A Survey of 2D and 3D Shape Descriptors

Ismail Khalid Kazmi, Lihua You, Jian Jun Zhang
National Center for Computer Animation
Bournemouth University
Bournemouth, UK
email: ikazmi@bournemouth.ac.uk, lyou@bournemouth.ac.uk, jzhang@bournemouth.ac.uk

Abstract - **3D objects play a vital role in computer games, movies, TV and military research. Researchers have studied 3D object retrieval techniques using 2D and 3D shape descriptors in the last several decades. In this paper, some of the most popular shape descriptors are reviewed. We first discuss the important criteria for a good shape descriptor. Then, we classify 2D and 3D shape descriptors into several categories. For each category, we review the state-of-the-art techniques along with their advantages and disadvantages. We also examine the recent contributions and improvements in these techniques and provide a comparison with other successful techniques based on previous literature and surveys. Finally we discuss the possible future directions.**

I. Introduction

Three dimensional objects possess superior importance in the fields of gaming, movies, TV, engineering design, biology, and military research. The demand has been growing to efficiently transmit them over the Internet with the rise of cloud computing technologies. Currently the common way to search for images and 3D models over the internet is via text queries and online catalogues. Retrieval of images using another image as a query is still not sophisticated enough to provide us with the desired results. The same is true for online 3D model retrieval. Moreover in entertainment, games and special effects industry, there is a huge demand of modelling highly realistic characters quickly and easily. Unfortunately, modelling realistic 3D characters is still a time consuming and laborious task which requires highly skilled and experienced modellers. Sketch based modelling techniques exist, however these techniques are prone to roughness, inaccuracy, and difficulty. Thus there exist a gap between traditional modelling techniques and sketch-based modelling techniques. This requires powerful content based retrieval techniques and sketch based modelling interfaces to be implemented.

Content based information retrieval (CBIR) has been studied by researchers from several decades. CBIR is the process of using shape descriptors to extract important visual features from the boundary and the interior of a shape from shape databases. Some important features include contours curves, shape signature, shape histogram, shape context, and spectral features [3]. Several notable databases have been developed in the past including MPEG-7 shape database [8], Princeton Shape Benchmark [44], Konstanze University (CCCC) 3D benchmark, McGill Shape Benchmark (MSB) [57], and National Taiwan University database (NTU) [28]. Several excellent surveys on 2D and 3D shape descriptors (SDs) exist including but are not limited to [2], [3], [30], and [54]. In this survey paper, we present and summarize several important 2D and 3D shape descriptors which have undergone extensive research in CBIR, and discuss their advantages and disadvantages. We provide taxonomies of the 2D and 3D shape descriptors. We also present a few hybrid techniques which have gained importance in the recent years because of their promising effectiveness and performance.

The reminder of the paper is organized as follows. The definition and characteristics of shape descriptors are presented in Section 2, 2D shape descriptors are discussed in Section 3, 3D shape descriptors are reviewed in Section 4, and 3D Shape Retrieval Contest is introduced in Section 5 followed by a section of Conclusion & future research directions.

II. Definition and Characteristics of Shape Descriptors

Generally speaking, a shape descriptor is a simplified representation of a 2D or 3D shape in the form of a vector containing a set of numerical values or a graph-like structure used to describe the shape geometrically or topologically [1], [31]. Shape descriptors are evaluated on the basis of several characteristics which define the overall quality and effectiveness of a shape descriptor. The following common characteristics of an effective shape descriptor have been proposed in [1], [2], [3], [4].

a) Discriminative accuracy: To accurately distinguish one shape from another based on subtle differences

b) Transformation (translation, scaling, and rotation) invariance: Also known as pose normalization

c) Robustness against model degeneracies / roughness

TABLE I. CLASSIFICATION OF 2D SHAPE DESCRIPTORS

Contour Based Descriptors	Fourier Descriptors (FD)
	Wavelet Descriptors (WD)
	Curvature Scale Space (CSS)
	Shape Context (SCD)
Region Based Descriptors	Zernike Moment Descriptors (ZMD)
	Scale Invariant Feature Transform (SIFT)
	Angular Radial Transform (ART)
Hybrid Descriptors	FD + ART
	FD + ZMD

d) Uniqueness: Each shape descriptor must be uniquely coupled with a unique shape

e) Performance and memory efficient

f) Partial matching: robust against incomplete shapes

g) Insensitive to noise: Small changes in the shape to lead to small changes in the shape descriptor

III. 2D SHAPE DESCRIPTORS

Over the past 2 decades, 2D SDs have been actively utilized in 3D search engines and sketch based modelling techniques. Some of the most popular 2D SDs are Curvature Scale Space (CSS), Fourier Descriptor, SIFT, ART, ZMD, and Shape Contexts. In existing literature, 2D SDs are classified into two broad categories: contour based and region based. Contour based SDs extract shape features from the contour of a shape only. In contrast, region based SDs obtain shape features from the whole region of a shape. In addition, hybrid techniques have also been proposed in the past, which are a combination of contour and region based techniques [9]. Therefore we classify 2D SDs into three categories and the third category in our taxonomy is called Hybrid 2D Shape Descriptors. Table I shows the classification of the 2D SDs.

A. Contour Based Descriptors

Contour based descriptors (CBD) only consider the boundary of the shape and neglect the information contained in the shape interior [58]. These descriptors are very efficient at filtering out the results based on the boundary points because of their low computation complexity. However, they are not good at handling image noise and thus not accurate.

