Neurocomputing 173 (2016) 181–195

Contents lists available at ScienceDirect

Neurocomputing

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

Dimensionality reduction for histogram features: A distance-adaptive approach

Miaomiao Cui, Jiangtao Cui, Hui Li*

School of Computer Science and Technology, Xidian University, #161 postbox, South Taibai Road #2, Xi'an, Shaanxi 710071, PR China

ARTICLE INFO

Article history:
Received 30 June 2014
Received in revised form
15 March 2015
Accepted 28 March 2015
Available online 3 September 2015

Keywords:
Histogram features
Dimensionality reduction
MDS
Similarity measure

ABSTRACT

Histogram representations of visual features, such as high dimensional Bag-of-Features (BOF) and Spatial Pyramid Matching (SPM) representation, have been widely studied and adopted in image classification and retrieval due to their simplicity and performance. Problems involving high dimensional feature vectors usually require much computational cost and huge storage space. Moreover, it may additionally suffer from low accuracy because of the noise in data. In this paper we propose a novel distance-adaptive dimensionality reduction framework, namely generalized Multidimensional Scaling, with linear coding time to create compact and discriminative BOF or SPM representations. Comparing with traditional MDS, our approach exhibits two advantages, on one hand it is adaptive to many measures; on the other hand, it is able to map arbitrary query points into the new space. Exhaustive experimental results show that a very low dimension of BOF or SPM is sufficient for the retrieval task without losing accuracy. Comparatively, the state-of-the-art methods cannot achieve high accuracy on the low dimension. Aside from image retrieval task, we also show that our approach is much more effective than the original histogram representations when applied in image classification task.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Image relevant applications [1–3], such as object retrieval, scene recognition, etc., have been widely studied during last decade. The problem can be generally described as finding the images in a database that are the most similar to a query image. The significant growth in both the number and size of digital images and video collections on the web has brought in imperious demands for more powerful image retrieval tools. In recent years, many solutions have been proposed to improve the quality of search result. In particular, a sustained line of research has been initiated by the histogram features such as Bag-of-Features representation (BOF) [4], which have been justified to be effective in datasets with up to millions of images [5]. The general process are as follows. Firstly, an image is represented by a collection of local descriptors such as Scale-Invariant Feature Transform (SIFT) [6]. Then these descriptors are aggregated into a single histogram representation, which collects the statistics of so-called "visual words". Additionally, BOF has been further extended to Spatial Pyramid Matching (SPM) [7], which is a spatial extension of an orderless BOF image representation.

However, aforementioned process may not be applicable in recent applications where the number of images to perform query may exceed tens of millions. In that case, to well approximate the distribution of visual words in an image, the BOF representation is chosen to be high dimensional, probably up to more than thousands of dimensions [8,9]. Consequently, the corresponding histogram representation of BOF is very sparse. Similar phenomenon also exists in SPM [7], where 2100–4200 is set as the range for optimal number of dimensions. Philbin et al. reported a memory usage of 4.3 G for approximately 1.1 M images [10], which indicates that each BOF vector needs huge storage space. Moreover, the computational efficiency of BOF in large scale datasets will succumb to the well-known "curse of dimensionality". To summarize, there are two factors limiting the number of images that can be indexed in practice: the efficiency of the search process and the memory required to represent an image.

To solve the problems above, dimensionality reduction techniques have been proposed to generate compact and discriminative histogram representations. Among them, linear algorithms are convenient to create a kernel to project the original points into the new low-dimensional space, including Principal Component Analysis (PCA)[11] and Linear Discrimination Analysis (LDA) [12]. While histograms can be viewed as points on a statistical manifold, it is typically concentrated on a lower dimensional manifold of the measurement space. The study of these low dimensional manifolds has led to a specific research topic of

* Corresponding author.
E-mail address: hli@xidian.edu.cn (H. Li).

http://dx.doi.org/10.1016/j.neucom.2015.03.123
0925-2312/© 2015 Elsevier B.V. All rights reserved.

machine learning, namely manifold learning, based on which series of nonlinear dimensionality reduction algorithms have been proposed, including Isomap [13], Local Linear Embedding (LLE) [14], and Laplacian Eigenmaps (LE) [15]. However, none of these approaches are specifically designed for histogram features, which is a representative method for image description and has shown its success in both image classification and retrieval. To the best of our knowledge, histogram features have not been thoroughly studied in the aspect of dimensionality reduction.

MDS has been one of the most popular approaches to reduce dimensionality [16], it is used as a way to represent perceived similarities between a pair of stimuli by minimizing the ratio of differences between inter-point distances in the original high-dimensional space and the projected low-dimensional space. However, state-of-the-art implementations of MDS are not suitable for the large-scale datasets nowadays, due to two limitations: (1) traditional MDS is not specifically designed for the histogram representations. As is well known to all, dimensionality reduction techniques which aim to map high-dimensional data into much lower dimensional space have been thoroughly studied [17,18]. However, most of existing dimensionality reduction techniques (including traditional MDS) are designed for the Euclidean space, which are not suitable for the scenarios such as BOF (resp. SPM), where data samples are represented using histogram. (2) MDS cannot map an arbitrary query into the lower dimensional space, thus are not applicable in image query tasks. Given an arbitrary query point, the MDS similarity matrix needs to be updated and it is impossible to make a direct mapping into low-dimensional space.

