Knowledge-Based Systems 92 (2016) 92–103

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Graph regularized sparse coding for 3D shape clustering

A. Ben Hamza*

Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada

ARTICLE INFO

Article history:
Received 4 July 2015
Revised 7 October 2015
Accepted 15 October 2015
Available online 22 October 2015

Keywords:
Clustering
Laplace–Beltrami
Biharmonic distance
Sparse coding

ABSTRACT

Feature descriptors have become an increasingly important tool in shape analysis. Features can be extracted and subsequently used to design robust signatures for shape retrieval, correspondence, classification and clustering. In this paper, we present a graph-theoretic framework for 3D shape clustering using the biharmonic distance map and graph regularized sparse coding. While this work focuses primarily on clustering, our approach is fairly general and can be used to tackle other 3D shape analysis problems. In order to seamlessly capture the similarity between feature descriptors, we perform shape clustering on mid-level features that are generated via graph regularized sparse coding. Extensive experiments are carried out on three standard 3D shape benchmarks to demonstrate the much better performance of the proposed clustering approach in comparison with recent state-of-the-art methods.

1. Introduction

The recent surge of interest in the spectral analysis of the Laplace–Beltrami operator (LBO) has resulted in a plethora of spectral shape signatures that have been successfully applied to a wide range of tasks, including object recognition and deformable shape analysis [1–6], medical imaging [7], multimedia protection [8], and shape classification [9]. Spectral shape signatures are feature vectors representing local and/or global characteristics of a shape and may be broadly classified into two main categories: local and global descriptors. Local descriptors (also called point signatures) are defined on each point of the shape and often represent the local structure of the shape around that point, while global descriptors are usually defined on the entire shape. Moreover, most point signatures can easily be aggregated to form global descriptors by integrating over the entire surface of the shape. Rustamov [2] proposed a local feature descriptor referred to as the global point signature (GPS), which is a vector whose components are scaled eigenfunctions of the LBO evaluated at each surface point. The GPS signature is invariant under isometric deformations of the shape, but it suffers from the problem of eigenfunctions' switching whenever the associated eigenvalues are close to each other. This problem was lately well handled by the heat kernel signature (HKS) [10], which is a temporal descriptor defined as an exponentially-weighted combination of the LBO eigenfunctions. HKS is a local shape descriptor that has a number of desirable properties, including robustness to small perturbations of the shape, efficiency and invariance to isometric transformations. The idea of HKS was also independently proposed by Gębal et al. [11] for 3D shape skeletonization and segmentation under the name of auto diffusion function. Using the Fourier transform's magnitude, Bronstein and Kokkinos [12] introduced the scale invariant heat kernel signature (SIHKS), which is constructed based on a logarithmically sampled scale-space. A generalized shape signature based on spectral graph wavelets was introduced in [4]. The SGW signature is a multiresolution local descriptor that is not only isometric invariant, but also compact, easy to compute and combines the advantages of both band-pass and low-pass filters. A comprehensive list of global spectral descriptors can be found in [13,14].

One of the simplest spectral shape signatures is Shape-DNA [1], which is an isometry-invariant global descriptor defined as a truncated sequence of the LBO eigenvalues arranged in increasing order of magnitude. Gao et al. developed a variant of Shape-DNA, referred to as compact Shape-DNA (cShape-DNA) [9]. The cShape-DNA is an isometry-invariant signature that is obtained by applying the discrete Fourier transform to the area-normalized eigenvalues of the LBO. Chaudhari et al. presented a slightly modified version of the GPS signature [7] by setting the LBO eigenfunctions to unity. This signature, called GPS embedding, is defined as a truncated sequence of inverse square roots of the area-normalized eigenvalues of the LBO. Inspired by the Shape Google framework for 3D shape retrieval [3], Bu et al. introduced a deep learning based approach (3D-DL) for 3D shape classification and retrieval. The 3D-DL framework uses a 2D global descriptor, which is represented by a full matrix defined in terms of the geodesic distance and eigenfunctions of the LBO. Unlike sparse representations, the matrix representation in 3D-DL may, however, require a large amount of memory for data storage compared to 1D global

* Tel.: +1 514-848-2424.
E-mail address: hamza@ciise.concordia.ca

http://dx.doi.org/10.1016/j.knosys.2015.10.019
0950-7051/

descriptors. Moreover, the 3D-DL approach suffers from the long running time of deep belief networks. More recently, a global descriptor called reduced biharmonic distance matrix (R-BiHDM) was proposed in [6] for nonrigid shape retrieval. The R-BiHDM signature has a number of attractive properties that makes it suitable for addressing other shape analysis problems. It is isometry invariant, computationally efficient, robust to various shape deformations, and possesses good discriminative capabilities.

While spectral signatures have received much attention in nonrigid 3D shape analysis [1–6], view-based methods, on the other hand, have also been successfully applied to 3D shape recognition and retrieval [15–18]. Gao et al. proposed a view-based 3D shape recognition and retrieval approach by exploring higher-order relationships among shapes via hypergraphs [15], where a vertex represents a shape and an edge delineates a cluster of views. Another view-based method for 3D shape retrieval was presented in [16], which requires no camera constraint for view capturing of shapes, i.e. a shape may be described by a set of views from any arbitrary direction. Given a query shape, all its query views are first clustered in an effort to generate view clusters, which are then used to construct the query model. More recently, Zhao et al. introduced a feature fusion framework for view-based 3D shape retrieval using multi-modal graph learning [17]. The idea is to extract multiple visual features with the aim of describing each view of a 3D shape, followed by constructing multiple graphs that are fused with different weights. The optimal weights for each graph are then learned via a graph Laplacian regularization approach. Zhao et al. also introduced a generalized 3D depth data matching strategy for action retrieval [18]. This approach employs a 3D shape context descriptor for extracting features of each static depth frame, and then uses dynamic time warping for computing the matching similarity between two 3D dynamic depth sequences. Other applications of shape features include times series visualization [19] and content-based image retrieval [20].

