Unsupervised Remote Sensing Image Retrieval using Probabilistic Latent Semantic Hashing

Ruben Fernandez-Beltran, Begüm Demir, *Senior Member, IEEE*, Filiberto Pla and Antonio Plaza, *Fellow, IEEE*

***Abstract*—Unsupervised hashing methods have attracted considerable attention in large-scale remote sensing (RS) image retrieval, due to their capability for massive data processing with significantly reduced storage and computation. Although existing unsupervised hashing methods are suitable for operational applications, they exhibit limitations when accurately modelling the complex semantic content present in RS images using binary codes (in an unsupervised manner). To address this problem, in this letter we introduce a novel unsupervised hashing method which takes advantages of the generative nature of probabilistic topic models to encapsulate the hidden semantic patterns of the data into the final binary representation. Specifically, we introduce a new probabilistic latent semantic hashing (pLSH) model to effectively learn the hash codes using three main steps: (A) data grouping, where the input RS archive is clustered into several groups; (B) topic computation, where the pLSH model is used to uncover highly descriptive hidden patterns from each group; and (C) hash code generation, where the data probability distributions are thresholded to generate the final binary codes. Our experimental results, obtained on two benchmark archives, reveal that the proposed method significantly outperforms state-of-the-art unsupervised hashing methods.**

***Index Terms*—Remote sensing, image retrieval, hash codes, unsupervised hashing, probabilistic topic models.**

I. Introduction

The fast development of satellite technologies has resulted in the availability of massive remote sensing (RS) image archives, which call for efficient and effective strategies for image search and retrieval. Traditional methods often exploit exact nearest neighbor search approaches, which exhaustively compare the query image with each image in the archive. This approach, which is also called *exhaustive linear scan*, is time-consuming and thus inappropriate for large-scale image search and retrieval problems. To overcome this issue, approximate nearest neighbor search strategies based on hashing techniques have been recently proven to be effective in order to reduce the cost of content-based image retrieval (CBIR) in terms of both processing time and storage requirements [1]. Hashing methods aim at learning hash functions that map the original high-dimensional image descriptors into low-dimensional binary codes, such that the similarity within the original image feature space can be well-preserved.

Hashing methods can be divided into two main categories: i) data-independent hashing methods; and ii) data dependent (also known as learning-based) hashing methods. Data-independent methods like Locality-Sensitive Hashing (LSH) [2] define hash functions by random projections that guarantee a high probability of collision for similar input images. Thus, they remain unaware of the data distribution, and require long codes to achieve a high retrieval performance [3], [1]. Data-dependent hashing methods can learn more compact binary codes by utilizing a set of data samples from the considered archive, and can be roughly divided into two sub-categories. The first one includes supervised hashing methods, in which supervised information (i.e., annotations of images) is necessary for learning the hash functions. For instance, Demir *et al.* presented, adapted to RS data properties and tested a supervised kernel-based hashing method [1], while Li *et al.* [4] introduced a deep hashing neural network (DHNN) to to address CBIR in RS. The DHNN jointly learns semantically accurate deep image features and binary hash codes by employing a high number of annotated images. Due to the use of such annotated images, supervised methods produce discriminative hash codes that satisfy the requirement of semantic similarity between the images. However, it is time consuming and expensive to obtain a sufficient number of high-quality annotated RS images, particularly for large-scale CBIR problems. The second sub-category comprises unsupervised hashing methods that do not require annotated images for learning hash codes.

