Pattern Recognition 52 (2016) 397–409

Contents lists available at ScienceDirect

Pattern Recognition

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

Orthogonal moment-based descriptors for pose shape query on 3D point cloud patches

Huaining Cheng [1], Soon M. Chung *

[a] Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson AFB, OH 45433, USA
[b] Department of Computer Science and Engineering, Wright State University, Dayton, OH 45435, USA

ARTICLE INFO

Article history:
Received 20 March 2015
Received in revised form
3 August 2015
Accepted 24 September 2015
Available online 3 October 2015

Keywords:
3D shape descriptor
Tchebichef moment
Fourier transform
Wavelet transform
Point cloud
LIDAR
3D shape reconstruction
3D shape retrieval

ABSTRACT

When 3D sensors such as Light Detection and Ranging (LIDAR) are employed in targeting and recognition of human actions from both ground and aerial platforms, the corresponding point clouds of body shape often comprise low-resolution, disjoint, and irregular patches of points resulted from self-occlusions and viewing angle variations. Many existing 3D shape descriptors designed for shape query and retrieval cannot work effectively with these degenerated point clouds because of their dependency on dense and smooth full-body scans. In this paper, a new degeneracy-tolerable, multi-scale 3D shape descriptor based on the discrete orthogonal Tchebichef moment is proposed as an alternative for single-view partial point cloud representation and characterization. To evaluate the effectiveness of our descriptor, named *Tchebichef moment shape descriptor* (TMSD), in human shape retrieval, we built a multi-subject pose shape baseline to produce simulated LIDAR captures at different viewing angles and conducted experiments of nearest neighbor search and point cloud reconstruction. The query results show that TMSD performs significantly better than the Fourier descriptor and is slightly better than the wavelet descriptor but more flexible to construct. In addition, we proposed a voxelization scheme that can achieve translation, scale, and resolution invariance, which may be less of a concern in the traditional full-body shape analysis but are crucial requirements for meaningful partial point cloud retrievals.

1. Introduction

During the last decade, some advanced 3D sensors, such as flash LIDAR (Light Detection and Ranging) [1], started appearing in various applications. Even though individual devices can have different designs, they usually provide 3D point clouds or gray/color scaled depth images. With quick accumulation of such data, there is a need to study compact and robust shape description models for content-based information retrieval (CBIR) applications. However, these sensor outputs are typically not as good and complete as the traditional 3D shape data of dense point clouds or watertight meshes generated by full-body laser scanners [2] or graphics software. Instead, they are partial views of 3D objects at a specific viewing angle. When human targets are involved, there are often self-occlusions that break a body's point cloud into random and disjoint patches. The low-resolution setting typically seen in standoff sensing systems further degrades meaningful point connectivity. These problems pose some significant challenges for CBIR systems because such shape degeneracy and sparsity make feature extraction and representation difficult. Many existing 3D descriptors may not be applicable or suitable under this circumstance. For example, without a smooth dense point cloud, it would be difficult to acquire stable first order (surface normal) and second order (surface curvature) geometric properties. Considering the point cloud degeneracy, a method approximating point cloud distribution may be more appropriate and robust. Thus, we conducted in-depth research on point cloud voxelization and discrete orthogonal transforms, which have not been studied much with respect to this type of freeform, irregular point cloud patches.

Specifically, we propose a new Tchebichef moment shape descriptor (TMSD), inspired by the work of [3,4] on applying Tchebichef moments to 2D image analysis, for multi-scale 3D feature representation of point cloud patches. TMSD is made of low-order 3D Tchebichef moments which compact information on shape patterns, so it enables shape search in an embedded subspace. This reduced-dimension search is made possible by TMSD's property of distance preservation in the subspace, which prevents false negatives in the nearest neighbor search. We also applied 3D discrete Fourier transform (3D DFT) and 3D discrete wavelet

* Corresponding author. Tel.: +1 937 775 5119.
E-mail addresses: huaining.cheng@wpafb.af.mil (H. Cheng), soon.chung@wright.edu (S.M. Chung).
[1] Tel.: +1 937 255 9333.

http://dx.doi.org/10.1016/j.patcog.2015.09.028
0031-3203/

transform (3D DWT) on these point cloud patches for a more comprehensive investigation.

Our experiments demonstrate that TMSD performs significantly better than 3D DFT and slightly better than 3D DWT. It is also more flexible than 3D DWT for multi-scale construction because it does not have the restriction of dyadic sampling. The experiments were designed as single-view nearest neighbor (NN) queries of human pose shapes using a newly constructed baseline of partial 3D point clouds, captured through biofidelic human avatars of individual volunteers performing three activities — jogging, throwing, and digging. The NN queries measure the similarity between the query pose shape's descriptor and the descriptors of other shapes in the pose shape baseline. The baseline provides a geometric simulation of LIDAR data at multiple viewing angles, organized into two subsets of horizontal (0°) and vertically-slant (45°) elevation angles. Each subset consists of more than 5500 frames of point cloud patches obtained at different azimuth angles, grouped into 200 plus pose shape classes according to the action pose segmentation and azimuth angle. The construction of this baseline offers us a unique advantage of performance evaluation at a full range of viewing angles, unlike many other studies that were limited to ground-level, frontal, or side views only. We were able to demonstrate that TMSD maintains consistent performance under different elevation angles. This has a particular significance for aerial platforms that have been seldom studied.