1) Fourier Descriptor (FD): A Fourier descriptor (FD) is a representation of the shape obtained after the application of Fourier transform on the coefficients of the shape signature of the shape. It can be obtained through extracting boundary coordinates, computing shape signatures, and generating Fourier descriptors. Here, a shape signature means any 1-D function used to represent 2-D shapes or boundaries.

Boundary coordinates are extracted from the planar closed curves such as sketches or silhouette images. One of the four frequently used shape signature functions: centroid distance, complex coordinates, curvature function, and cumulative angular function, can be used to compute shape signatures. Among them, Fourier descriptor method using centroid distance performs the best [10].

When using centroid distance, the centroid of N boundary points is first found. Then, a shape signature $s(t)$ (t = 0, 1, …, N-1) is obtained to represent the distances of the N boundary points to their centroid. Finally, the discrete Fourier transform of $s(t)$

$$u_n = \frac{1}{N}\sum_{t=0}^{N-1} s(t)\exp\left(\frac{-j2\pi nt}{N}\right), n = 0,1,\ldots,N-1 \qquad (1)$$

is used to determine Fourier coefficients u_n (n = 0, 1, …, N-1), which are usually called Fourier descriptors (FDs) of the shape, denoted as FD_n (n = 0, 1, …, N-1) [5]. Shape signature computation is a preliminary step of FD in which a 1-D function is obtained from the boundary coordinates of the shape.

Advantages of FDs [3] are: 1) simple to compute, 2) simple to normalize for simplifying the shape matching process, 3) captures both global and local features, 4) insensitive to noise.

Recently, Zhang et al. [10] proposed a modified FD (MFD). Their technique rests on the idea that geometry information along with non-geometry information in the contour of an object is important for accurate discrimination which according to the authors is ignored in the previous works.

2) Wavelet Descriptor (WD): Wavelet descriptor (WD) was initially proposed by Chuang and Kuo [6] in which the authors employed wavelet transforms to describe the shape of planar closed curves. It is multi-resolution approach which decomposes the shape into several components in multiple scales. The components in the higher resolutions contain the global information and the components in the finer resolutions contain the more detailed local information.

Wavelet descriptors have several useful advantages over other descriptors, including insensitive to noise, invariance, uniqueness, and stability against boundary variations. Chung and Kuo [6] concluded that as compared to FDs, WDs are more efficient in representing and detecting local features of a curve.

Recently, Khoje and Bodhe [11] used a combination of FD and WD to evaluate the irregularity of the mango fruit shape, and named the combination Wavelet Fourier Descriptor (WFD). WFD outperforms region based and contour based FD with classification efficiency of 89.83% making it a promising 2D descriptor to be used in view based 3D object retrieval. However WFD has not been tested for shape recognition where the shape resembles a human character. To evaluate this technique in human character shape retrieval and 3D search engines could be a possible future research direction.

3) Curvature Scale Space (CSS): CSS is one of the most widely used algorithms in content based image retrieval (CBIR) and has undergone a substantial amount of refinement since it was introduced by Brady and Asada [12]. The authors in [7] and [8] have clearly described the CSS algorithm. CSS divides the shape into convex and concave segments by identifying a set of inflection points – points where the curvature of the shape is zero. The CSS algorithm involves calculating the curvature of the contour while progressively smoothing down the curve and then finally generating the CSS image. Smoothing is done till the point when there are no zero curvature points left and the whole contour becomes convex. The CSS image is a curve on a 2 dimensional plane which plots all the zero crossing points. The horizontal axis contains the position of the points on the contour while the vertical axis contains the level of iteration / smoothing performed on the contour. A two dimensional vector is associated with each inflection (zero crossing) point expressed as (s,l) where s corresponds to the amount of smoothing applied while l is the position of the point on the contour curve. To compute l, we take the distance from an arbitrary chosen point to the inflection point on the contour, while moving clockwise on the contour. The positions of the inflection points mapped onto the CSS image are the indices of these positions, which are peak values. Mokhtarian et al [13] were the earliest to improve the CSS algorithm. They proposed a new way for extracting the feature vector by taking into account the maxima of the zero crossing points on the contour. They proposed a fast indexing method using the aspect ratio of the CSS image along with eccentricity and circularity of the contour.

4) Shape Context Descriptor (SCD): Shape context was used as a novel shape descriptor for measuring shape similarity by Belongie et al [14]. The main idea of shape context is to find the correspondence between the two shapes and find out the dissimilarity measure between the two shapes. To find the correspondence between the two shapes, N points are sampled from the contour of the shape and a reference point is fixed. The points are sampled using an edge detector algorithm. Then a set of vectors are computed originating from the reference point to all the other sampled points. The shape context for each point pi is defined as a histogram hi of the relative polar coordinates of the remaining sampled points.

$$h_i(k) = \#\{Q \neq P_i : (Q - P_i) \in bin(k)\} \qquad (2)$$

Relative to the reference point the distribution of the remaining points is computed. To find the correspondence between the two shapes we find for each sample point on one shape a sample point on the other shape for which the shape context is most similar. After the correspondence of points has been calculated, the shapes are mapped onto each other using alignment transformation estimations. Belongie et al [14] have used the regularized Thin Plate Spline (TPS) for representing flexible coordinate transformations. This transformation idea derived from the classical work by D'Arcy Thompson (On Growth and Form), in which he explained that the similarity of two similar biological forms can be measured by simple mathematical transformation between corresponding features. The dissimilarity between the two shapes is found by summing up the matching errors between corresponding points. After the dissimilarity measurements, object recognition is performed using nearest neighbour techniques. Shape context is a very robust shape descriptor which is highly discriminative. Besides being highly discriminative, shape context is transformation invariant, robust against shape variations, and they have few outliers [1].

