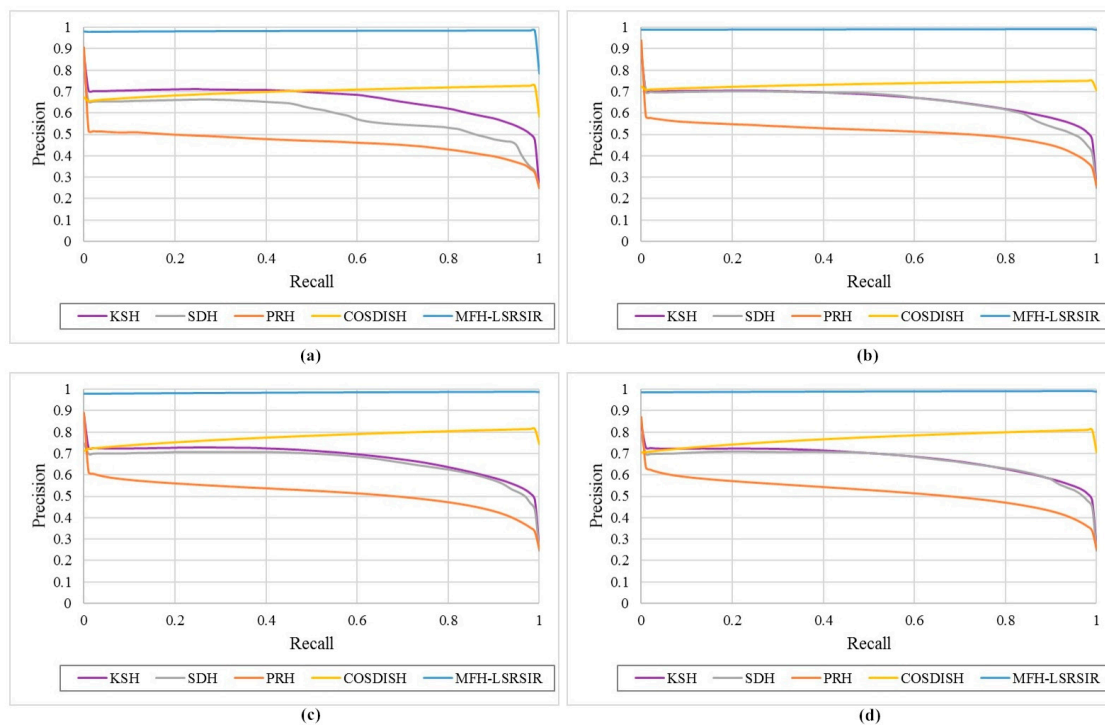
Figure 8. Compared precision-recall curves between different hashing methods with varied hash code lengths on the SAT-6 datasets: (**a**) 8 bits; (**b**) 16 bits; (**c**) 32 bits; and, (**d**) 64 bits.

The Query Image	(Building)								
MFH-LSRSIR									
CODISH									
SDH									
KSH									
PRH									
	1st	500th	1000th	1500th	2000th	2500th	3000th	3500th	4000th	4500th

Figure 9. Visualized retrieval results on the SAT-6 dataset by different hashing methods with 64-bit hash codes. The red rectangles indicate incorrect retrieval results. The abscissa represents the sequence number of the image.

4. Conclusions

In this paper, we propose a general multiple feature hashing learning framework for large-scale RS image retrieval, called MFH-LSRSIR. In order to achieve a comprehensive description of complex remote sensing images, we extract multiple features with different receptive fields by unsupervised multi-layer feature learning, which can fully mine the spectra and spatial context cues. On the hashing learning stage, MFH-LSRSIR utilizes the column sampling to learn the hybrid feature hash functions by iteration. Through comparing with the existing hashing method, the proposed approach can achieve a mean average precision up to 99.52% on the SAT-4 dataset, and 99.00% on the SAT-6 dataset. Therefore, our proposed MFH-LSRSIR is a competent method for large-scale remote sensing image retrieval.

The experiments performed on both the SAT-4 and SAT-6 datasets confirmed that the proposed MFH-LSRSIR framework is a simple but effective framework. In this work, the adopted large-scale remote sensing image datasets just have a small number of land cover categories. In order to fulfill the demand of real remote sensing retrieval tasks, we will exploit supervised deep networks, such as supervised deep hashing networks [37–40] to address the retrieval problem on more complex remote sensing image datasets in our future work. In addition, we will explore more applications of the proposed MFH-LSRSIR, such as hyperspectral remote sensing image classification [41], image matching [42–44], high-resolution remote sensing image built-up area detection [45], high-resolution remote sensing image urban villages detection [46], infrared target detection [47], and so forth.

