Title:
Learned 3D Shape Descriptors for Classifying 3D Point Cloud Models

Authors:
Xiaojun Zhao, Xiaojun.zhao@uconn.edu, University of Connecticut
Horea T. Ilies, ilies@engr.uconn.edu, University of Connecticut

Keywords:
Point Clouds, Classification, Shape Descriptors, Zernike Moments.

DOI: 10.14733/cadconfP.2016.142-146

Introduction:
The recent developments in 3D sensing devices that deliver high-quality raw 3D data in real time offer growing opportunities to explore the usage of this data in a variety of 3D perception and reasoning tasks. In this paper, we focus on the problem of classifying 3D point clouds, and we are integrating different supervised machine learning classifiers with several capable yet underexplored shape descriptors based on visual similarity (light-field), angular radial transform (ART) and Zernike moments. Specifically, we investigate the use of 3D Zernike descriptors as well as a combination of 2D ART descriptors with light field techniques to construct and compare the performance of practical descriptors for 3D point cloud classification. We train our classifiers with a database of point clouds corresponding to several common objects obtained by sampling polygonal models obtained from Google's 3D Warehouse and by post-processing them to attain controlled levels of density and noise. We show that these descriptors provide a promising alternative to the current shape descriptors employed for classifying point clouds in the presence of noise.

The classification of shapes relies on the existence of a similarity or dissimilarity measure between shapes. A good representation of the shape features in terms of a shape descriptor must be discriminating, efficient to compute and compare, invariant under isometries, insensitive to geometric as well as topologic noise, and robust to degeneracies. Many 3D shape descriptors have been recently proposed, and applied primarily to tessellated models as detailed in the recent reviews that appear in Tangelder et al. [9] and Kazmi et al. [4]. At the highest level, the descriptors can be grouped based on their representation into: global features (e.g., volume, statistical moments), global feature distributions (e.g., histograms), spatial maps (e.g., spherical harmonics), and local features (e.g., shape spectra) as detailed in [9]. However, almost all existing shape descriptors have been defined for tessellated models, and very few exist that can be applied to native point clouds. For example, Williams et al. [10] developed a practical and convergent estimate of the Laplace-Beltrami operator for point clouds, which is symmetric under real-world conditions, and used it to construct compact shape signatures of point cloud models. These signatures were them used in conjunction with topological clustering techniques via Vietoris-Rips clustering to segment point cloud models of engineering artifacts into geometric features of engineering interest.

Two-dimensional image moments have been traditionally used for image recognition, but they suffer from noise sensitivity, and information suppression. These difficulties have been addressed in 2D by introducing the Zernike moments defined with Zernike polynomials. For example Chen et al. [3] proposed the Light Field Descriptors (LFD) which compute 2D Zernike moments and Fourier coefficients based on the silhouettes images taken from cameras on the vertices of a dodecahedron. These Zernike moments have been extended to 3D by Canterakis [2], and have been applied to tessellated model retrieval by Novotni et al. [7], where it is argued that the 3D Zernike moment-based descriptors lead to better retrieval performance and robustness against topological and geometrical artifacts of tessellated models than state of the art descriptors.

To the best of our knowledge, the work presented in this paper is the first application of 3D Zernike moments to point cloud classification.

Main Idea*:*

In this paper we construct shape descriptors for 3D point clouds with 3D Zernike moments, develop a computational framework to compute practical descriptors of point cloud models, and compare the classification performance against established descriptors. Specifically, we compare the classification performance against that of the Light Field Descriptors (LFD) based on the Angular Radial Transform, or ART. These descriptors are then used to train machine learning classifiers on a database of point cloud models to perform effective classification of point clouds that may or may not contain geometric and topological noise. While here we simply summarize our new paradigm to classify 3D point clouds with Zernike descriptors, additional details are provided in the accompanying journal paper.

