Integration of semantic and visual hashing for image retrieval

Songhao Zhu, Dongliang Jin, Zhiwei Liang, Qiang Wang, Yajie Sun, Guozheng Xu

PII: S1047-3203(16)30172-9
DOI: http://dx.doi.org/10.1016/j.jvcir.2016.08.013
Reference: YJVCI 1842

To appear in: *J. Vis. Commun. Image R.*

Received Date: 1 March 2016
Revised Date: 4 July 2016
Accepted Date: 20 August 2016

Please cite this article as: S. Zhu, D. Jin, Z. Liang, Q. Wang, Y. Sun, G. Xu, Integration of semantic and visual hashing for image retrieval, *J. Vis. Commun. Image R.* (2016), doi: http://dx.doi.org/10.1016/j.jvcir.2016.08.013

INTEGRATION OF SEMANTIC AND VISUAL HASHING FOR IMAGE RETRIEVAL

Songhao Zhu[12], *Dongliang Jin*[1*], *Zhiwei Liang*[1], *Qiang Wang*[1], *Yajie Sun*[2], *Guozheng Xu*[1]

[1]School of Automation, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China
[2]Nanjing University of Information Science and Technology, Nanjing, 210044, China
njuptzsl@yeah.net

ABSTRACT

With the rapid proliferation of large-scale web images, recent years have witnessed more and more images labeled with user-provided tags, which leads to considerable effort made on hashing based image retrieval in huge databases. Current research efforts focus mostly on learning semantic hashing functions which design compact binary codes to map semantically similar images into similar codes; however the visual similarity is not well explored for constructing semantic hashing functions. Here a novel approach is proposed to learn hashing functions that preserve semantic and visual similarity between images. Specifically, semantic hashing codes are first learned by leveraging the similarity between textual structure and visual structure; then, the maximum entropy principle is exploited to achieve compact binary codes; finally, the function decay principle is introduced to remove noisy visual attributes. Experimental results conducted on a widely-used image dataset demonstrate the superior performance of the proposed method over the examined state-of-the-art techniques.

Index Terms—Image Retrieval, Semantic Similarity, Visual Structure, Hashing Code

1. INTRODUCTION

Due to the rapid advance of the Internet and digital cameras, many community-contributed image websites, such as Flickr[1], Zooomr[2], Picasa[3] and YouTube[4], facilitate users to upload and share their images for various purposes. This makes rapid growing scale of massive image collections with rich user-provided metadata available online. Therefore, how to retrieve visually relevant images effectively and efficiently from large scale databases has attracted more and more attentions over the past decades, and is increasingly becoming a challenging but urgent issue in multiple retrieval communities.

Over the past decades, numerous techniques have been proposed to retrieve the most similar patterns with respect to a given pattern in a high-dimensional feature space [5-13]. Recently, binary hashing learning has attracted considerable attention in computer vision, information retrieval, and data mining due to the computational complexity and storage efficiency of binary hashing codes [14-17]. The key principle in devising hashing methods is to encode high-dimensional image data to compact binary codes in the Hamming space, while maintaining some aspects of the structure of the original data as much as possible, such as the metric similarity in the feature space or the semantic similarity in form of labels. Having done so, one can perform efficient similarity search in the generated low-dimensional code space by simply calculating the Hamming distance between compact binary codes.

Early hashing methods are data-independent, such as kernelized locality-sensitive hashing[15], locality sensitive Hashing[18], min-wise Hashing[19], which use random projections as the hashing functions without taking the data distribution into consideration. Therefore, these randomized hashing methods usually require long binary codes to achieve high retrieval accuracy and have inferior search accuracy for large-scale image search [20].

To deal with this issue, data-dependent hashing techniques have been proposed to learn hashing functions, where these techniques can be categorized into the following two types: unsupervised and supervised (including semi-supervised) methods. On the one hand, unsupervised methods aim to learn hashing functions by taking into account data distribution information, such as spectral Hashing[21], isotropic Hashing[22], iterative quantization[23], spherical Hashing[24]. On the other hand, supervised methods aim to learn hashing functions by exploring supervisory semantic information, such as supervised Hashing[20], binary reconstructive embedding[25], semi-supervised Hashing[26], multi-Index Hashing[27], weighted component Hashing[28].

Although various data-dependent hashing approaches have shown their effectiveness and efficiency for large-scale image retrieval tasks, it is argued that few of them can be applied to deal with the retrieval of images with user-uploaded labels. That is, existing data-dependent hashing approaches seldom take the content of images into account, which indeed has a great effect on the improvement of the performance of image retrieval. Therefore, it is necessary to combine visual representation into semantic representation for learning more effective and efficient hashing codes.

To this end, in this paper a novel data-dependent hashing approach is proposed by taking into account the semantic representation with high-level labels and the visual representation with low-level features, as illustrated in