Pattern Recognition 48 (2015) 3160–3168

Contents lists available at ScienceDirect

Pattern Recognition

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

Learning predictable binary codes for face indexing

Ran He [a,*], Yinghao Cai [b], Tieniu Tan [a], Larry Davis [c]

[a] The Center for Research on Intelligent Perception and Computing and the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, No. 95 ZhongGuanCun East St, HaiDian District, Beijing 100190, China
[b] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[c] The Institute for Advanced Computer Studies and the Department of Computer Science, University of Maryland, College Park, MD 20742, United States

ARTICLE INFO

Article history:
Received 26 September 2014
Received in revised form
29 January 2015
Accepted 17 March 2015
Available online 1 April 2015

Keywords:
Binary codes
Hashing
Face index
Large scale
Feature learning

ABSTRACT

High dimensional dense features have been shown to be useful for face recognition, but result in high query time when searching a large-scale face database. Hence binary codes are often used to obtain fast query speeds as well as reduce storage requirements. However, binary codes for face features can become unstable and unpredictable due to face variations induced by pose, expression and illumination. This paper proposes a predictable hash code algorithm to map face samples in the original feature space to Hamming space. First, we discuss the 'predictability' of hash codes for face indexing. Second, we formulate the predictable hash coding problem as a non-convex combinatorial optimization problem, in which the distance between codes for samples from the same class is minimized while the distance between codes for samples from different classes is maximized. An Expectation Maximization method is introduced to iteratively find a sparse and predictable linear mapping. Lastly, a deep feature representation is learned to further enhance the predictability of binary codes. Experimental results on three commonly used face databases demonstrate the superiority of our predictable hash coding algorithm on large-scale problems.

1. Introduction

A face recognition system aims to identify or verify one person by comparing an input facial image with registered images in a database. Computational requirements are aggravated by the high dimensionality of discriminative features [1] as well as the large number of registered people [2]. Many strategies, such as social network context [3], two-stage strategies [4], and cascade structures [5], have been applied to speed up training or searching in a large-scale face database.

Recently, hashing methods have drawn attention in large-scale image retrieval and face recognition, where the terminology 'hashing' refers to learning compact binary codes with Hamming distance computation. For image retrieval, similarity-sensitive hashing or locality-sensitive hashing algorithms [6–11], support vector machine [12,13], decision trees [14] and deep learning [15,16] have been studied to map high-dimensional data into a similarity-preserving low-dimensional Hamming space. Jegou et al. [17] used Hamming embedding to replace vector quantization in bag-of-feature construction. Wang et al. [18,19] introduced sequential projection learning and semi-supervised learning for hashing with compact codes. Biswas et al. [20] developed an efficient and robust algorithm to map shape features to a hash table. To discover or preserve the neighborhood structure in the data for compact codes, Liu et al. [21] and Kong et al. [22] presented a graph-based hashing method and a Manhattan Hashing method respectively. Based on similar and dissimilar data pairs, Gong and Lazebnik [23] developed an iterative procrustean approach to learning binary codes, and Liu et al. [24] further proposed a kernel-based supervised hashing model. In LDAHash, Strecha et al. [25] performed linear discriminant analysis (LDA) or difference of covariances on the descriptors before binarization. And for multi-view or cross-view retrieval, deep multi-view hashing [26], predictable dual-view hashing [27], co-regularized hashing [28], and collective matrix factorization hashing [29] were developed. For a brief review of binary hash codes for large-scale image search, refer to [16,27].

For face recognition, Ngo et al. [30] discretized the PCA coefficients of a face image to binary codes by using a bit-extraction method. In BioHashing methods [31,32], randomized dimension reduction or optimal linear transformations are generated to calculate the dot product of test features. Zeng et al. [33] addressed the hashing problem of high dimensional SIFT vectors based on the p-stable distribution locality sensitive hashing scheme. Shi et al. [34] built a connection between hashing kernels and compressed sensing, and applied hashing to speed up sparse representation based face recognition. Then Yan et al. [35] made use of a group of hashing function to learn similarity binary codes. Sattar et al. [36] proposed to use the 2-D discrete cosine transform and K-means clustering to learn hash codes. In addition, Wu et al.

* Corresponding author. Tel.: +86 10 82544641; fax: +86 10 82544485.
E-mail address: rhe@nlpr.ia.ac.cn (R. He).

http://dx.doi.org/10.1016/j.patcog.2015.03.016
0031-3203/

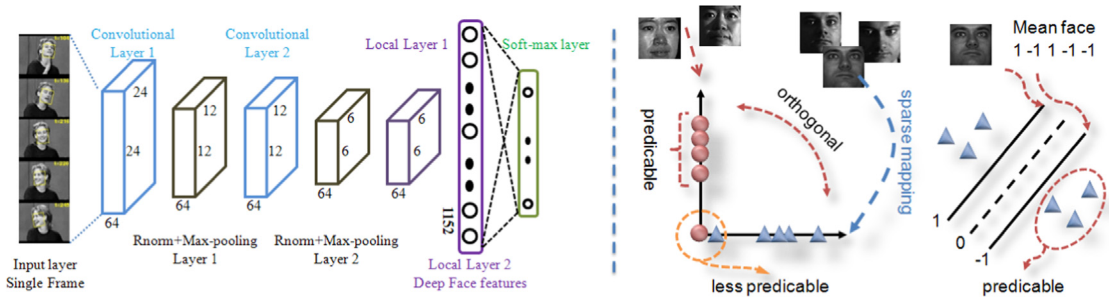

Fig. 1. The proposed scheme to learn predictable binary codes. 'Predictability' indicates that hash codes are predictable across facial variations. Our basic idea (right figure) is that the facial codes from one person are similar to the code of the mean face of this person with a predictable margin; meanwhile the codes of mean faces from different persons are significantly different and tend to be orthogonal (or low correlation [9,27]) to one another. Convolutional neural networks (left figure) are adopted to learn deep face features to improve the predictability of binary codes.