ART descriptors were proposed by Kim et al. [6]. As a region-based shape descriptor, the original ART is defined as a set of normalized magnitudes of the ART moments or coefficients computed on a 2D image, and is capable of describing both connected and disconnected regions with rotational invariance. These shape descriptors possess several desirable properties, such as compact size, invariance to similarity transformations, and robustness against noise and scaling, and are able to capture features of 2D color images [9]. The ART coefficients, F_{nm} of order n and m, are defined by:

$$F_{nm} = \int_0^{2\pi}\int_0^1 V_{nm}(\rho,\theta)\, f(\rho,\theta)\rho\, d\rho\, d\theta \quad , \tag{1}$$

where $f(\rho,\theta)$ is the image function expressed in a polar coordinate system, and $V_{nm}(\rho,\theta)$ are the separable ART basis functions. The ART has been generalized to the indexing of 3D tessellated models (see for example [8]).

Zernike moments are mappings of the function that defines the shape (or image) onto a set of orthogonal polynomials over a unit ball. Similar to the ART descriptor, the Zernike descriptors can be defined from the magnitudes of a set of orthogonal complex moments of objects, and are rotationally invariant.

Two dimensional ZD of images can be obtained by computing the magnitudes of Zernike moments, which are given by [5]:

$$A_{\mathrm{nm}} = \frac{n+1}{\pi}\sum_x\sum_y f(x,y)V_{nm}^*(\rho,\theta)\text{, for } x^2+y^2 \le 1 \tag{2}$$

where V_{nm} are the 2D Zernike polynomials. The 3D Zernike moments Ω_{nl}^m of an object are defined as [7]:

$$\Omega_{nl}^m = \frac{3}{4\pi}\int_{|\mathrm{x}|} f(\mathrm{x})\overline{Z_{nl}^m(\mathrm{x})}d\mathrm{x} \tag{3}$$

where Z_{nl}^m are the 3D Zernike polynomials. We note that the Zernike moments Ω_{nl}^m are not invariant under rotations, but 3D rotationally invariant Zernike descriptors can be defined by using spherical harmonics [7].

LFD uses the observation that two similar objects look similar from similar viewing angles. We set 20 view-points (or cameras) on 20 vertices of a regular dodecahedron. Since the cameras on the opposite vertices would produce the same silhouettes, 10 object views are needed for each model. To achieve the rotational invariance property, each 3D model is “observed” by 10 cameras in 10 different orientations. Therefore, a total of 100 silhouettes are determined for each model, and each 2D silhouette is encoded by a feature vector extracted by ART. To compute the Zernike descriptors, we first convert 185 meshed models from Google’s 3D Warehouse (see below) into point clouds models via Poisson disk sampling, and estimate the surface normal at each point from the k-nearest neighbors of each query point. Then, the model is scaled to fit inside a unit cube and then translated to the origin of the coordinate system. Next, point cloud models are voxelized to compute the 3D Zernike descriptors [7]. A feature vector is then computed for each model by using LFD (with ART) and 3D ZD. This allows the construction of a feature matrix having a size of 185×L, where L is the length of a single feature vector for a database of 185 models.

We tested the descriptors on a database of 3D point cloud models obtained by downloading and pre-processing 3D models from Google's 3D Warehouse, which contains 185 models belonging to six categories, namely cars, planes, mugs, tables, chairs and lamps. We sampled the downloaded polygonal models with and without noise to generate the point cloud database. The general pipeline of our experimentations is shown in Fig 1.