Complementary to the NN search, a new voxelization scheme was designed to provide translation, scale, and resolution invariance. The inclusion of scale and resolution normalization distinguishes our work from many existing 3D shape search methods. The majority of existing methods only deal with full-body models in which the complete surface, rather than individual patches and their spatial relationships, defines shape similarity. Therefore, rotational invariance is the main concern. However, in the case of partial point clouds, rotational invariance is meaningless because the point clouds are viewing angle dependent; instead the scale and resolution differences are important variations. This implies that many existing 3D shape descriptors may be ill-suited for the tasks at hand.

Finally, we also investigated whether our native 3D shape analysis is superior to 2D depth image analysis, with respect to the transform-based descriptors. An experiment was designed as a paired performance comparison between a 3D moment descriptor applied to a voxelization of point cloud and its corresponding 2D moment descriptor applied to the depth image made from the same voxelization. This comparison is termed as 3D-outperform-2D hypothesis test for simplicity of discussion.

In summary, this paper exploits LIDAR-like freeform point cloud patches in their native 3D mode from arbitrary viewing angles, which is rarely studied before. This paper's main contributions are four folds: (1) a new simulated human pose shape baseline with biofidelic anthropometry and locomotion to enable performance assessment per viewing angle, (2) the introduction of TMSD as a robust and effective shape descriptor for the representation of irregular point cloud patches, (3) a new voxelization scheme with translation, scale, and resolution invariance, and (4) an experiment supporting the 3D-outperform-2D hypothesis.

The organization of the paper is as follows. Section 2 provides a brief review on 3D shape descriptors with an emphasis on transform-based ones. Section 3 describes the preprocessing of voxelization and normalization. Section 4 introduces orthogonal moment-based descriptors with a focus on TMSD. Experimental results and analyses are given in Section 5. Section 6 contains some conclusions.

2. Related work

2.1. Overview of 3D shape descriptors

This section provides a brief review of various groups of feature-based 3D shape descriptors, which can be built upon 3D spatial relationships, surface geometry, and transform coefficients (see Fig. 1). The review shows why the group of discrete orthogonal moments is a necessary and superb choice for describing point cloud patches. Among the three groups, the transform-based one is discussed separately in the next subsection.

The use of spatial relationship leads to either the global spatial map or local feature density distribution. The spatial map typically maps each surface mesh or sampling point into a global partition framework. For example, the cord-based descriptor [5] captures the entire surface curvature of a shape by binning the angles between a cord (a vector from an object's center of mass to the center of a mesh triangle) and the first two principal axes as well as the radius of a cord. The 3D shape histogram [6] counts the spatial occupancy of a point cloud in a global partition grid of concentric shell, pie-shaped sector, or spider web. Using the similar global partition grid, the 3D shape context [7] records the relative coordinates among N surface sample points. The shape impact descriptor [8] describes a shape through the shape's surrounding Newtonian and relativistic fields, according to the gravity law.

The group of local feature density distributions usually take the form of density distribution of pair-wise spatial relationships among surface sampling points, unrelated to any global partition. One popular approach is the 3D shape distribution [9], which can be constructed from pair-wise Euclidean distances or angles among surface sample points. Isometry-invariant geodesic distance also has been introduced to produce a probabilistic shape description [10].

These descriptors of spatial relationship usually tolerate degeneracy and sparsity; hence they are applicable to our case. However, they may not be ideal candidates. Since each bin is equally important, we are forced to make a tradeoff between descriptor size and performance. Moreover, a large number of histogram bins or a refined partition scheme could result in undesirable high dimensionality. Even though some data reduction techniques such as the Principal Component Analysis (PCA) have been used to reduce the dimensionality, it is very difficult to acquire datasets that are sufficiently large to achieve consistency [11]. In other words, the outcome of data reduction is tied to the specific dataset and not scalable.

The surface geometry type of descriptors are generated from local geometric properties, such as radial distance (zero order), surface normal (first order), surface curvature (second order), and so on. These local geometric properties can form both global and local shape descriptors, depending on whether they are aggregated to a global partition framework or collected as a bag of features.

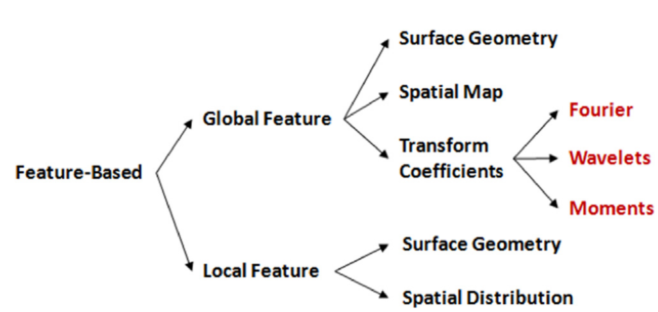

Fig. 1. Taxonomy of feature-based 3D shape descriptors. Three transform-based ones are studied in this paper.