Sparse coding, on the hand, is a family of unsupervised algorithms [21] that are essentially employed for learning an overcomplete set of bases, where an image or shape can be represented by a high-dimensional but sparse feature vector. The goal of sparse coding is to represent a feature vector by a linear combination of a sparse set of basis vectors. Sparse coding has been successfully employed in a slew of computer vision and geometry processing applications such as face recognition [22], image denoising [23], image classification [24–27] and 3D shape retrieval [28], to name just a few. In recent years, various coding schemes have been proposed in the literature, and have proven to be effective in a wide range of computer vision tasks. Wang et al. [25] introduced locality-constrained linear coding, which enforces locality instead of sparsity. In [26,27], the graph regularized sparse coding, also known as the Laplacian sparse coding, was proposed. This coding scheme takes into account the geometric structure of the data space by using a graph Laplacian regularizer in an effort to preserve the locality of the features to be encoded. Unlike sparse coding in which each feature is encoded independently, the graph regularized sparse coding encodes similar features with similar sparse codes, thereby preserving the locality information of the features to be encoded.

Motivated by the good performance of the reduced biharmonic distance matrix in shape retrieval, we propose a robust graph-theoretic clustering approach, called graph biharmonic distance map (GraphBDM), which uses the R-BiHDM signature in conjunction with graph regularized sparse coding. Unlike classification in which objects are assigned to predefined classes, clustering is different in the sense that the number (and labels) of clusters or the cluster structure are not known in advance. The core goal of 3D shape clustering is to organize a dataset of 3D shapes into homogeneous subgroups or clusters in a unsupervised manner using a pre-defined similarity of dissimilarity measure. These clusters are formed in such a way that objects in the same cluster are very similar, while objects in different clusters are very dissimilar.

In addition to exploiting the dependence among the features of shape signatures via the graph Laplacian matrix, the proposed GraphBDM framework performs clustering on sparse mid-level feature vectors that are learned via graph regularized sparse coding (i.e. in the sparse codes domain), thereby seamlessly capturing the similarity between these features. We not only show that our formulation allows us to incorporate feature dependencies, but we also demonstrate that the proposed framework yields better clustering accuracy results compared to state-of-the-art methods, while remaining computationally attractive.

The rest of this paper is organized as follows. In the next section, we briefly overview the Laplace–Beltrami operator and the basics of sparse coding. In Section 3, we introduce a graph-theoretic framework for 3D shape clustering, and we discuss in detail its main algorithmic steps. Experimental results are presented in Section 4. Finally, we conclude in Section 5 and point out some future work directions.

2. Background

A 3D shape is usually modeled as a triangle mesh $\mathbb{M}$ whose vertices are sampled from a Riemannian manifold. A triangle mesh $\mathbb{M}$ may be defined as a graph $\mathbb{G} = (\mathcal{V}, \mathcal{E})$ or $\mathbb{G} = (\mathcal{V}, \mathcal{T})$, where $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ is the set of vertices, $\mathcal{E} = \{e_{ij}\}$ is the set of edges, and $\mathcal{T} = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_g\}$ is the set of triangles, as depicted in the enlarged view of Fig. 1 (left). Each edge $e_{ij} = [\boldsymbol{v}_i, \boldsymbol{v}_j]$ connects a pair of vertices $\{\boldsymbol{v}_i,$

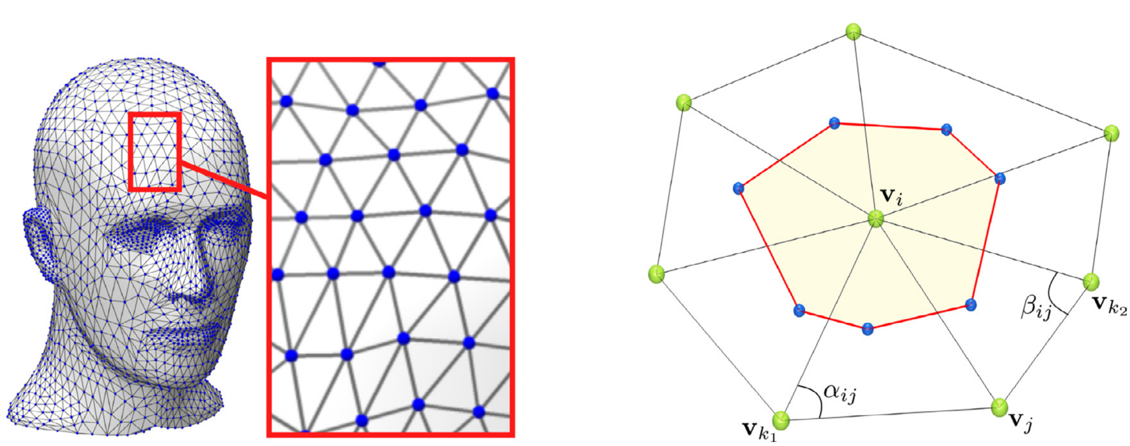

Fig. 1. Triangular mesh representation (left); cotangent scheme angles (right).