B. Region Based Descriptors

Region based descriptors (RBD) take into account the boundary as well as the internal information of the image and are more robust against noise and other shape variations. Most popular RBDs are Zernike moments descriptor (ZMD), Angular Radial Transform (ART), and Scale Invariance Feature Transform (SIFT) descriptor.

1) Zernike Moments Descriptor: The applications of Zernike moments as a 2D shape descriptor in image analysis began with the pioneering work by Teague in [15]. Zernike moment descriptor is one of the most commonly used region based descriptors and has undergone constant improvement since it was first introduced. Zernike moments are continuous orthogonal moments derived from Zernike polynomials, a ground breaking work by the Nobel laureate Fritz Zernike [17]. Liao and Pawlak [16] provided a very clear definition and explanation of Zernike moments. In order to understand Zernike moments, it is important to understand the concept behind Zernike functions. Liao and Pawlak [16] defined Zernike functions as: "A set of complex orthogonal functions with a simple rotational property which forms a complete orthogonal basis over the class of square integrable functions defined over the unit disk." [16]. Celebi and Aslandogan [18] pointed out the following advantages and disadvantages of Zernike moments. The advantages are

a) Rotation Invariance
b) Robustness against small changes in shape
c) Insensitive to noise

d) Expressiveness: There is minimum information redundancy as the basis are orthogonal

The disadvantages are:

a) Coordinate space normalization: The image coordinate space must be transformed to the domain where the orthogonal polynomial is defined (unit circle for the Zernike polynomial).

b) Discrete approximation of continuous integrals: The continuous form of the Zernike moments must be approximated to discrete form. This approximation leads to errors in the computations as investigated by [16].

As the order of the Zernike moments increases, the computation complexity increases.

2) Scale Invariant Feature Transform (SIFT) Descriptor: The research into the SIFT algorithm began with the pioneer work by Lowe [19]. Lowe observed that image recognition should involve extracting local features which should be invariant to transformation, occlusion, illumination and affine transformations and highly distinctive in nature. Lowe based his approach on the observations and contribution from Schmid and Mohr [20] that efficient object recognition can be achieved by using local image descriptors that can be sampled at a large number of repeatable locations. The SIFT algorithm converts the image into a huge collection of location feature vectors that are scale, rotation and translation invariant. The first step of the algorithm is to extract the scale invariant features in the image using the staged feature approach. Lowe used a scale space analysis approach and Gaussian kernels because it allows high efficiency and rotation invariance. The feature vectors extracted from the image are called SIFT keys. SIFT keys are then used for indexing for identifying candidate object models by using a nearest neighbour algorithm. The advantages of SIFT algorithm are: 1) invariant to scale, rotation and translate, 2) partially invariant to illumination changes, 3) robust against occluded objects, 4) robust against object degeneracies, 5) insensitive to noise.

3) Angular Radial Transform (ART) Descriptor: Angular Radial Transform (ART), proposed by Kim and Kim [21], is a popular region based descriptor which is used in the MPEG-7 standard [8]. It was concluded by Core Experiments that among the descriptors, the overall performance of ART descriptor is the best for region-based similarity [8]. ART is defined by Bober et al. as "the orthogonal unitary transform defined on a unit disk that consists of the complete orthogonal sinusoidal basis functions in polar coordinates" [8]. Mathematically the ART coefficients are determined by:

$$F_{nm} = \int_0^{2\pi} \int_0^1 V_{nm}(\rho,\theta), f(\rho,\theta)\rho d\rho d\theta \quad (3)$$

where F_{nm} is an ART coefficient of order n and m, $f(\rho,\theta)$ is an image function in polar coordinates, and $V_{nm}(\rho,\theta)$ is an ART basis function that are separable along the angular and radial directions, i.e,

$$V_{nm}(r,q) = A_m(q)R_n(r) \quad (4)$$

In order to achieve rotation invariance, an exponential function is used for the angular basis function,

$$A_m(q) = \frac{1}{2p}\exp(jmq) \quad (5)$$

C. *Hybrid 2D Shape Descriptors*

Wei et al. [49] proposed the two-component solution (TCS), in which centroid distances, contour curvature and Zernike moments were selected as the shape features, while a two-component strategy was applied in feature matching. They used their approach in trademark image retrieval (TIR). Singh [9] proposed a hybrid approach which combines Fourier descriptor with ART. Fourier descriptor is used to extract local features while ART is used to extract global features. The authors performed experiments with FD + ART and FD + ZMD combinations. They showed that their techniques perform better than the TCS technique.

IV. 3D Shape Descriptors

3D shape descriptors have been extensively used by researchers since the 1990's in 3D search engines and sketch based modelling systems. Tangelder et al. [2] and Zhang et al. [30] have publishesd comprehensive surveys of 3D SDs. Table II shows a classification of 3D SDs.