In this letter, we focus on learning-based unsupervised hashing methods due to their relevance for operational RS image retrieval scenarios. In the RS community, there are few unsupervised hashing strategies available. As an example, the kernel-based unsupervised hashing (KULSH) [5] that defines hash functions for high-dimensional nonlinearly separable RS image descriptors was adopted for RS-based CBIR problems in [1]. 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 [3] 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.* propose in [6] a multi-code hashing method that initially

This work was supported by Generalitat Valenciana (APOSTD/2017/007), the Spanish Ministry of Economy (ESP2016-79503-C2-2-P, RTI2018-098651-B-C54, TIN2015-63646-C5-5-R), Junta de Extremadura (Ref. GR18060), the European Union under the H2020 EOXPOSURE project (No. 734541) and the European Research Council under the ERC Starting Grant BigEarth (No. 759764). *(Corresponding author: R. Fernandez-Beltran.)*

R. Fernandez-Beltran and F. Pla are with the Institute of New Imaging Technologies, University Jaume I, 12071 Castellón de la Plana, Spain. (e-mail: rufernan@uji.es, pla@uji.es). B. Demir is with the Faculty of Electrical Engineering and Computer Science, Technische Universitat Berlin, D-10587 Berlin, Germany (e-mail: demir@tu-berlin.de). A. Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain.(e-mail: aplaza@unex.es).

characterizes the images by using descriptors of primitive-sensitive clusters, and then constructs the multi-hash codes from these descriptors using the KULSH. It is worth noting that, unlike in the RS community, in the computer vision and multimedia communities the use of hashing is more extended and widely studied. The anchor graphs hashing (AGH) [7], the isotropic hashing (IsoH) method [8], the compressed hashing (CH) [9], the harmonious hashing (HamH) [10] and the density sensitive hashing (DSH) [11] methods are examples of widely used unsupervised hashing methods in that context.

Although the above-mentioned unsupervised hashing methods are relevant for operational applications, the hash codes produced by these algorithms might be not discriminative and descriptive enough to model the high-level semantic content of images under complex RS image retrieval tasks. To address this problem, this letter presents a new unsupervised hashing technique based on probabilistic topic models. These models [12] have been recently used to provide RS imagery with a higher level of semantic understanding [13]. This work takes advantage of the generative nature of probabilistic topic models to produce highly descriptive binary codes using the latent semantic patterns pervading the RS data. Specifically, we define a new topic model called probabilistic latent semantic hashing (pLSH) that learns the binary representation of RS images with complex semantic content. Our experiments on two benchmark archives demonstrate that the proposed method outperforms state-of-the-art unsupervised hashing methods presented in both the RS and computer vision communities.

II. Probabilistic Latent Semantic Hashing

Let $\mathbf{X} = \{\mathbf{X}_1, ..., \mathbf{X}_N\}$ be a complete RS archive with N images, which are characterized according to a specific K-dimensional original feature space, i.e. $\mathbf{X}_i = \{x_i^1, ..., x_i^K\}$. Let $\mathbf{\Lambda}^b = \{\mathbf{\Lambda}_1^b, ..., \mathbf{\Lambda}_N^b\}$ be the corresponding set of L-dimensional binary codes, where $\mathbf{\Lambda}_i^b \in \mathbb{Z}_2^L$ is the hash code of $\mathbf{X}_i$. Given a query image $\mathbf{X}_q$, the objective is to find the most similar images from the archive (using their hash codes) in order to reduce search time and storage requirements. The proposed unsupervised hashing method (Fig. 1) is defined based on the three steps described in the following subsections.

A. Data grouping

This step aims at dividing the RS archive into different groups, according to the similarities between the images in the original feature space. In order to achieve this goal, we cluster the initial RS image archive $\mathbf{X}$ into G disjoint groups, such that $\mathbf{X} = \cup_{i=1}^{G} \mathbf{C}_i$. On the one hand, this approach allows us to process the whole archive using a mini-batch scheme, where the number of images that are simultaneously handled can be substantially reduced. Note that this scheme provides important advantages in operational RS frameworks. On the other hand, it also pursues to detect certain data similarities in the original feature space, to make the topics highly specialized in terms of the hidden feature patterns associated to each group. Note that each group contains similar images in the original feature space, and thus the extracted topics can accurately characterize the semantic differences among potentially ambiguous images. In this letter, we use the well-known k-means clustering for grouping the images of the archive, whereas any clustering algorithm could be exploited.