Author Contributions: Dongjie Ye wrote the source code, performed the experiments, and wrote the original manuscript; Yansheng Li provided the idea for this study, supervised the work, and revised the manuscript; Chao Tao, Xunwei Xie, and Xiang Wang improved the manuscript, and contributed to the discussion of experimental results. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, L.; Zhong, H.; Ranjan, R.; Zomaya, A.; Liu, P. Estimating the statistical characteristics of remote sensing big data in the wavelet transform domain. *IEEE Trans. Emerg. Top. Comput.* **2014**, *2*, 324–337. [CrossRef]
2. Ma, Y.; Wu, H.; Wang, L.; Huang, B.; Ranjan, R.; Zomaya, A.; Jie, W. Remote sensing big data computing: Challenges and opportunities. *Future Gener. Comput. Syst.* **2015**, *51*, 47–60. [CrossRef]
3. Rathore, M.M.U.; Paul, A.; Ahmad, A.; Chen, B.W.; Huang, B.; Ji, W. Real-time big data analytical architecture for remote sensing application. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *8*, 4610–4621. [CrossRef]
4. Bosilj, P.; Aptoula, E.; Lefèvre, S.; Kijak, E. Retrieval of remote sensing images with pattern spectra descriptors. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 228. [CrossRef]
5. Sebai, H.; Kourgli, A.; Serir, A. Dual-tree complex wavelet transform applied on color descriptors for remote-sensed images retrieval. *J. Appl. Remote Sens.* **2015**, *9*. [CrossRef]
6. Bretschneider, T.; Cavet, R.; Kao, O. Retrieval of remotely sensed imagery using spectral information content. In Proceedings of the International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 24–28 June 2002; pp. 2253–2255.
7. Scott, G.J.; Klaric, M.N.; Davis, C.H.; Shyu, C.R. Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1603–1616. [CrossRef]
8. Ma, A.; Sethi, I.K. Local shape association based retrieval of infrared satellite images. In Proceedings of the IEEE International Symposium on Multimedia, Irvine, CA, USA, 12–14 December 2005; pp. 551–557.
9. Ferecatu, M.; Boujemaa, N. Interactive remote-sensing image retrieval using active relevance feedback. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 818–826. [CrossRef]
10. Hongyu, Y.; Bicheng, L.; Wen, C. Remote sensing imagery retrieval based-on gabor texture feature classification. In Proceedings of the International Conference on Signal Processing, Beijing, China, 31 August–4 September 2004; pp. 733–736.
11. Newsam, S.D.; Kamath, C. Retrieval using texture features in high-resolution multispectral satellite imagery. In Proceedings of the SPIE—The International Society for Optical Engineering, Orlando, FL, USA, 12 April 2004; pp. 21–32.
12. Samal, A.; Bhatia, S.; Vadlamani, P.; Marx, D. Searching satellite imagery with integrated measures. *Pattern Recognit.* **2009**, *42*, 2502–2513. [CrossRef]
13. Yang, Y.; Newsam, S. Geographic image retrieval using local invariant features. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 818–832. [CrossRef]
14. Zhou, W.; Shao, Z.; Diao, C.; Cheng, Q. High-resolution remote-sensing imagery retrieval using sparse features by auto-encoder. *Remote Sens. Lett.* **2015**, *6*, 775–783. [CrossRef]
15. Du, Z.; Li, X.; Lu, X. Local structure learning in high resolution remote sensing image retrieval. *Neurocomputing* **2016**, *207*, 813–822. [CrossRef]
16. Wang, Y.; Zhang, L.; Tong, X.; Zhang, L.; Zhang, Z.; Liu, H.; Xing, X.; Mathiopoulos, P.T. A three-layered graph-based learning approach for remote sensing image retrieval. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6020–6034. [CrossRef]
17. Chaudhuri, B.; Demir, B.; Bruzzone, L.; Chaudhuri, S. Region-based retrieval of remote sensing images using an unsupervised graph-theoretic approach. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 987–991. [CrossRef]
18. Li, Y.; Zhang, Y.; Tao, C.; Zhu, H. Content-based high-resolution remote sensing image retrieval via unsupervised feature learning and collaborative affinity metric fusion. *Remote Sens.* **2016**, *8*, 709. [CrossRef]
19. Li, Y.; Tao, C.; Tan, Y.; Shang, K.; Tian, J. Unsupervised multilayer feature learning for satellite image scene classification. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 157–161. [CrossRef]
20. Datar, M.; Immorlica, N.; Indyk, P.; Mirrokni, V.S. Locality-sensitive hashing scheme based on p-stable distributions. In Proceedings of the Twentieth Symposium on Computational Geometry, Brooklyn, NY, USA, 8–11 June 2004; pp. 253–262.
21. Weiss, Y.; Torralba, A.; Fergus, R. Spectral Hashing. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–11 December 2008; pp. 1753–1760.