TABLE II: CLASSIFICATION OF 3D SHAPE DESCRIPTORS

View Based	Adaptive Views Clustering
	Compact Multi-View Descriptor
	LightField Descriptor (LFD)
Histogram Based	Shape Spectrum
	Generalized shape distributions
	Bag-of-Features (BoF)
Transform Based	Spherical Harmonics Descriptor
	PCA Spherical Harmonics Trns.
	Spherical Trace Transform
Graph Based	Skeletal Graph Based
	Reeb Graph Based
Hybrid 3D Descriptors	CMVD + STT
	Depth-Buffer + Silhouette + REXT
	SIFT + Bag of Features
	Depth-Buffer + Spherical Harmonics

A. *View Based Descriptors (VBD)*

View based descriptors (VBD) use silhouette, greyscale or depth-buffer images extracted from multiple views of 3D objects. According to Liu [22], the advantages of a VBD are twofold: 1) It does not require the explicit virtual model information, which makes the method robust to real practical applications. 2) The view-based 3D model analysis methods can be benefited from existing image processing technologies which has been studied from several decades. These views are then used in the second step to retrieve 3D models from the database which best match the photos and is referred to as the clustering problem by K-means. The smaller set is found using Bayesian Information Criteria (BIC) which gives scores to the sets and the one with the highest score is the most optimal characteristic views set. The third step employs Bayesian Probability distributions for indexing and retrieving 3D models based on the input photos. The authors found that the Light Field Descriptor (LFD) algorithm is 2nd best in terms of retrieval accuracy and performance.

1) Compact Multi-View Descriptor (CMVD): In [25] and [27], Daras and Axenopoulos have presented a new method for 3D shape retrieval called Compact Multi-View Descriptor (CMVD). Their method accepts multi-modal queries (2D images, sketches, and 3D models). The first step of the method is pose estimation in which Principal Component Analysis (PCA) and Visual Contact Area (VCA) methods are used to estimate the pose and apply rotation invariance to the 3D object. The next step is to generate 24 sets of 2D image views of the 3D object from 18 different viewpoints of a 32-hedron surrounding the 3D object, by rotating the object 24 times in 90 degrees along the three principal axes. Two types of views are extracted, binary views (silhouette views) and depth views. After extraction of the 2D views, the method generates 2D shape descriptors of the extracted views by applying three rotation invariant 2D functionals namely 2D Polar Fourier Transforms, 2D Zernike Moments, and Krawtchouk Moments [26]. The final step, 3D to 3D matching is achieved by computing the total dissimilarity of the extracted 2D views of the two 3D objects. 2D to 3D matching is achieved by computing the dissimilarity of the 2D image query and the extracted 2D view of the 3D object that is the most similar to the query 2D image. According to the authors, the CMVD method when combined with the Spherical Trace Transform (STT) algorithm outperforms other well-known algorithms.

2) Light-Field Descriptor (LFD): The Light-field Descriptor (LFD) was introduced by Chen et al. [28], which is based on the idea that two 3D objects are similar if they look similar from all viewing angles. In their approach, ten silhouette images are taken from 10 viewing angles distributed evenly on a dodecahedron. To extract the features of the silhouette images, they used Zernike moments and Fourier descriptors. To calculate the dissimilarity, they found the minimal dissimilarity obtained from rotating the viewing spheres of one lightfield descriptor relative to the other lightfield descriptor. Experiments indicate that LFD performs better than Spherical Harmonics Descriptors [29]. Advantages of LFD include invariant to translation, scale and rotation, and robust against noise and degeneracies.

B. *Histogram Based Descriptors (HBD)*

Histogram based descriptors collect the features of a 3D shape in numerical values in bins defined over the feature domain.

1) 3D Shape Spectrum Descriptor (3D SSD): A 3D shape spectrum descriptor (3D SSD) is a shape descriptor which contains a shape index distributed over the entire mesh [32]. A shape index is defined as a local geometric feature of the shape expressed as the angular coordinate of the polar representation of the principal curvature vector. Regarding the original feature the shape index is invariant to scale and Euclidean transform, and it represents by salient elementary shapes (convex, concave, rut, ridge, saddle, and so on). 3D SSD locally characterizes free form discrete polygon 3D meshes. 3D SSD possesses the characteristics of: (1) generality, since 3D meshes may include open surfaces that have not an associated volume; (2) invariance to scale and Euclidean transforms; (3) robustness that different triangulations of the same object are permitted and it successfully retrieves articulated objects with different postures. Since this descriptor is a simple local feature representation, it should be combined with some global representation schemes.

2) Generalized Shape Distributions (GSD): Liu et al [33] proposed a novel technique for 3D shape retrieval called the Generalized Shape Distributions (GSD), which is a 3D histogram. GSD is based on local and global shape signatures / descriptors of a 3D model. Before generating a GSD histogram, there are some preliminary steps, which involves the generation of a dictionary of local shape descriptors / signatures using Spin Images approach. 50,000 points are sampled on the surface of the 3D shape and these sampled points are then accumulated to create spin images. These spin images are then clustered into 1500 clusters using k-means algorithm. Then each spin image is assigned an index based on the index of its nearest cluster. After these steps, GSD representation is generated. The GSD histogram has three dimensions. The first dimension stores the Euclidean distance of the 2 point pairs, while the other two dimensions store the index value of the two points. This technique has proved to be more accurate and efficient than its predecessor technique 'Shape distributions' [59] and Bag-of-Features technique.

3) Bag-of-Features Histogram: The Bag-of-Features (BoF) technique is a method of accumulating the visual features of a 3D model in a histogram where thousands of visual features are extracted from range images thus improving retrieval efficiency. Moreover bag-of-features has proved to be robust against articulated or non-rigid 3D models.