B. Topic computation

The aim of this step is to extract the hidden feature patterns (topics) of each image group, and to represent the whole archive into a set of uncovered topics. To this end, we introduce the pLSH topic model (Fig. 1-B). This model has been designed to sequentially extract $Z = [L/G]$ topics from each group as $\Theta^* = \{\Theta_1^*, ..., \Theta_L^*\}$, and to represent the images into these topics as $\Lambda = \{\Lambda_1, ..., \Lambda_N\}$. The pLSH consists of three observable random variables d, z^* and w, one hidden random variable z, and one regularization parameter δ. In detail, d represents the images of a particular group; z^* denotes the topics extracted from the previous groups; w symbolizes the features of the original feature space, and z is the set of topics extracted from the current group. In addition, four directed connections relate images to topics and topics to features. Unlike the regular probabilistic latent semantic analysis (pLSA) [12], our model design is able to simultaneously express the whole RS image archive in terms of two different sets of hidden patterns, given by the diverging random variables z^* and z. In addition, the δ regularizer also promotes those topics that exhibit a significant contribution within the RS archive. Note that these two properties are key factors in hashing, since they allow us to disregard redundant feature patterns while encapsulating the most relevant semantic content through a sequential processing scheme.

Our pLSH estimates two conditional probability distributions: (i) $\theta \sim \{p(w|z)\}$, which represents the description of topics in features, and (ii) $\lambda \sim \{p(z^*|d), p(z|d)\}$, which denotes the description of images in observed and hidden topics. In this letter, we estimate the θ and λ distributions by maximizing the complete log-likelihood via the Expectation-Maximization (EM) algorithm [12]. Initially, we define the log-likelihood expression according to the z^* and z random variables. Then, we apply Jensen's inequality and insert three different Lagrange multipliers: i) two for maintaining the θ and λ within the probability simplex, and ii) another one for maximizing the Kullback-Leibler divergence between λ and the uniform distribution, weighted by δ. Note that the parameter δ acts as a sparsity regularizer to enhance the dominant topics in λ, because the smallest probability values in the topic space logically become uninformative in the context of a binary characterization. Finally, we compute the partial derivatives, set them to zero, and isolate the model conditional probability distributions to obtain the final expressions for the E-step (Eqs. (1)-(2)) and the M-step (Eqs. (3)-(5)),

$$p(z^*|w,d) = \frac{p(w|z^*)p(z^*|d)}{\sum_z p(w|z)p(z|d) + \sum_{z^*} p(w|z^*)p(z^*|d)}, \quad (1)$$

$$p(z|w,d) = \frac{p(w|z)p(z|d)}{\sum_z p(w|z)p(z|d) + \sum_{z^*} p(w|z^*)p(z^*|d)}, \quad (2)$$

$$p(w|z) = \frac{\sum_d n(w,d)p(d)p(z|w,d)}{\sum_w \sum_d n(w,d)p(d)p(z|w,d)}, \quad (3)$$

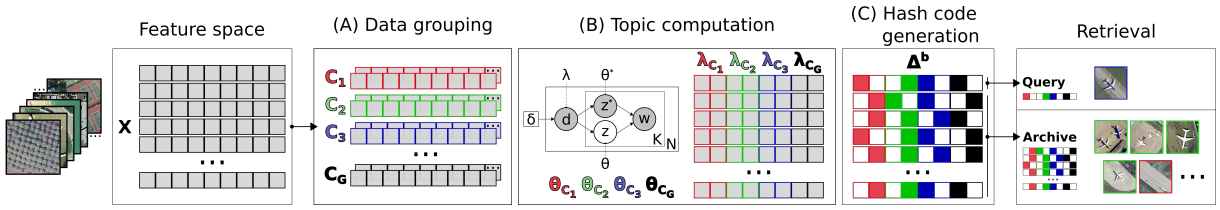

Fig. 1. Framework of the proposed unsupervised hashing method.

