Computer Vision and Image Understanding 124 (2014) 12–21

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Semi-supervised multi-graph hashing for scalable similarity search

Jian Cheng [a,*], Cong Leng [a], Peng Li [a], Meng Wang [b], Hanqing Lu [a]

[a] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] School of Computer and Information, Hefei University of Technology, Hefei 230009, China

ARTICLE INFO

Article history:
Received 18 August 2013
Accepted 1 April 2014

Keywords:
Hashing
Multiple graph learning
Multiple modality
Semi-supervised learning

ABSTRACT

Due to the explosive growth of the multimedia contents in recent years, scalable similarity search has attracted considerable attention in many large-scale multimedia applications. Among the different similarity search approaches, hashing based approximate nearest neighbor (ANN) search has become very popular owing to its computational and storage efficiency. However, most of the existing hashing methods usually adopt a single modality or integrate multiple modalities simply without exploiting the effect of different features. To address the problem of learning compact hashing codes with multiple modality, we propose a semi-supervised Multi-Graph Hashing (MGH) framework in this paper. Different from the traditional methods, our approach can effectively integrate the multiple modalities with optimized weights in a multi-graph learning scheme. In this way, the effects of different modalities can be adaptively modulated. Besides, semi-supervised information is also incorporated into the unified framework and a sequential learning scheme is adopted to learn complementary hash functions. The proposed framework enables direct and fast handling for the query examples. Thus, the binary codes learned by our approach can be more effective for fast similarity search. Extensive experiments are conducted on two large public datasets to evaluate the performance of our approach and the results demonstrate that the proposed approach achieves promising results compared to the state-of-the-art methods.

1. Introduction

Recently, with the rapid advances of digit devices and the Internet, the amount of multimedia data (e.g. images, videos) are explosively increasing. These huge databases have posed a significant challenge in terms of scalable similarity search to many multimedia applications, such as content based multimedia retrieval (CBMR) [1], and classification.

In order to perform nearest neighbor search, many content based multimedia retrieval applications exhaustively compare the query with each sample in the database, which is infeasible for large-scale cases because the linear complexity is not scalable in practical situations. Besides, many large-scale CBMR applications suffer from the curse of dimensionality since feature descriptors usually have hundreds or even thousands of dimensions. Therefore, beyond the infeasibility of exhaustive search, storage of the original data also becomes a challenging problem. Fortunately, in CBMR and many other applications, finding approximate nearest neighbors is sufficient. To overcome these problems, a lot of research efforts have been devoted to investigate the alternative solution – approximate nearest neighbor (ANN) search, which trades off a little search accuracy to greatly speed up the search process.

Over the past decades, many advances have been made on ANN techniques for large-scale applications. In the earlier works, many tree-based methods [2–4] can perform similarity search effectively for low-dimensional data. However, for the high-dimensional cases and applications, their performance is not satisfactory and does not guarantee faster search compared to linear scan methods. Therefore, hashing techniques have been actively studied to provide efficient solutions for such high-dimensional data in recent years [5–7]. Hashing-based methods are promising in accelerating similarity search for their capability of generating compact binary codes for a large number of images in the dataset so that similar images will be mapped to close binary codes. Retrieving similar neighbors is then accomplished simply by finding the images that have codes with a small Hamming distance from the query, which is extremely fast to calculate. In addition, the binary codes representation can save much more storage space than the feature descriptors for large databases.

Many hashing approaches have been developed in recent years. A notable hashing method is Locality Sensitive Hashing (LSH) [5,6] which projects the data point to a random hyperplane and then conducts random thresholding. Inspired by spectral clustering,

* Corresponding author.
E-mail address: jcheng@nlpr.ia.ac.cn (J. Cheng).

http://dx.doi.org/10.1016/j.cviu.2014.04.001
1077-3142/

Spectral Hashing (SH) [7] and its extensions [8–10] attempt to apply machine learning approaches to find good data-aware hash functions. All of these hashing methods have shown success in scalable similarity search. However, most of the existing hashing methods only adopt single feature modality to learn the binary codes in hamming space. As we all know, the real-world images are often represented by multiple modalities, such as different kinds of visual features, and the importance of the roles that different modalities play can also be significantly different in hash function learning.

In recent years, more and more researchers are attracted to study how to fuse multiple modality to improve the performance of hashing algorithms [11]. Zhang et al. proposed a *Composite Hashing* (CHMS) [12]. In CHMS, one graph is established for each source and then combine them to learn linear hashing functions for each source. Song et al. presented a *Multiple Feature Hashing* (MFH) [13] to tackle the near-duplicate video retrieval problem. Liu et al. formulated the hashing problem as a similarity preserving hashing with linearly combined multiple kernels [14]. Xia et al. extended KLSH to Multi-Kernel Locality Sensitive Hashing (MKLSH) [15] by using multiple kernels. However, most of the multi-modal hashing methods which employ graph learning might suffer from the high complexity of large graph construction and most of them learn the hash functions in an unsupervised manner, which cannot effectively preserve the semantic similarity. There exist supervised hashing methods that can handle such semantic similarity but they are prone to overfitting when labeled data is small or noisy.