C. Transform Based Descriptors (TBD)

The Princeton Group and Konstanz Group have undertaken considerable research on transform based descriptors (TBD). The theoretical foundations of TBD are in classical processing such as spherical harmonics, and Fourier transform. Usually, for pose normalization Principal Component Analysis (PCA) [41] is applied as a preliminary step in TBDs.

1) Spherical Harmonics Descriptor (SHD): Funkhouser et al. [29] are the first researchers to use spherical harmonics to describe a 3D shape, and considerable work has been done in this domain following that. With this descriptor, the 3D model is first voxelized with each voxel having either a 1 or 0 value. The binary voxel grid of the 3D model is then placed under concentric spheres and decomposed into spherical functions. Next, a set of harmonic functions are computed from each concentric sphere which are rotation invariant. Each harmonic function is represented as a histogram called the spherical signature. The spherical signatures are then combined to generate a rotation invariant 3D shape descriptor. To compare two spherical harmonics descriptors the authors used Euclidean distance. The disadvantage of SHD is that a 3D model cannot be reconstructed from the feature vector. In [29], Funkhouser et al. avoided the PCA step because they believe that PCA is an unstable approach for pose normalization.

2) PCA Spherical Harmonics Transform: The independent work on spherical harmonic based descriptors carried out by the Konstanz University group is in parallel with the Princeton University group. There has been a debate around whether to use PCA for pose normalization or not. Vranic et al. [41], [42] proposed another spherical based shape descriptor that uses PCA as its pose estimation step and claimed that this descriptor has outperformed the spherical harmonic descriptor proposed by the Princeton Group. Vranic's spherical harmonics descriptor [41] is different from that of Princeton Group [29] in a way that this descriptor involves a generalized PCA step for pose estimation not only considering the vectors and coordinate axes, but also all the points on the mesh with equal weights. For feature extraction, Fourier transform is applied after the spherical harmonics functions are computed which finally generates a feature vector. According to the authors [41], their PCA approach is slightly more expensive but more accurate than the original approach.

3) Spherical Trace Transform Descriptor (STTD): The STTD proposed by Daras et al. [43] is an extension of the 'Trace Transform' with is actually a generalization of 'Radon Transform'. STTD does not employ PCA as its preliminary step but it uses rotation invariant spherical functions to produce a completely rotation invariant shape descriptor. The first step which is the pre-processing stage is to achieve translation and scaling normalization. In this stage the 3D model is placed inside a bounding cube and the cube is partitioned in equal cube shaped voxels. Voxelization allows us to achieve translation and scale normalization. Then, a set of initial 2D functions are applied to the model which creates a set of concentric spheres. After applying initial functions a set of spherical functions are applied to produce the final descriptor vector. For similarity matching, weights are assigned to each descriptor to achieve effectiveness in shape retrieval.

D. Graph Based Descriptors (GBD)

GBDs tend to represent the topology of a 3D shape in the form of a graph or a tree structure. These descriptors are easy to compute. However they are not computationally efficient. According to [4], an advantage of GBDs is that they allow representation at multiple levels of detail and facilitate matching of local geometry.

1) Skeletal Graph Descriptors: Skeletal graph-based techniques utilize a skeletal graph of a shape as its shape descriptor by computing the 'skeleton' for a model. Blum proposed the concept of a skeleton in [34]. A skeleton in 2D is the medial axis, while in 3D it is the medial surface. Several methods have been proposed to perform Skeletonization such as distance transform [35], thinning [36], or Voronoi-based methods [37]. Additionally, curve skeletonization [38] methods have been proposed to convert a 3D model into a medial axis type representation. The skeletal graph stores the various entities obtained after skeletonization in a graph data structure. Advantage of skeletal graph-based methods is topology preserving. Hence, they can be used for subgraph isomorphism at a very low computational cost. Additionally, local part attributes can be stored for a more accurate comparison.

2) Reeb Graph Descriptors: Reeb [39] defined a skeleton structure, called the Reeb graph, which is determined using a continuous scalar function on an object. Three types of scalar functions have been used namely height function, curvature function, and geodesic distance. Geodesic distance has been used in many applications because it provides invariance against rotation and robustness against noise and small perturbations. The function is integrated over the whole body to make it invariant to the starting point and is also normalized to achieve scale invariance.

E. Hybrid 3D Shape Descriptors

Recently researchers in the field of 3D object retrieval have combined several algorithms to produce hybrid algorithms in order to improve the quality of 3D shape retrieval and analysis. Vranic [40] proposed one of the earliest hybrid algorithms, and indicated that the cross breeding of depth-buffer, silhouette, and REXT descriptor (Depth-Buffer + Silhouette + REXT) has superior performance over all other state-of-the-art descriptors.

1) CMVD + STT: Daras and Axenopoulos [25] demonstrated that Compact Multi-View Descriptor (CMVD) when combined with Spherical Trace Transform (STT) performs better than all the algorithms proposed in the past. They tested their hybrid algorithm on the three databases i.e. Princeton Shape Benchmark (PSB) [44], ITI database used in Victory 3D Search Engine [45], and the Engineering Shape Benchmark (ESB) [46]. The authors compared their hybrid algorithm (CMVD + STT) with three other successful algorithms 1) The LightField Descriptor, 2) SIFT + bag of features, and 3) Depth-Buffer + Silhouette + REXT (DSR). The precision-recall results have shown that the proposed hybrid algorithm performs better than all other algorithms.