[2] and Chen et al. [37] resorted to hash codes for inverted indexing to speed up scalable face image retrieval.

Learning binary hash codes has been a key step to facilitate face recognition or image retrieval. And all the above hashing related methods indeed speed up retrieval or searching time. However, for face recognition, the hash codes learned from facial features tend to be unstable and unpredictable due to face variations induced by pose, expression and illumination. When one applies hashing methods to encode high-dimensional facial features, the learned codes should be predictable[1] across facial variations. This is due to the fact that two facial images and their corresponding binary codes from one person will generally not be the same. In addition, to the best of our knowledge, although several hashing methods have been used for face recognition, the fundamental question of what type of binary hash codes is good for face recognition has not been addressed.

This paper discusses the application of predictable hash codes to face indexing, and proposes a predictable hash code (PHC) learning scheme to embed high-dimensional dense facial features into Hamming space. First, based on the concept of linear discriminant analysis (or Fisherfaces in face recognition), we discuss the 'predictability' of hash codes for face indexing. Second, as illustrated in Fig. 1, we require that the distance between the codes from the same person (within-class distance) is minimized while the distance between the codes (between-class distance) from different classes is maximized. To achieve this goal, we relax the notion of within-class distance and between-class distance to the similarity of codes for images of the same person to the code of the mean face of that person and orthogonality (or low correlation [9,27]) of the codes of mean faces of different people, respectively. This allows us to formulate the predictable hash coding problem as a non-convex combinatorial optimization problem, which can be solved with an Expectation Maximization (EM) method to iteratively find a sparse and predictable linear mapping. Lastly, to further enhance the predictability of binary codes in real-world scenarios, convolutional neural networks (CNN) are adopted to learn a deep face representation. Experimental results on three commonly used face recognition databases demonstrate the superiority of our predictable hash coding algorithms on large-scale face indexing problems (the number of comparisons is larger than 900 million). Particularly on the YouTube Celebrities dataset, our proposed algorithms only use a 128-bits representation to achieve state-of-the-art results.

The rest of this paper is organized as follows. We discuss the 'predictability' of hash codes and present our predictable hash coding (PHC) algorithm in Section 2. Section 3 provides a series of experiments to validate our PHC algorithm, prior to summary in Section 4.

[1] Predicability indicates that maximum within-class Hamming distance H_w is smaller than minimum between-class Hamming distance H_b.

Table 1
Important notations used in this paper.

Notations	Descriptions
d	The number of features
n	The number of total training samples
C	The number of classes
k	The length of hash codes
X	The data matrix $X=[x_1,\ldots,x_n]\in R^{d\times n}$
M	The mean face matrix $M\in R^{d\times c}$
H_w	Maximum within-class Hamming distance
H_b	Minimum between-class Hamming distance
H_m	Margin of codes H_b-H_w

2. Predictable hash codes

Unsupervised and supervised dimensionality reduction methods have been widely used in face recognition. But many methods do not scale to large datasets because their complexity is quadratic (or worse) in the number of data points [16]. Hence hashing has been used to improve query speeds and reduce storage costs. However, previous hashing based methods [30–32,25] often treat dimensionality reduction and hashing as two independent steps, which makes the learned binary codes less discriminative. In this section, we apply the concept of dimensionality reduction into Hamming space and study predictable hash codes for face indexing.

2.1. Problem formulation

Table 1 summarizes the notation needed to present the method. The generic learning problem of dimensionality reduction for face recognition is formulated as follows. Consider a dataset X from C classes, which consists of n samples x_i $(1\le i\le n)$ in a high-dimensional Euclidean space R^d. Each class has n_c samples with that set denoted as X^c. That is $X=[X^1,\ldots,X^C]=[x_1,\ldots,x_n]$. Let matrix M contain mean faces m_c. That is $M=[m_1,\ldots,m_C]$. A dimensionality reduction method aims to learn a linear or non-linear mapping (or projection) matrix $W\in R^{d\times k}$ to project samples into a low-dimensional Euclidean space R^k.

One of the most widely used dimensionality reduction methods in face recognition is linear discriminant analysis (LDA). It maximizes a loss function that encourages a large separation between the projected class means while also encouraging a small variance within each class. Inspired by LDA, we define a within-class distance and a between-class distance for binary codes in Hamming space. H_w denotes the maximum within-class Hamming distance between any two codes for samples from the same class. H_b denotes the minimum between-class Hamming distance between any two codes from different classes. Given H_w and H_b, the margin of binary