$$p(z^*|d) = \frac{\sum_w n(w,d)p(z^*|w,d) - \delta/Z^*}{\sum_{z^*}\sum_w n(w,d)p(z^*|w,d) + \sum_z \sum_w n(w,d)p(z|w,d)}, \quad (4)$$

$$p(z|d) = \frac{\sum_w n(w,d)p(z|w,d) - \delta/Z}{\sum_{z^*}\sum_w n(w,d)p(z^*|w,d) + \sum_z \sum_w n(w,d)p(z|w,d)}, \quad (5)$$

where $n(w_k, d_n)$ represents the number of times that the feature w_k appears in the image d_n, according to the original feature space. Note that, for the sake of simplicity, we omit in Eqs. (1)-(5) the summation indices of the random variables. Given an image group C_i in $n(w,d)$, a number of topics per group Z, a sparsity factor δ, and the set of previously extracted topics Θ^* in $p(w|z^*)$, the EM process is performed according to Algorithm 1 as follows. Initially, $p(w|z)$, $p(z^*|d)$ and $p(z|d)$ are randomly initialized. Then, the E-step (Eqs. (1)-(2)) and the M-step (Eqs. (3)-(5)) are alternated during I iterations.

This EM optimization is embedded into Algorithm 2 to extract the topics of the complete archive in $\Theta^* \in \mathbb{R}^{L\times K}$ and to represent all the images into this topic space in $\Lambda \in \mathbb{R}^{N\times L}$. From the image archive $\mathbf{X}$ (divided in G groups), the number of hash bits L and the sparsity factor δ, Algorithm 2 sequentially learns the set of observable topics (lines 2-6) and the representation of the archive into these topics (lines 7-11). Note that Θ^* is set to the void distribution for the first image group (line 1). In addition, $n(w|d)$ and $p(w|z^*)$ are fixed to the original feature representation of C_i and the distribution of previously observed topics Θ^* (lines 3 and 8). Finally, $p(w|z)$ is fixed to the zero distribution in the second loop of Algorithm 2 (line 9), to allow representing the whole image archive in the complete set of extracted topics. Note that the complexity of the topic computation algorithm is $\mathcal{O}(IKNZ)$.

Algorithm 1: EM-based optimization for the pLSH

input: $n(w,d)$, Z, δ, $p(w|z^*)$
output: $p(w|z)$, $p(z^*|d)$
1 $p(w|z)$, $p(z^*|d)$, $p(z^*|d) \Leftarrow$ Random initialization
2 **for** $i \leftarrow 1$ *to* I **do**
3 | **E-step**: $p(z^*|w,d)$, $p(z|w,d) \Leftarrow$ Eqs. (1)-(2)
4 | **M-step**: $p(w|z)$, $p(z^*|d)$, $p(z|d) \Leftarrow$ Eqs. (3)-(5)
5 **end**

C. Hash code generation

The aim of this step is to generate the final binary codes of the archive as $\mathbf{\Lambda}^b = \{\mathbf{\Lambda}^b_1, ..., \mathbf{\Lambda}^b_N\}$, where $\mathbf{\Lambda}^b_i \in \mathbb{Z}_2^L$ represents the binary code of $\mathbf{X}_i$. In order to achieve this objective, the previously uncovered probability distributions

Algorithm 2: Proposed topic computation step

input: $\mathbf{X}$, L, δ
output: Λ
1 $Z = [L/G]$, $\Theta^* = \emptyset$, $\Lambda = \emptyset$
2 **for** $i \leftarrow 1$ *to* G **do**
3 | $n(w,d) \leftarrow \mathbf{C}_i$, $p(w|z^*) \leftarrow \Theta^*$
4 | $p(w|z)$, $p(z^*|d) \Leftarrow$ **Algorithm 1**
5 | $\Theta^* \leftarrow \Theta^* \cup \{p(w|z^*)\}$
6 **end**
7 **for** $i \leftarrow 1$ *to* G **do**
8 | $n(w,d) \leftarrow \mathbf{C}_i$, $p(w|z^*) \leftarrow \Theta^*$
9 | $p(w|z)$, $p(z^*|d) \Leftarrow$ **Algorithm 1** $\langle p(w|z) \leftarrow$ Set to $0\rangle$
10 | $\Lambda \leftarrow \Lambda \cup \{p(z^*|d)\}$
11 **end**