22. Gong, Y.; Lazebnik, S.; Gordo, A.; Perronnin, F. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2916–2929. [CrossRef] [PubMed]

23. Kong, W.; Li, W.J. Isotropic Hashing. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1646–1654.

24. Liu, W.; Mu, C.; Kumar, S.; Chang, S.F. Discrete Graph Hashing. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–13 December 2014; pp. 3419–3427.

25. Norouzi, M.E.; Fleet, D.J. Minimal loss hashing for compact binary codes. In Proceedings of the International Conference on Machine Learning, Bellevue, WA, USA, 28 June–2 July 2011; pp. 353–360.

26. Liu, W.; Wang, J.; Ji, R.; Jiang, Y.G. Supervised hashing with kernels. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2074–2081.

27. Lin, G.; Shen, C.; Shi, Q.; Hengel, A.V.D.; Suter, D. Fast supervised hashing with decision trees for high-dimensional data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1971–1978.

28. Shen, F.; Shen, C.; Liu, W.; Shen, H.T. Supervised discrete hashing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 37–45.

29. Kang, W.C.; Li, W.J.; Zhou, Z.H. Column sampling based discrete supervised hashing. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 1230–1236.

30. Zhang, C.; Zheng, W. Semi-supervised Multi-view Discrete Hashing for Fast Image Search. *IEEE Trans. Image Process.* **2017**, *99*, 2604–2617. [CrossRef] [PubMed]

31. Wang, M.; Song, T. Remote sensing image retrieval by scene semantic matching. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2874–2886. [CrossRef]

32. Demir, B.; Bruzzone, L. Hashing-based scalable remote sensing image search and retrieval in large archives. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 892–904. [CrossRef]

33. Li, P.; Ren, P. Partial randomness hashing for large-scale remote sensing image retrieval. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 464–468. [CrossRef]

34. Han, X.-H.; Chen, Y.-W. Generalized aggregation of sparse coded multi-spectra for satellite scene classification. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 175. [CrossRef]

35. Yang, J.; Yang, J.Y.; Zhang, D.; Lu, J.F. Feature fusion: Parallel strategy vs. Serial strategy. *Pattern Recognit.* **2003**, *36*, 1369–1381. [CrossRef]

36. Basu, S.; Ganguly, S.; Mukhopadhyay, S.; Dibiano, R.; Karki, M.; Nemani, R. Deepsat: A learning framework for satellite imagery. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 3–6 November 2015; p. 37.

37. Li, Y.; Zhang, Y.; Huang, X.; Zhu, H.; Ma, J. Large-scale remote sensing image retrieval by deep hashing neural networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *PP*, 1–16. [CrossRef]

38. Gao, L.; Song, J.; Zou, F.; Zhang, D.; Shao, J. Scalable multimedia retrieval by deep learning hashing with relative similarity learning. In Proceedings of the ACM International Conference on Multimedia, Brisbane, Australia, 26–30 October 2015; pp. 903–906.

39. Zhu, H.; Long, M.; Wang, J.; Cao, Y. Deep hashing network for efficient similarity retrieval. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 2415–2421.

40. Li, W.J.; Wang, S.; Kang, W.C. Feature learning based deep supervised hashing with pairwise labels. In Proceedings of the International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 1711–1717.

41. Zhong, Z.; Fan, B.; Ding, K.; Li, H.; Xiang, S.; Pan, C. Efficient multiple feature fusion with hashing for hyperspectral imagery classification: A comparative study. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4461–4478. [CrossRef]

42. Ma, J.; Zhou, H.; Zhao, J.; Gao, Y.; Jiang, J.; Tian, J. Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6469–6481. [CrossRef]

43. Ma, J.; Zhao, J.; Tian, J.; Bai, X.; Tu, Z. Regularized vector field learning with sparse approximation for mismatch removal. *Pattern Recognit.* **2013**, *46*, 3519–3532. [CrossRef]

44. Ma, J.; Zhao, J.; Tian, J.; Yuille, A.; Tu, Z. Robust point matching via vector field consensus. *IEEE Trans. Image Process.* **2014**, *23*, 1706–1721.

45. Li, Y.; Tan, Y.; Deng, J.; Wen, Q.; Tian, J. Cauchy graph embedding optimization for built-up areas detection from high-resolution remote sensing images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2015**, *8*, 2078–2096. [CrossRef]

46. Li, Y.; Huang, X.; Liu, H. Unsupervised deep feature learning for urban village detection from high-resolution remote sensing images. *Photogramm. Eng. Remote Sens.* **2017**, *83*, 567–579. [CrossRef]

47. Li, Y.; Zhang, Y.; Yu, J.; Tan, Y.; Tian, J. A novel spatio-temporal saliency approach for robust DIM moving target detection from airborne infrared image sequences. *Inf. Sci.* **2016**, *369*, 548–563. [CrossRef]