2) SIFT + Bag of Features (BF-SIFT): Ohbuchi et al [47] have proposed a hybrid descriptor based on extracting local visual features of a 3D model using SIFT algorithm and efficiently integrating them in a histogram using the Bag-of-Features approach. In their algorithm, several 2D range images are obtained from of the 3D model. Then, SIFT algorithm is used to extract local features. Each feature is a vector quantized using a visual codebook. K-means learning is used to cluster the local features into a bag of visual words. Then a histogram is generated using the frequencies of visual words, which acts as the feature vector for the 3D model. Some advantages of BF-SIFT are 1) suitable for articulated models, 2) high discriminative power, 3) suitable for 2D image and sketch based queries, and 4) effective for partial matching.

3) Depth-Buffer + Spherical Harmonics: Papadakis et al. [48] proposed a hybrid descriptor which is composed of depth buffer algorithm for extracting 2D features and spherical harmonics for encoding 3D features. For pose normalization two alignment methods namely CPCA and NPCA are used while compactness of the feature vector is supported via scalar feature quantization to a set of values that is further compressed using Huffman coding. The authors have demonstrated superior performance of the proposed retrieval technique through an extensive comparison against state-of-the-art methods such as LFD and DSR on standard datasets.

V. SHREC - 3D Shape Retrieval Contest

SHREC is an international shape retrieval contest started in 2006, which aims to compare and evaluate 3D shape retrieval algorithms using a benchmark. In 2012 [50] the performance of five hybrid algorithms were compared on a generic shape benchmark (SHREC'12) devised from other well-known benchmarks including PSB, NTU, and CCCC. These algorithms include Local Shape Distribution descriptor, ZFDR (hybrid descriptor based on Zernike, Fourier, depth-buffer, and ray based descriptors) [51], 3D Spatial Pyramid descriptor, Dense Voxel Spectrum descriptor (DVD) [52], and Dense Grid SIFT (DGSIFT) descriptor [53]. The precision-recall results showed that DGSIFT performed the best amongst others followed by DVD and ZFDR respectively.

VI. Results & Comparisons

Yang et al. [54] have compared the computational complexity of all the major categories of 3D shape descriptors. According to their findings, transform based methods have computational complexity as $O(b.\ N^3)$ where b is the number of spherical functions used, and N is the number of voxels in each axis. As compared to these, histogram based methods are faster with $O(t.\ N.\ \log N)$ complexity, where t is the number of histogram bins and N is the number of mesh vertices in each axis. However, transform based methods have higher discriminative power than histogram based methods. Graph based methods capture shape features very well however they are computationally expensive with $O(N.\ \log N)$ complexity. Histogram based descriptors are compact are compact, fast, and very robust, and are therefore suitable for hybrid descriptors, an example being the Bag-of-Features (BoF) approach.

VII. Conclusion & Future Research Directions

Over the past few years, the research in content based 3D model retrieval has moved towards inventing and utilizing hybrid descriptors, and dealing with objects that are non-rigid or articulated. 3D model retrieval is still in its infancy and much work needs to be done to improve the accuracy of the techniques. Moreover, semantic based retrieval has also become a hot topic amongst researchers in recent times. Hybrid techniques have proved to be more successful than the original individual techniques.

Researchers have found that region based 2D shape descriptors are great tools for local feature matching in 3D shape retrieval. Amongst the 3D SDs, graph based algorithms are the least preferred algorithms because of their performance inefficiency. View based and histogram based algorithms are the most favourite amongst the researchers due to their decency in accuracy and performance.

Much debate has happened between global feature based matching and local feature based matching. Global feature based methods take the whole geometry into consideration and produce a feature vector describing the whole shape. Thus they are not suitable for local or partial matching. However these are easy and faster to compute. On the other hand, techniques like SIFT have paved a way for robust matching of local features.

Bustos et al. [56] identified that there is currently a lack of databases dedicated to domain specific models such as models of human characters of several levels of details. This could be a very important future direction for benchmark databases. Moreover, techniques to retrieve articulated human character models (models in different poses) also need to be further investigated.

In surveys by Yang et al. [54] and Qin et al. [55], the authors have identified the following areas that deserve future investigation.

a) 3D models these days are available in different formats however there is a lack of algorithms that extracts features directly compressed format of the 3D models. Partial matching of 3D objects that utilize the local feature of a model at different resolutions requires different kinds of feature vectors.

b) Graph based techniques need to be improved in terms of computation complexity and efficiency. Graph based techniques can describe an object very accurately on several levels of details, however current approaches take topology of the mesh into account which is computationally impractical. Moreover graph comparison is proportional to graph-size.

c) Non-shape features of a model such as color, texture, and material should be investigated further.

d) The animation industry needs artist friendly and intuitive tools for creating 3D models and currently the interfaces are not much artist friendly.

e) Very few techniques exist that retrieve 3D objects from scenes containing multiple 3D objects. This requires a novel hierarchical object structure to be investigated to recognize 3D objects from cluttered scenes.

ACKNOWLEDGMENT

This research is supported by the grant of UK Royal Society International Joint Projects / NSFC 2010.

REFERENCES

[1] M. Yang, K. Kpalma, and J. Ronsin, "A survey of shape feature extraction techniques," Pattern Recognition, vol. 2008.

[2] J. W. H. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," Multimedia Tools and Applications, vol. 39, no. 3, pp. 441–471, Dec. 2007.

[3] D. Zhang and G. Lu, "Review of shape representation and description techniques," Pattern Recognition, vol. 37, no. 1, pp. 1–19, Jan. 2004.