$\Lambda = \{\Lambda_1, ..., \Lambda_N\}$, where $\Lambda_i = \{\Lambda_i^1, ..., \Lambda_i^L\}$, are thresholded as follows. Initially, we associate each topic to a particular hash bit (since the total number of extracted topics is L). Then, we compute the marginal probabilities of all the topics in Θ^*, as Eq. (6) shows. Finally, we obtain the hash code for a given image $\mathbf{X}_i$ by thresholding each element of its corresponding Λ_i distribution, according to the piece-wise function defined in Eq. (7). The target of this function is to transfer the most discriminating semantic patterns of $\mathbf{X}_i$ to its final binary characterization. Therefore, H only activates the hash bits associated to those topics that exhibit a significant contribution in $\mathbf{X_i}$ with respect to their occurrence frequency.

$$p(z^*) = \frac{\sum_d p(z^*|d)}{\sum_{z^*}\sum_d p(z^*|d)} \quad (6)$$

$$H(\Lambda_i^{z^*}) = \begin{cases} 1 & \text{if } \Lambda_i^{z^*} > p(z^*) \\ 0 & \text{else} \end{cases} \quad (7)$$

III. Dataset description and experimental design

In this work, the UCMerced [14] and the EuroSAT [15] datasets have been utilized because they are two important benchmark RS archives. On the one hand, the UCMerced archive contains 2100 RGB aerial images with a size of 256×256 pixels and a spatial resolution of 0.3 m. To evaluate the performance of the proposed method in the UCMerced archive, we have used the multi-label annotations of each image available at http://bigearth.eu/datasets. These annotations include 17 different semantic classes. Each UCMerced image is associated with a number of labels that varies between 1 and 7. To characterize the UCMerced archive, we have used a bag-of-visual-words (BOVW) representation of the local invariant features extracted by the scale-invariant feature

transform (SIFT). To obtain the BOVW representation of images, initially the images have been converted to gray-scale. Then, the SIFT descriptor of each image is obtained. Subsequently, the k-means clustering with $k = 512$ has been applied to 100 000 randomly selected SIFT descriptors. Finally, each image has been encoded as a histogram of visual words, normalized by the L2-norm.

On the other hand, the EuroSAT archive includes 27 000 Sentinel-2 images with a size of 64×64 pixels. In this work, we have used the RGB bands which have a spatial resolution of 10 m. The retrieval assessment in the EuroSAT has been conducted using the single-label annotations available at https://github.com/phelber/EuroSAT, which comprise 10 classes. The number of samples per class in this archive varies from 2000 to 3000. To characterize the EuroSAT archive, we have used the deep features extracted by the pretrained ResNet-18 convolutional neural network [15]. Specifically, the images have been initially scaled to $224 \times 224 \times 3$. Then, the ResNet-18 feature maps have been extracted and normalized by the softmax function to characterize each input image as a 1×512 feature vector. Note that the input feature space is independent from the presented unsupervised hashing appraoch.

The proposed method has been compared with seven state-of-the-art unsupervised hashing methods: the AGH [7], the CH [9], the DSH [11], the HamH [10], the IsoH [8], the KULSH [5] and the LSH [2]. These methods have been selected because they are relevant single-hash-code methods which have been also employed in other related works [1], [3]. The standard pLSA [12] has been also included as baseline. The experimental parameters have been set according to the suggestions made in the corresponding papers. In the case of the proposed method, we have considered a general configuration with $Z = 8$, $\delta = 1/Z^2$ and $I = 100$. The retrieval results are provided in terms of the average precision and recall metrics, obtained when considering the top-20 retrieved images in both archives. For the UCMerced, we select each image from the archive as a query, whereas 100 random image queries per class are selected for the EuroSAT. Moreover, five Monte Carlo runs have been conducted to obtain the average and standard deviation results and two different codelength values have been tested as $L = \{16, 32\}$.

