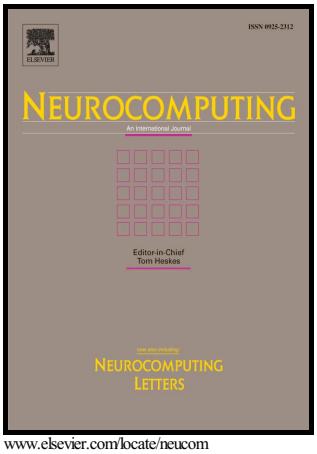

www.elsevier.com/locate/neucom

Binary representation learning in computer vision

Fumin Shen, Hanwang Zhang, Yang Yang, Chunhua Shen

PII: S0925-2312(16)30693-2
DOI: http://dx.doi.org/10.1016/j.neucom.2016.06.053
Reference: NEUCOM17305

To appear in: *Neurocomputing*
Accepted date: 11

Cite this article as: Fumin Shen, Hanwang Zhang, Yang Yang and Chunhua Shen, Binary representation learning in computer vision, *Neurocomputing* http://dx.doi.org/10.1016/j.neucom.2016.06.053

Binary Representation Learning in Computer Vision

Fumin Shen[a,*], Hanwang Zhang[b], Yang Yang[a], Chunhua Shen[c]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology, Chengdu, PR China
[b] School of Computing, National University of Singapore, Singapore
[c] School of Computer Science, The University of Adelaide, Australia

Keywords: Binary representation learning, hashing, computer vision

The volume of data has been dramatically enlarged in the big data era. The traditional representation of data or feature learning algorithms may not work well or be computationally tractable for large-scale applications, such as image retrieval, object recognition, etc. It is desirable to develop new, efficient data representation or feature learning/indexing techniques, which can be easily performed with big data and achieve promising performance in the related tasks. In most recent years, the data-dependent hashing or compact binary code learning techniques have attracted broad research interests in computer vision and related areas, due to the high efficiency of storage and pairwise comparison with the Hamming distance. Benefiting from the nature of binary codes, these methods can well help perform various vision tasks (e.g., retrieval, classification), especially the ones with large-scale data. Recently, the hashing techniques have been shown to achieve promising performance in various applications in computer vision, such as image retrieval, object recognition and classifier training.

This special issue focuses on the most recent progress on binary representation learning or hashing methods [1, 2, 3, 6, 18, 20] for various visual tasks with large-scale data, such as content-based image/video classification/retrieval/detection/tracking [4, 8, 10, 14, 16, 18], image/video annotation [12], multimedia processing [5, 7, 13, 15] and visual semantic analysis [8, 16]. This special issue will also target on related fast feature extraction or representation learning techniques [11, 17, 19], which can well handle large-scale visual tasks. The primary objective of this special issue fosters focused attention on the latest research progress in this interesting area.

On these interesting topics, totally 30 submissions have been received. After two or three rounds of reviews, 20 papers were accepted for publication. The introductions of the accepted papers are summarized as follows:

The paper by Dai et al. [1] focused on encoding high-dimensional features. In this paper, the authors introduced a novel method called Bayesian Hashing, which learned an optimal Hamming embedding to encode high-dimensional features to binary bits, and discussed its application to the challenging problem of face recognition. The learned hashing representation was modeled with a well-designed supervised Bayesian learning framework, which consisted of three ingredients. First, local bit correlations were elaborately modeled using Naive Bayesian model (FERN), which was boosted to obtain a classifier for the hashing bit stream. Second, without incurring additional storage cost, the authors are imposed hashing bit-stream permutations to obtain a series of classifiers, which could achieve better performance. Third, the sequential forward floating search (SFFS) algorithm was introduced to perform model selection on multiple-permutation models, gaining further performance improvement. Extensive evaluations and comparative studies were conducted, which demonstrate that the proposed approach gives superior performance in both accuracy and speed. State-of-the-art results were achieved on several well-known face recognition benchmarks.

In the paper by Shen et al. [2], the authors proposed a novel hashing method, named Semi-paired Hashing (SPH), to deal with a more challenging cross-view retrieval task, where only partial pairwise correspondences were provided in advance. Specifically, SPH aimed to preserve within-view similarity and cross-view correlation among multi-view data. Similarity structures within each view were obtained via anchor graph. As limited samples were paired, correlations between unpaired samples were exploited via a simple yet effective approach, which estimated cross-view correlations by partial cross-view pairwise information and within-view similarity structure. Besides, the authors further incorporated two regression terms between original features and target binary codes to reduce the quantization loss. An efficient iterative algorithm was presented to simultaneously solve hash functions and binary codes. Extensive experiments on two benchmark datasets demonstrated the superiority of SPH over the state-of-the-art methods, especially in the semi-paired scenarios.

The paper by Zhang et al. [3] proposed a surprisingly simple method to solve the Approximate Nearest Neighbor search (ANN) problem with high accuracy results and requiring only a limited number of random I/O. ANN is one of the most frequently used and yet expensive operations in the high-dimensional database, especially the multimedia database involving massive high-dimensional feature vectors. Recently,

*Corresponding author. Tel.: +86 28 61831783.
Email addresses: fumin.shen@gmail.com (Fumin Shen), hanwangzhang@gmail.com (Hanwang Zhang), dlyyang@gmail.com (Yang Yang), chunhua.shen@adelaide.edu.au (Chunhua Shen)