[4] N. Iyer, S. Jayanti, K. Lou, Y. Kalyanaraman, and K. Ramani, "Three-dimensional shape searching: state-of-the-art review and future trends," Computer-Aided Design, vol. 37, no. 5, pp. 509–530, Apr. 2005.

[5] Zhang, Dengsheng, and Guojun Lu. "A comparative study on shape retrieval using Fourier descriptors with different shape signatures," In Proc. of international conference on intelligent multimedia and distance education (ICIMADE01), 2001, pp. 1-9.

[6] G. Chuang and C. Kuo, "Wavelet descriptor of planar curves: Theory and applications," IEEE Transactions on Image Processing, vol. 5, no. 1, 1996.

[7] Berrada, F., D. Aboutajdine, S. E. Ouatik, and A. Lachkar. "Review of 2D shape descriptors based on the curvature scale space approach," In Multimedia Computing and Systems (ICMCS), 2011 International Conference on, 2011, pp. 1-6. IEEE.

[8] Bober, and Miroslaw. "MPEG-7 visual shape descriptors," Circuits and Systems for Video Technology, IEEE Transactions on 11, no. 6 (2001): 716-719.

[9] C. Singh and Pooja. "An Effective Image Retrieval System using Region and Contour based Features," IJCA Proceedings on International Conference on Recent Advances and Future Trends in Information Technology (iRAFIT 2012) iRAFIT(2):7-12, April 2012.

[10] G. Zhang, Z. Ma, L. Niu, and C. Zhang, "Modified Fourier descriptor for shape feature extraction," Journal of Central South University of Technology, pp. 488–495, 2012.

[11] S. Khoje and S. Bodhe, "Performance Comparison of Fourier Transform and its Derivatives as Shape Descriptors for Mango Grading," International Journal of Computer Applications, vol. 53, no. 3, pp. 17–22, Sep. 2012.

[12] H. Asada and M. Brady, "The curvature primal sketch," Pattern Analysis and Machine Intelligence, IEEE Transactions on, no. 1, pp. 2–14, 1986.

[13] F. Mokhtarian, S. Abbasi, and J. Kittler, “Robust and E cient Shape Indexing through Curvature Scale Space,” Proceedings of the 1996 British Machine and Vision Conference BMVC, 1996.

[14] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[15] M. Teague, “Image analysis via the general theory of moments,” Optical Society of America, Journal, 70(8), 920-930, 1980.

[16] S. Liao and M. Pawlak, “On the accuracy of Zernike moments for image analysis,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 20, no. 12, pp. 1358–1364, 1998.

[17] F. Zernike., “Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode,” Physica, vol. 1, pp. 689-701, 1934.

[18] M. E. Celebi and Y. a. Aslandogan, “A comparative study of three moment-based shape descriptors,” International Conference on Information Technology: Coding and Computing (ITCC’05) - Volume II, pp. 788–793 Vol. 1, 2005.

[19] D. G. Lowe, “Object recognition from local scale-invariant features,” Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, pp. 1150–1157 vol.2.

[20] Schmid, Cordelia, and R. Mohr. "Local grayvalue invariants for image retrieval," Pattern Analysis and Machine Intelligence, IEEE Transactions on 19, no. 5: 530-535, 1997

[21] W.-Y. Kim, and Y.-S. Kim, "A New Region-Based Shape Descriptor," ISO/IEC MPEG99/M5472, Maui, Hawaii, December 1999.

[22] Q. Liu, “A Survey of Recent View-based 3D Model Retrieval Methods,” arXiv preprint arXiv:1208.3670, pp. 1–15, 2012.

[23] T. Ansary, M. Daoudi, and J. Vandeborre, “3D model retrieval based on adaptive views clustering,” Pattern Recognition and Image Analysis, no. Icapr, 2005.

[24] Ansary, Tarik Filali, Mohamed Daoudi, and Jean-Phillipe Vandeborre. "3D model retrieval based on adaptive views clustering," In Pattern Recognition and Image Analysis, pp. 473-483. Springer Berlin Heidelberg, 2005.

[25] P. Daras and A. Axenopoulos, “A 3D Shape Retrieval Framework Supporting Multimodal Queries,” International Journal of Computer Vision, vol. 89, no. 2–3, pp. 229–247, Jul. 2009.

[26] P.-T. Yap, R. Paramesran, and S.-H. Ong, “Image analysis by Krawtchouk moments,” IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, vol. 12, no. 11, pp. 1367–77, Jan. 2003.

[27] P. Daras and A. Axenopoulos, “A Compact Multi-view Descriptor for 3D Object Retrieval,” 2009 Seventh International Workshop on Content-Based Multimedia Indexing, pp. 115–119, Jun. 2009.

[28] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, “On Visual Similarity Based 3D Model Retrieval,” Computer Graphics Forum, vol. 22, no. 3, pp. 223–232, Sep. 2003.

[29] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs, “A search engine for 3D models,” ACM Transactions on Graphics, vol. 22, no. 1, pp. 83–105, Jan. 2003.

[30] L. Zhang, M. Fonseca, and A. Ferreira, “Survey on 3D shape descriptors,” 2007.

[31] C. Akgül, “Density-based shape descriptors and similarity learning for 3D object retrieval,” 2007.

[32] T. Zaharia and F. Preteux., "Three-dimensional shape-based retrieval within the MPEG-7 framework," In Proceedings SPIE Conference on Nonlinear Image Processing and Pattern Analysis XII, San Jose, CA, January 2001, Vol. 4304, pp. 133-145.