IV. Experimental Results

A. Results: UCMerced

Table I reports the multi-label average precision and the average recall scores obtained by the proposed pLSH and the compared state-of-the-art unsupervised hashing methods for the UCMerced archive, with $L = 16$ and $L = 32$. From the table, one can see that the proposed pLSH provides the highest precision and recall compared to the other methods under both hash bits. In addition, the IsoH and the HamH achieve the second and third best average performance, respectively. When $L = 16$, the improvement of the pLSH with respect to the second best method is 6.24% in terms of average precision and 6.64% in terms of average recall. When $L = 32$, the pLSH outperforms the second best method by 3.51% in terms of precision and 2.05% in terms of recall. By analyzing the Table I in more detail, one can also see that standard deviations associated to the pLSH are always among the three lowest values, being all of them below 0.5%. These results show that the proposed method provides competitive advantages with respect to other state-of-the-art methods in the experiments with the UCMerced archive. Regarding the retrieval results of the methods used for comparison, the IsoH obtains (on average) the second best performance, followed by the HamH, the LSH, the CH, the KULSH, the DSH and the AGH. As an illustration, it is possible to see in Table I that the performance improvement of the IsoH with respect to the HamH is 1.92% in terms of precision, and 1.35% in terms recall (when $L = 32$). The average processing times of the three best methods (pLSH, IsoH and HamH) are 4.14s, 0.88s and 1.62s, respectively.

Fig. 2 shows an example of the images retrieved by the two best hashing methods, that are the proposed pLSH and the IsoH, for the UCMerced archive. Specifically, Fig. 2.(a) displays the selected query image, while Fig. 2.(b) and Fig. 2.(c) show the top-5 retrieved samples of the IsoH and the pLSH, respectively. Note that the retrieval order is given above the corresponding images, and the land-cover class labels associated to each image are given below such images. By analyzing

TABLE I
AVERAGE PRECISION, RECALL AND TIME OF THE PROPOSED METHOD AND SOME STATE-OF-THE-ART COMPETITORS FOR THE UCMERCED ARCHIVE.

Method	Precision (%)		Recall (%)		T(s)
	$L = 16$	$L = 32$	$L = 16$	$L = 32$	
AGH [7]	36.56 ± 2.63	53.78 ± 1.06	31.39 ± 3.54	50.14 ± 1.76	2.80
CH [9]	50.72 ± 0.38	55.73 ± 0.49	47.60 ± 0.78	54.48 ± 0.76	2.07
DSH [11]	40.71 ± 3.02	54.25 ± 0.76	38.18 ± 1.90	52.61 ± 0.62	0.59
HamH [10]	49.53 ± 0.57	57.29 ± 0.53	46.69 ± 0.91	56.69 ± 0.37	1.62
IsoH [8]	50.86 ± 0.63	59.21 ± 0.31	47.52 ± 0.67	58.04 ± 0.46	0.88
KULSH [5]	47.44 ± 0.84	51.97 ± 0.81	45.03 ± 0.98	51.42 ± 0.70	1.72
LSH [2]	49.57 ± 0.91	55.92 ± 0.29	48.13 ± 0.70	55.04 ± 0.42	0.50
pLSA [12]	50.14 ± 0.73	52.27 ± 0.87	47.87 ± 0.58	50.67 ± 1.29	7.57
Our pLSH	**57.10** ± 0.49	**62.72** ± 0.34	**54.77** ± 0.45	**60.09** ± 0.44	4.14