In this paper, we propose a semi-supervised Multi-Graph Hashing (MGH) approach for scalable similarity search. Different form the traditional methods, our approach can effectively integrate the multiple modalities with optimized weights in a multi-graph learning framework. In this way, the effects of different modalities can be adaptively modulated. It is worth to note that, in the existing multi-view (modality) based hashing method, [12,13] are also graph based, but they treat each graph equally. Besides, in our approach, anchor graph is employed for fast large graph construction and semi-supervised information is also incorporated into the unified learning scheme. Thus, the binary codes learned by our method can be more effective for fast similarity search. Fig. 1 illustrates the flowchart of our proposed hashing approach. The characteristics of our work can be summarized as follows:

(1) We propose a semi-supervised multi-graph hashing framework for binary code learning in scalable similarity search.
(2) The effects of different modalities can be adaptively modulated in our unified learning scheme. Therefore, our hashing approach can integrates different sources of information with optimized weights.

The remainder of this paper is organized as follows. In Section 2, we give a brief review of the related work. Section 3 describes our semi-supervised Multi-Graph Hashing (MGH) approach in detail. We conduct extensive experiments on two large publicly available datasets in Section 4. Finally, conclusion and future work are given in Section 5.

2. Related Work

There has been extensive research on fast similarity search due to its critical importance in many information retrieval and computer vision fields. For a low-dimensional feature space, similarity search can be carried out efficiently by some tree-based methods [2–4], e.g., KD-tree, M-tree. They usually partition the data space recursively to implement an exact similarity search in the low-dimension feature space. However, the high dimensionality will significantly corrupt the efficiency of tree-based methods, and make them even perform worse than the naive methods (e.g. linear scan) [16]. Thus, they will encounter difficulties in practical applications where the number of dimensions may be hundreds or even thousands. To overcome this issue, hashing-based approximate nearest neighbor (ANN) search approaches have become more and more popular recently. By encoding the high-dimensional data points into binary codes in a low-dimensional Hamming space, hashing methods provide compact data representation and efficient indexing mechanism for scalable similarity search. In general, we summarize them into two categories, single-view hashing and multi-view hashing.

2.1. Single-view hashing

Most of the existing hashing work utilize a single view to generate binary codes. They can be broadly categorized into two groups: data-independent and data-dependent schemes.

One of the most well known data-independent methods is Locality Sensitive Hashing (LSH) [5,6], which utilizes a batch of locality sensitive functions to embed the data into Hamming space. The LSH algorithm typically guarantees the high probability for any two similar samples falling into the same bucket and two dissimilar samples into the different buckets. One popular scheme in LSH is to generate a batch of random vectors that preserves inner-product similarity from a particular probabilistic distribution [17]. LSH has been extended to several similarity measures, such as, p-norm distances for $p \in (0, 2]$ [18], the Mahalanobis distance [19,20] and the kernel similarity [21]. However, since the random vector is data-independent, LSH may lead to quite inefficient codes in practice as it requires multiple tables with long

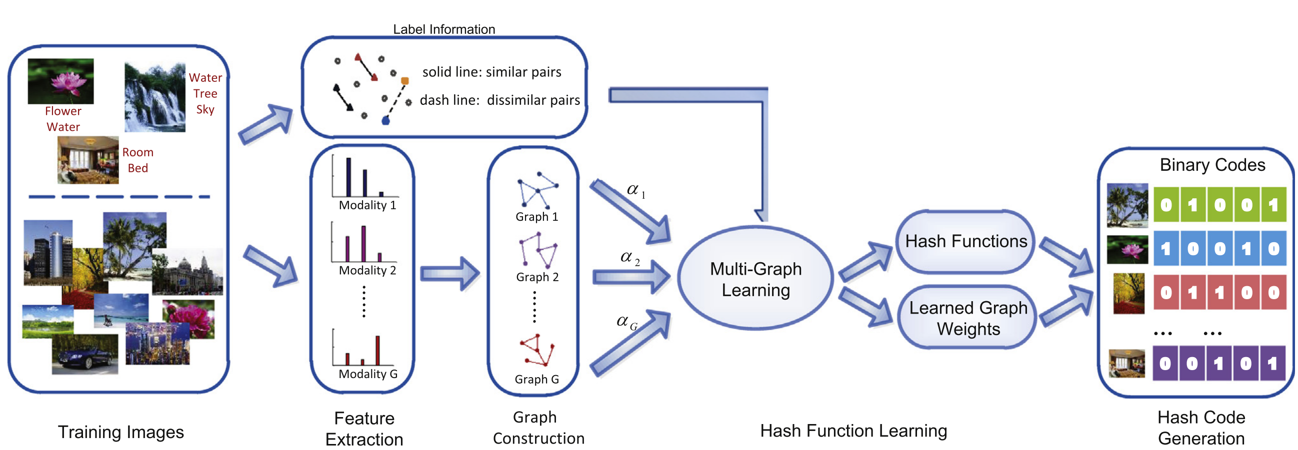

Fig. 1. The flowchart of our proposed semi-supervised Multi-Graph Hashing (MGH) approach.