1 Previous work

Shape analysis, classification, and retrieval has been developed quite extensively, especially in last decade, in part due to a rise in the number of available models and techniques to process them. In order to effectively exploit this rise in usable shapes, the development of a comparison and retrieval system is necessary, prominently in the 3D model search engine developed at Princeton[?], as well as at the National Research Council of Canada[?], the National Taiwan University[?], and others[?, ?, ?].

There is a broad swath of literature dedicated to 3D retrieval shape retrieval, with many different methods with their particular strengths and weaknesses. All of these must be evaluated with respect to several salient attributtes: foremost, robustness and discrimination, but also efficiency, necessity of preprocessing, ability to partially match, and strictness of requirements for data (e.g. mesh versus point cloud, closed meshes, etc.). Evaluating robustness and performance can be done with several standard datasets, including Princeton's Shape Benchmark[?] and the MPEG-7 dataset[?, ?].

1.1 Global features

One category of papers that aid in 3D retrieval are those which use global features, i.e. features which look at the whole shape.

For example, Zhang and Chen[?] proposes an algorithm to extract global features like volume, moments, and Fourier transform coefficients from a mesh representation. It does this by simply summing up the features on each of the elementary shapes that form this mesh. Though efficient, it uses nonsophisticated features and requires a mesh shape representation. They then extend[?] this with a clever trick with manual annotation, which selects the least-known shape (based on their feature representation) and chooses that as the next annotation.

Other global features are explored by Corney et al.[?], who create an online (Internet-based) 3D shape search engine named ShapeSifter. Their coarse filters, which are used to filter and classify before more fine distinctions and distances are made, are based on a novel application of convex hulls. The three indices they use to do a preliminary filtering of candidate 3D shapes include *hull crumpliness* (the ratio of the object's surface area to the surface area of its convex hull), *hull packing* (the percent of the convex hull volume not occupied by the original object), *hull compactness* (the ratio of the convex hull's surface area cubed over the convex hull's volume squared). Though extremely quick, this work is at best a rough first approximation that require the use of more differentiating methods.

In the same vein of speedy similarity measures, Kolonias et al.[?] propose the use of the following shape descriptors: the aspect ratio, a binary 3D shape mask, and (more complex) the edge paths of the models. The set of paths that outline the shape of the 3D object is extracted using a proposed algorithm which extracts those geometric properites from an input VRML file. These sets are compared between 3D objects by first checking the similarity of angles and lengths, and then finding some metrics on corresponding paths.

A method which is only superficially similar to ours is that of Horn[?], who represent the shapes of surfaces by mapping each polygon of the 3D shape to the corresponding unit normal point on the Gaussian (unit) sphere, weighted by the surface area. This extended Gaussian image uniquely defines a convex polyhedron, up to a translation. This is extended by Kang and Ikeuchi[?], which simply stops discarding the distance to the origin of the polygon by encoding it onto the imaginary part of the complex sphere. However, these both require pose normalization.

One famous and highly cited innovation in using global feature for shape retrieval involves the concept of a global feature distribution, which uses the distribution of features rather than the features directly. Osada et al.[?] explores this concept by measuring properites like distance, area, volume, and angle between random points of the relevant shape, constructing a distribution, and using this to compare to other shapes. The most effective is the D2 shape distribution, which is the distribution of random points on the surface of the shape.

Shape distributions are quite good at interclass differentiation (i.e. between classes of shapes), though not as good at intraclass differentiation[?] (i.e. between shapes which belong to the same broad class but still require distinction) because they require finer representations.

1.2 Spatial features

Spatial features distinguish themselves by somehow capturing and using the spatial location of the 3D model for use in comparisons. In general, they requires pose normalization to be effective since they aren't rotationally invariant.

One early approach in spatial features is proposed by Ankerst et al.[?], who simply use 3D shape histograms. This was applied to molecular surfaces by taking a histogram of the points in concentric shells (and subdividing those shells into sectors) around a 3D shape's centroid.

Kazhdan et al.[?] propose an approach that can transform any rotation dependent shape descriptors to rotation independent one through the use of spherical harmonics. They esssentially calculate spherical harmonics on concentric spheres of the shape and then use frequency analysis to construct a descriptor indexed only by radius and frequency. This is invariant to rotations.

Another highly cited paper is the the work by Novotni and Klein[?], which uses 3D Zernike descriptors instead of spherical harmonics to obtain rotational invariance. This is an improvement over spherical harmonics because spherical harmonics cannot detect internal rotations of the object.

1.3 Local features

An approach that has some similarities with our work is that of Shum et al., which presents a way to measure similarity between 3D shapes. They map local curvatures a distribution on a sphere, and then construct a distance metric between these spherical approximations of 3D shapes. However, this requires closed surfaces.

Ohbuchi et al.[?] use SIFT features extracted from range images and put into a bag-of-features representation. The range images are sampled uniformly from the view sphere.

1.4 View-based features

The idea of view-based features is that much of intrinsic information of a 3D shape can be encoded into multiple 2D views from different angles of that shape. Therefore, it's possible to compare two shapes by simply comparing the views.

A recent paper by Zhu et al.[?] uses deep learning to learn a feature representation of a shape using autoencoders on 2D views of a 3D shape. They combine this with the local descriptors by the previous Ohbuchi et al[?] to obtain state-of-the-art results.