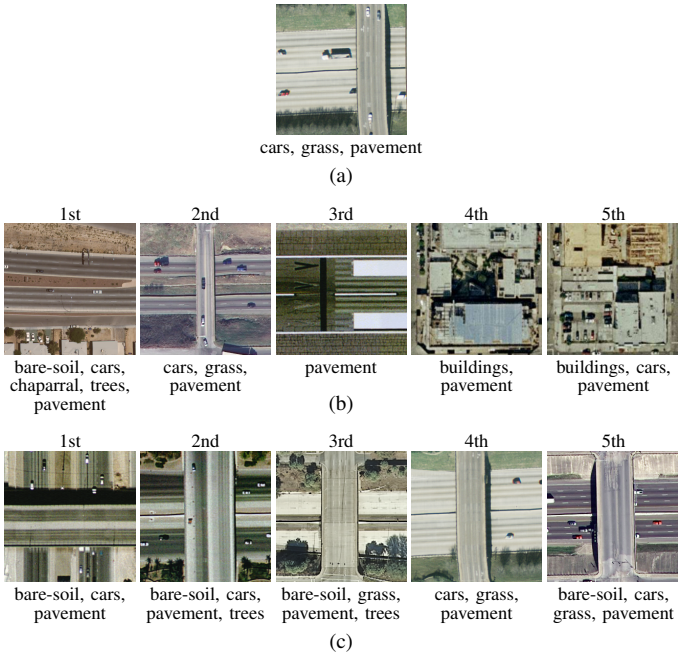

Fig. 2. UCMerced retrieval example when $L = 16$. (a) Query image. (b) Images retrieved by the IsoH. (c) Images retrieved by the proposed method.

TABLE II
AVERAGE PRECISION, RECALL AND TIME OF THE PROPOSED METHOD AND SOME STATE-OF-THE-ART COMPETITORS FOR THE EUROSAT ARCHIVE.

Method	Precision (%)		Recall (%)		T(s)
	$L = 16$	$L = 32$	$L = 16$	$L = 32$	
AGH [7]	24.62 ± 2.15	40.79 ± 0.92	0.17 ± 0.01	0.29 ± 0.01	20.05
CH [9]	38.80 ± 0.65	47.51 ± 0.23	0.28 ± 0.01	0.35 ± 0.00	9.65
DSH [11]	26.21 ± 2.16	45.71 ± 1.09	0.18 ± 0.02	0.33 ± 0.01	2.44
HamH [10]	28.49 ± 0.69	41.74 ± 0.70	0.20 ± 0.01	0.31 ± 0.00	6.14
IsoH [8]	43.19 ± 0.75	57.35 ± 0.69	0.31 ± 0.01	0.42 ± 0.01	3.36
KULSH [5]	34.03 ± 0.88	43.86 ± 0.35	0.24 ± 0.01	0.32 ± 0.00	14.73
LSH [2]	28.67 ± 2.86	45.94 ± 0.56	0.21 ± 0.02	0.34 ± 0.00	1.17
pLSA [12]	39.07 ± 0.97	45.95 ± 0.72	0.28 ± 0.01	0.33 ± 0.01	28.17
Our pLSH	**55.54** ± 0.57	**63.15** ± 0.36	**0.41** ± 0.00	**0.47** ± 0.00	13.85

these visual results, one can observe that the proposed method is able to retrieve images which are semantically more similar to the query. For instance, the fourth and fifth retrieved images by the pLSH contain cars, grass and pavement that are all present in the query image, whereas those images retrieved by the IsoH contain buildings, which are unrelated to the concept of the query. From Fig. 2, we can also observe that the proposed method tends to retrieve images containing land-cover classes that are more closely related to the query. As an example, the third retrieved image by the pLSH contains the bare-soil and the tree classes, which are not present in the query image but are certainly very related to the query classes: pavement and grass. The same behavior is observed in the retrieval results of many other query images.

B. Results: EuroSAT

Table II provides the single-label average precision and the average recall obtained in our experiments with the EuroSAT archive. From the table, one can see that the pLSH achieves the highest precision and recall scores with respect to the other hashing methods. In details, the IsoH and the CH obtain the second best and the third best performances, respectively. The improvement provided by the pLSH with respect to the IsoH is 9.07% in terms of precision and 0.07% in terms of recall. Additionally, the pLSH outperforms the CH by 16.19% in terms of precision and 0.12% in terms of recall. The average time of the three best methods (pLSH, IsoH and CH) is 13.85, 3.36s and 9.65s, respectively. The results confirm that the proposed method provides significant advantages for RS CBIR problems. Nevertheless, the considered EM-based optimization is a computationally demanding process and further research could be developed in this regard.