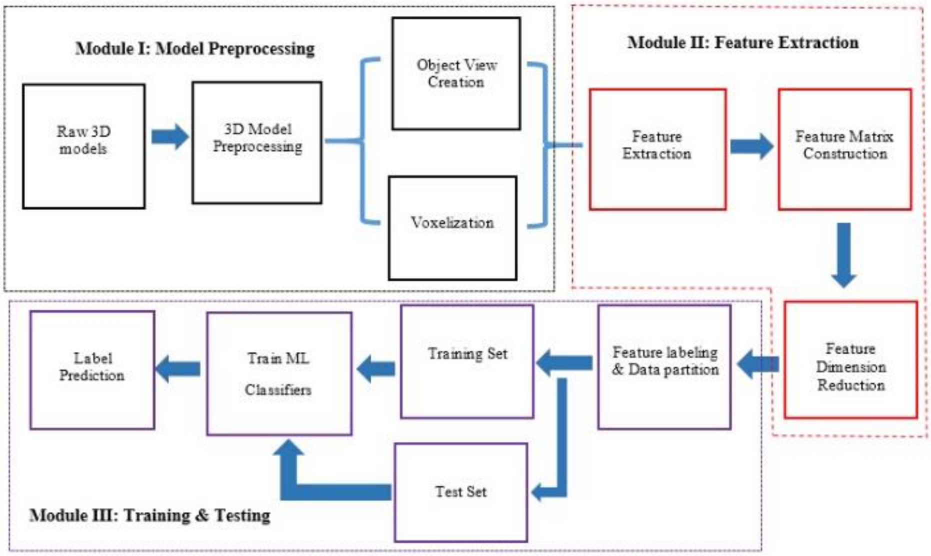

Fig. 1: Pipeline of the 3D model classification system.

To reduce the high computational complexity of training machine learning classifiers, we use PCA to reduce the dimensionality of the feature vectors to 20-30 feature components. We partition the dataset into training and test sets so that the training set contains 125 models, and the remaining 60 models are left in the test set. To prevent the overfitting issue, we repeat the whole process 50 times (with and without the k-fold cross validation) with different partitions of the data set. The final step is to compare the classification effectiveness and performance among different classifiers, and an illustration of this process for a chair model is shown in Fig.2.

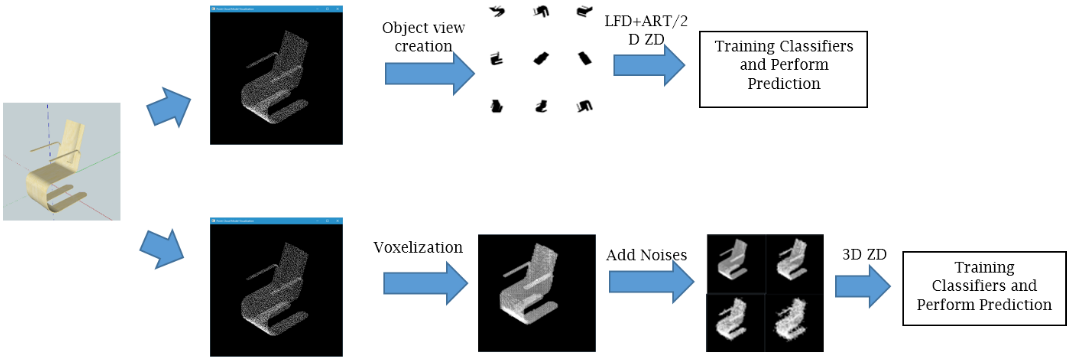

Fig. 2: The classification procedure applied to a chair model.

The average classification accuracy for LFD+2D ART among different classifiers is presented in Tab.1. Additional data is included in our accompanying journal paper. Our experiments with the object database described above show that the classification accuracy is relatively uniform across the model categories.

Light-Field Descriptors		*Avg. Classification Accuracy*	
		With 10-fold cv (repeat 50 times)	*Without cv (repeat 50 times)*
MLP (Input layer size: 40; Hidden layer size: 25; Training Function: Sigmoid)	LFD+ART	80.54%±1.0%	80.00%±1.0%
KNN (Feature size: 30; k=1)	LFD+ART	98.22%±1.0%	98.57%±1.0%
Random Forests (#Trees: 120)	LFD+ART	98.70%±1.0%	98.08%±1.0%

Tab. 1: Classifcation accuracy of LFD+2D ART with three different machine learning algorithms: Multi-Layer Perceptron (MLP), k-Nearest Neighbor (kNN) and Random Forests.

The classification accuracy for 3D Zernike descriptors with different classifiers is presented in Tab.2, which shows the results for 3D Zernike polynomials order of 15. The accompanying journal paper discusses the influence of this polynomial order on the classification accuracy for our database.