[33] Y. Liu, H. Zha, and H. Qin, “The generalized shape distributions for shape matching and analysis,” proc. IEEE int'l conf. shape modelling and applications, june 2006.

[34] H. Blum, "A transformation for extracting new descriptors of shape," in Models for the Perception of Speech andVisual Forms, W. Whaten-Dunn (Ed.), MIT Press, Cambridge, MA, pp. 362–380, 1967.

[35] G. Borgefors, "Distance transformations in arbitrary dimensions," Comput Vis, Graph, Image Process. 27:321–45, 1984.

[36] L. Lam, S. W. Lee, and C. Y. Suen, "Thinning methodologies: a comprehensive survey," IEEE Trans Pattern Anal Mach Intell. 14(9): 869–85, 1992.

[37] R. L. Ogniewicz, and O. Kubler, "Hierarchical voronoi skeletons. Pattern Recogn," 28(3):343–59, 1995.

[38] Sanniti di Baja, Gabriella, and S. Svensson. "A new shape descriptor for surfaces in 3D images," Pattern Recognition Letters 23, no. 6: 703-711, 2002.

[39] G. Reeb, "Sur les Points Singuliers d’une Forme de Pfaff Completement Integrable ou d’une Fonction Numerique (On the singular points of a completely integrable Pfaff form or of a numerical function)," Comptes Randus Acad Sci Paris, 222:847–9, 1946.

[40] D. V. Vrani¢. "3D Model Retrieval," PhD thesis, University of Leipzig, 2004.

[41] D. Vranic, D. Saupe, and J. Richter, “Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics,” In Proc. of the IEEE fourth workshop on multimedia signal processing (pp. 293–29), 2001.

[42] D. Vranic, “An improvement of rotation invariant 3D-shape based on functions on concentric spheres,” in Proceedings of 2003 International Conference on Image Processing, ICIP 2003, Vol. 3, pp. 757–760.

[43] P. Daras and D. Zarpalas, “3D model search and retrieval based on the spherical trace transform,” In Multimedia Signal Processing, 2004 IEEE 6th Workshop on, 2004, pp. 335–338.

[44] P. Shilane and P. Min, “The princeton shape benchmark,” Shape Modeling Applications, 2004. Proceedings, vol. 08540, 2004.

[45] P. Daras, D. Tzovaras, and S. Dobravec, “Victory: a 3d search engine over p2p and wireless p2p networks,” Proceedings of the 4th Annual International Conference on Wireless Internet, 2008.

[46] S. Jayanti, Y. Kalyanaraman, N. Iyer, K. Ramani, "Developing an engineering shape benchmark for CAD models," Computer-Aided Design 38 (9) 939–953, 2006.

[47] R. Ohbuchi, “Salient local visual features for shape-based 3D model retrieval,” 2008 IEEE International Conference on Shape Modeling and Applications, pp. 93–102, Jun. 2008.

[48] P. Papadakis and I. Pratikakis, “3D object retrieval using an efficient and compact hybrid shape descriptor,” 3D object retrieval using an efficient and compact hybrid shape descriptor. In Eurographics Workshop on 3D object retrieval, 2008.

[49] Wei, Chia-Hung, Yue Li, Wing-Yin Chau, and Chang-Tsun Li. "Trademark image retrieval using synthetic features for describing global shape and interior structure," Pattern Recognition 42, no. 3, 386-394, 2009.

[50] B. Li, A. Godil, M. Aono, X. Bai, and T. Furuya, “SHREC’12 Track: Generic 3D Shape Retrieval,” Eurographics 2012 Workshop on 3D Object Retrieval, pp. 119–126, 2012.

[51] B. Li and H. Johan, “3D model retrieval using hybrid features and class information,” Multimedia Tools and Applications, Sep. 2011.

[52] A. Tatsuma, M. Aono, "Multi-Fourier spectra descriptor and augmentation with spectral clustering for 3D shape retrieval," Vis. Comput. 25, 785–804, 2009.

[53] R. Ohbuchi and T. Furuya, “Distance metric learning and feature combination for shape-based 3D model retrieval,” Proceedings of the ACM workshop on 3D object retrieval - 3DOR ’10, p. 63, 2010.

[54] Y. Yang, H. Lin, and Y. Zhang, “Content-Based 3-D Model Retrieval: A Survey,” IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 37, no. 6, pp. 1081–1098, Nov. 2007.

[55] Z. Qin, J. Jia, and J. Qin, “Content based 3D model retrieval: A survey,” 2008 International Workshop on Content-Based Multimedia Indexing, pp. 249–256, 2008.

[56] B. Bustos, D. Keim, D. Saupe, and T. Schreck, "Content-based 3D object retrieval," Computer Graphics and Applications, IEEE, 27(4), 22-27, 2007.

[57] Zhang, Juan, K. Siddiqi, D. Macrini, A. Shokoufandeh, and S. Dickinson. "Retrieving articulated 3-d models using medial surfaces and their graph spectra," In Energy minimization methods in computer vision and pattern recognition, pp. 285-300. Springer Berlin Heidelberg, 2005.

[58] C. Singh, "Local and global features based image retrieval system using orthogonal radial moments," Optics and Lasers in Engineering, 2011.

[59] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin,. "Shape distributions," ACM Transactions on Graphics (TOG), 21(4), 807-832, 2002.