At this point, it is worth noting that the semantic intricacy of the complex semantic content present in RS images using binary hash codes becomes a major challenge in operational retrieval applications. In this sense, the obtained results show that the proposed method is suitable for operational RS image retrieval scenarios, where the images are expected to contain highly complex semantic content. Many existing unsupervised hashing methods try to characterize this semantic complexity using some sort of projection or clustering mechanism. For instance, the IsoH learns different projection functions that allow equal variance across the projected space dimensions. In this way, the number of bits used for each projection can be balanced with respect to the data variance in the original feature space. However, our obtained retrieval results show that the IsoH (as well as the other methods used for comparison) may provide a limited retrieval performance in RS problems, since they are unable to effectively extract and exploit the hidden relationships among different feature patterns given in RS imagery. In contrast, the proposed method aims at enhancing the semantic information carried by each single hash bit by taking advantage of generative semantic nature of probabilistic topic models. The obtained quantitative and qualitative results demonstrate the effectiveness of our newly developed unsupervised hashing approach, demonstrating its relevance for operational RS image retrieval scenarios.

V. CONCLUSIONS

In this letter, a novel unsupervised hashing method based on topic models has been presented for large-scale CBIR problems using RS imagery. Taking advantage of the generative semantic nature of probabilistic topic models, the proposed method defines a new model (pLSH) to learn the binary hash codes of images in large archives, in a fully unsupervised manner. The proposed method enables a detailed modeling of the semantic content of an image without requiring annotated images. Our experiments demonstrate the potential of topic models to extract and exploit semantic information present in RS images through binary codes. In the future, we plan to extend this work to deep probabilistic models, with additional experiments and efficient parallel implementations.

REFERENCES

[1] B. Demir and L. Bruzzone, "Hashing-based scalable remote sensing image search and retrieval in large archives," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 892–904, 2016.

[2] M. S. Charikar, "Similarity estimation techniques from rounding algorithms," in *Proc. ACM Sym. Theory Comp.*, 2002, pp. 380–388.

[3] P. Li and P. Ren, "Partial randomness hashing for large-scale remote sensing image retrieval," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 464–468, 2017.

[4] Y. Li, Y. Zhang, X. Huang, H. Zhu, and J. Ma, "Large-scale remote sensing image retrieval by deep hashing neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 950–965, 2017.

[5] B. Kulis and K. Grauman, "Kernelized locality-sensitive hashing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1092–1104, 2011.

[6] T. Reato, B. Demir, and L. Bruzzone, "An unsupervised multicode hashing method for accurate and scalable remote sensing image retrieval," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 276–280, 2019.

[7] W. Liu, J. Wang, S. Kumar, and S.-F. Chang, "Hashing with graphs," in *Proc. Int. Conf. Mach. Learn.* Omnipress, 2011, pp. 1–8.

[8] W. Kong and W.-J. Li, "Isotropic hashing," in *Adv. Neural. Inf. Process. Syst.*, 2012, pp. 1646–1654.

[9] Y. Lin, R. Jin, D. Cai, S. Yan, and X. Li, "Compressed hashing," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 446–451.

[10] B. Xu, J. Bu, Y. Lin, C. Chen, X. He, and D. Cai, "Harmonious hashing," in *International Joint Conference on Artificial Intelligence*, 2013.

[11] Z. Jin, C. Li, Y. Lin, and D. Cai, "Density sensitive hashing," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1362–1371, 2014.

[12] R. Fernandez-Beltran and F. Pla, "Prior-based probabilistic latent semantic analysis for multimedia retrieval," *Multimed Tools Appl.*, vol. 77, no. 13, pp. 16 771–16 793, 2018.

[13] R. Fernandez-Beltran, A. Plaza, J. Plaza, and F. Pla, "Hyperspectral unmixing based on dual-depth sparse probabilistic latent semantic analysis," *IEEE TGRS*, vol. 56, no. 11, pp. 6344–6360, 2018.

[14] Y. Yang and S. Newsam, "Geographic image retrieval using local invariant features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 818–832, 2012.

[15] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2019.