3D ZD	**KNN (k=1)**	**Random Forest (# tree = 120)**	**MLP**	**RBF Network**	**Adaptive Boosting**	**kNN (5-fold cv)**	**RBF network (5-fold cv)**
n=15 (72 feature components)	87.87%	82.37%	75% (hidden layer	85.67%	76.17%	89.52%	89.18%

Tab. 2: Performance comparisons of different machine learning algorithms for Zernike polynomials of order 15.

Furthermore, our investigations detailed in the journal version of this paper indicate that the classification accuracy does not change significantly as the level of noise increases. This suggests that the 3D Zernike descriptors are robust against random noise levels similar to those observed in 3D sensing with commercial cameras.

Conclusion:

This work explores the classification task for 3D point cloud models by incorporating supervised machine learning approaches with powerful shape descriptors that have traditionally been used for classifying 3D polygonal models. We consider two different approaches to feature extraction from point clouds, namely Light-Field Descriptors built with ART moments, and 3D Zernike Descriptors computed directly on the point clouds. The major difference between the two types of approaches is that the LFD-based approaches rely on view-similarities and extract lower-dimensional features (i.e. 2D shapes), while 3D ZD compute features directly based on 3D data (i.e. 3D point clouds). Our preliminary experimental results showed that LFD+ART outperform 3D ZD in terms of classification performance, but they both have the potential to robustly and effectively classify 3D point cloud models without requiring a mesh of the point cloud. Furthermore, our experiments show that 3D Zernike descriptors are robust against noise levels typically found in point cloud data output by current commercial RGB-D cameras.

Our preliminary experiments show that the 3D Zernike descriptors provide a promising alternative to the current shape descriptors employed for classifying noisy point clouds. Furthermore, the practicality of 3D Zernike descriptors coupled with their potential for parallel implementations on the GPU [1] makes them capable candidates for real-time applications in 3D sensing and perception.

References:

[1] Berjón, D.; Arnaldo, S.; Morán, F.: A parallel implementation of 3D Zernike moment analysis, IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics, 2011. http://dx.doi.org/10.1117/12.876683

[2] Canterakis, N.: 3D Zernike moments and Zernike affine invariants for 3D image analysis and recognition, In 11th Scandinavian Conf. on Image Analysis. 1999.

[3] Chen, D-Y.; et al: On visual similarity based 3D model retrieval, Computer Graphics Forum, 22(3), 2003, 223-232. http://dx.doi.org/10.1111/1467-8659.00669

[4] Kazmi, I. K.; You, L.; Zhang, J. J.: A survey of 2D and 3D shape descriptors, Computer Graphics, Imaging and Visualization (CGIV), 2013 10th International Conference. IEEE, 2013. http://dx.doi.org/10.1109/cgiv.2013.11

[5] Khotanzad, A.; Yaw, H-H.: Invariant image recognition by Zernike moments, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(5), 1990, 489-497. http://dx.doi.org/10.1109/34.55109

[6] Kim, W.-Y.; Kim, Y.-S.: A new region-based shape descriptor, ISO/IEC MPEG99 M, 1999, 5472.

[7] Novotni, M.; Klein, R.: 3D Zernike descriptors for content based shape retrieval, Proceedings of the eighth ACM symposium on Solid modeling and applications, ACM, 2003. http://dx.doi.org/10.1145/781606.781639

[8] Julien, R.; Coeurjolly, D.; Baskurt, A.: Generalizations of angular radial transform for 2D and 3D shape retrieval, Pattern Recognition Letters, 26(14), 2005, 2174-2186. http://dx.doi.org/10.1016/j.patrec.2005.03.030.

[9] Tangelder, J. W. H.; Remco C. V.: A survey of content based 3D shape retrieval methods, Multimedia tools and applications, 39(3), 2008, 441-471. http://dx.doi.org/10.1007/s11042-007-0181-0

[10] Williams, Reed; Ilies, H.: Practical Shape Analysis and Segmentation Methods for Point Cloud Models, Technical Report, University of Connecticut, December 2015.