In this paper, we propose a distance-adaptive approach, namely generalized MDS (gMDS) method. Our method is not only applicable for different distance measures (e.g. intersection kernel between histograms from BOF or SPM) but also exhibits limited computation time and space, which makes it suitable for large scale image retrieval. Our method is inspired by MDS techniques, but is in fact a linear approach in the aspect that it learn from MDS mapping an approximate transforming matrix such that an arbitrary query can be mapped into the new lower dimensional space. Moreover, it costs linear time to code an arbitrary query point, which makes our method more useful and scalable in large scale retrieval tasks than other approaches or distance measures in most cases. Remarkably, rather than focusing on reducing the dimensionality of histogram directly, we attempt to find a low-dimensional Euclidean embedding such that distances or the dissimilarity between BOFs or SPMs can be preserved.

In general, our contributions in this paper can be summarized as follows:

- We propose a general framework of dimensionality reduction for scenarios where data points are represented using histograms and corresponding variants. In this framework, we firstly generate a similarity matrix between each pair of histograms using some distance measure, and then find a low-dimensional Euclidean embedding of the original histograms such that similarity between histograms can be preserved in the new low-dimensional space. Inspired by the similarity measurement between SPMs using the kernel function, we use the kernel to build the similarity matrix. Furthermore, we find more latent semantic information of images by embedding the original high dimensional histograms into a more befitting low-dimensional Euclidean space.
- We have significantly improved MDS framework to be applicable in general scenarios. As we all know, for an arbitrary query points, MDS needs to update the similarity matrix, and it is not tolerated for big data applications. In our new framework, we learn a transition matrix between the original high-dimensional space and new low-dimensional space. Through the learned transition matrix, arbitrary query data can be mapped into the new low-dimensional space simply and fast. To the best of our knowledge, we are the first to develop a mapping scheme for MDS and apply it in image retrieval task. On the other hand, our framework is adaptive to many distance measures other, which makes it applicable in more general tasks.
- Some impressive findings based on exhaustive experiments have been observed by our novel dimensionality reduction algorithm on the histogram representation. Extensive experimental study shows that using intersection kernel to build similarity matrix can achieve better performance than Euclidean distance in most cases. Therefore, a small number of dimension of BOF or SPM is sufficient for the learning and retrieval tasks without loss of accuracy. Comparatively, the state-of-the-art methods cannot achieve high accuracy on the same low dimension.

The remainder of the paper is organized as follows. Section 2 briefly presents some related work of dimensional reduction technologies for BOF and SPM. In Section 3, we introduce some preliminary knowledge of dimensionality reduction. Our method is presented in Section 4, including the construction of similarity matrix and the generalized MDS (gMDS) framework. In Section 5, we show experimental results conducted using several real-world datasets from the aspects of both image retrieval and classification problems. Within the experimental study, we report and compare our work with state-of-the-art methods in the aspect of accuracy, time and storage. Finally, in Section 6 we conclude this work and propose some future directions.

2. Related works

In this section, we briefly review some unsupervised dimensionality reduction techniques for the histogram features, including linear dimensionality reduction and nonlinear dimensionality reduction.

A straightforward way of reducing the dimensionality of BOF is to create small-sized codebooks. However, this will quickly bring down the discriminability of BOF representations and degrade recognition performance. Simply selecting a small number of most discriminative visual words or linearly combining the bins will not work well neither [19].

Typical linear dimensional reduction methods mainly include PCA, LDA and pLSA. PCA is a widely used statistical technique for unsupervised dimension reduction. PCA calculates eigenvalues and eigenvectors of the data, then a covariance matrix containing the first n (i.e. the number of desire dimensions) eigenvectors with highest eigenvalues are selected. Mathematically, this is equivalent to finding the best low rank approximation (in L_2 norm) of the data via singular value decomposition (SVD). However, many high dimensional datasets cannot be easily mapped linearly to low-dimensional space. In these cases the high-dimensional data lie on or near a nonlinear manifold (not a linear subspace) and thus PCA cannot model the variability of the data correctly. Therefore, PCA has also been extended to many variants, such as non-circular PCA [20], nonlinear PCA [21], kernel PCA [22], etc.

As Bag-of-Words technology have been popular in computer vision area recently, some methods aiming to generate compact and discriminative representations of Bag-of-Words, like LDA [12] and pLSA [23], have been proposed. Both LDA and pLSA attempt to explain the distribution of visual words in an image as a mixture of a few semantic scene topics or aspects [12,24]. Lazebnik et al. used pLSA as dimension reduction technology [7] since the high dimensional character of the SPM for effective image classification,

Download English Version:

https://daneshyari.com/en/article/407276

Download Persian Version:

https://daneshyari.com/article/407276

Daneshyari.com