

Density-Based Shape Descriptors for 3D Object Retrieval

Ceyhun Burak Akgül^{1,2}, Bülent Sankur¹, Francis Schmitt², and Yücel Yemez³

¹ Department of Electrical and Electronics Engineering,
Boğaziçi University, Istanbul, Turkey

² GET-Télécom - CNRS UMR 5141, France

³ Department of Computer Engineering, Koç University, Istanbul, Turkey

Abstract. We develop a probabilistic framework that computes 3D shape descriptors in a more rigorous and accurate manner than usual histogram-based methods for the purpose of 3D object retrieval. We first use a numerical analytical approach to extract the shape information from each mesh triangle in a better way than the sparse sampling approach. These measurements are then combined to build a probability density descriptor via kernel density estimation techniques, with a rule-based bandwidth assignment. Finally, we explore descriptor fusion schemes. Our analytical approach reveals the true potential of density-based descriptors, one of its representatives reaching the top ranking position among competing methods.

1 Introduction

There is a growing interest in 3D shape classification, matching and retrieval as 3D object models become more commonplace in various domains such as computer-aided design, medical imaging, molecular analysis and digital preservation of cultural heritage. The research efforts in this field mainly focus on judicious design of discriminating shape features and on pragmatic computational schemes. Representations used for shape matching are referred to as *3D shape descriptors*, which are usually based on direct shape features or some function of these features [1,2].

We present a framework for 3D shape description based on probability density function of shape features. We first define a geometric feature over the surface of the 3D object. This geometric feature can be a scalar or a vector, and it is intended to measure a local property of the 3D surface. In this work, we limit ourselves to triangular mesh representations, however the proposed features can be computed for point cloud representations as well. We calculate the geometric feature on each triangle of the mesh and obtain a set of observations, each providing a local characterization. Using the set of observations and kernel density estimation (KDE) [3], we then estimate the probability density of the local geometric feature at target points chosen on the domain of the feature. The vector of the estimated density values becomes our 3D shape descriptor. This density-based approach collects local evidence about the shape information and then,

using KDE, it accumulates this evidence at target points so as to end up with a global shape description.

In the previous works on 3D shape descriptors, the idea of gathering and accumulating local surface information is implemented by histograms [1,2,4,5,6,7,8]. Paquet et al. use the cord length and the angles between a cord and the principal axes as geometric features to construct univariate histograms [4]. The resulting 3D shape descriptor consists of concatenated univariate histograms, called *Cord and Angle Histograms*. Osada et al. follow a random sampling approach to acquire a large set of observations so as to measure a global property of the surface, such as the Euclidean distance between two surface points (D2) [5]. Among other histogram-like approaches, *Extended Gaussian Images* (EGI) and their variants [6,7] are based on the distribution of surface normals over a spherical grid. These are not true histograms in the rigorous sense of the term, but they share the philosophy of accumulating a geometric feature. The *3D Hough Transform Descriptor* (3DHT), presented in [8], is based on the parameterization of the local tangent plane. The 3DHT can be viewed as a generalization of EGI. We have experimentally verified the conjecture that the 3DHT descriptor captures the shape information better than the EGI descriptor in [9].

The main motivation of the present work is to develop a probabilistic framework to compute histogram-based descriptors in a more rigorous and accurate manner by the KDE technique. The resulting framework is a general one that can be applied to any local feature vector of any dimension. In the light of the proposed framework, we also reformulate the existing local shape features discussed above in order to achieve an improved shape characterization. These features, when combined with a new set of shape features that we propose, result in shape descriptors that outperform all of its histogram-based competitors existing in the literature.

2 Local Geometric Features

We assume that each 3D shape is represented with a triangular mesh and that its center of mass coincides with the origin of the coordinate system. In what follows, capital italic letter P stands for a point in 3D, a small case boldface letter \mathbf{p} for its vector representation, \mathbf{n}_P for the unit surface normal vector at P whenever P is an element of some surface $\mathcal{M} \subset \mathbb{R}^3$ and $\langle \cdot, \cdot \rangle$ for the usual dot product. We define a local geometric feature, as a mapping S from the points of a surface \mathcal{M} into a d -dimensional space, usually constrained into a finite subspace of \mathbb{R}^d . Each dimension of this space corresponds to a specific geometric measure, characterizing the shape locally. In this work, we consider three different sorts of multidimensional local geometric features, as introduced next.

The *radial feature* S_r at a point P is a 4-tuple defined as $S_r(\mathbf{p}) \triangleq (r, \mathbf{r})$ where $r \triangleq \|\mathbf{p}\|$ and $\mathbf{r} = (r_x, r_y, r_z) \triangleq \mathbf{p} / \|\mathbf{p}\|$. S_r consists of a *magnitude* component r , measuring the distance of the point P to the origin; and a *direction* component \mathbf{r} , pointing to the location of the point P . The direction component \mathbf{r} is a 3-vector with unit-norm; hence it lies on the unit sphere.

The *tangent plane-based feature* S_t at a point P is a 4-tuple defined as $S_t(\mathbf{p}) \triangleq (d_t, \mathbf{n}_P)$ where $d_t \triangleq |\langle \mathbf{p}, \mathbf{n}_P \rangle|$ and $\mathbf{n}_P = (n_{P,x}, n_{P,y}, n_{P,z})$. Similar to the S_r -feature, S_t has a magnitude component d_t , which stands for the distance of the tangent plane at P to the origin, and a direction component \mathbf{n}_P . The normal \mathbf{n}_P is a unit-norm vector by definition and lies on the unit sphere.

The *cross-product feature* S_c at a point P is defined as $S_c(\mathbf{p}) \triangleq (r, \mathbf{c}_P)$ where $\mathbf{c}_P = (c_{P,x}, c_{P,y}, c_{P,z}) \triangleq \mathbf{p} \times \mathbf{n}_P$. This third feature encodes the interaction between the first two features above, namely, the radial feature S_r and the tangent plane feature S_t . In much the same way as in S_r and S_t , S_c is decoupled into its magnitude component r and its direction component \mathbf{c}_P . Notice however that \mathbf{c}_P , according to its definition, is not generally a unit-norm vector. Its norm satisfying $0 \leq \|\mathbf{c}_P\| \leq 1$, \mathbf{c}_P lies inside the unit ball.

Note that the above three features are neither scale- nor rotation-invariant. Accordingly, any method making use of them must assume prior scale and pose normalization of the mesh.

3 Density-Based Shape Description

Given a set of observations $\{s_k\}_{k=1}^K$ for a random variable (scalar or vector) S , the kernel approach to estimate the probability density function (pdf) f_S of S is formulated in its most general form as

$$f_S(s) = \sum_{k=1}^K w_k |H_k|^{-1} \mathcal{K}(H_k^{-1}(s - s_k)). \quad (1)$$

where $\mathcal{K} : \mathbb{R}^d \rightarrow \mathbb{R}$ is a kernel function, H_k is a design parameter called the $d \times d$ bandwidth matrix and w_k is the importance weight associated with the k th observation. We intend to apply this classical kernel scheme to derive probability distributions of 3D shape features. In this context, $\{s_k\}_{k=1}^K$ correspond to measurements of some local geometric feature S , and the array of f_S -values at a predefined set of target points corresponds to the descriptor. For a triangular mesh consisting of K triangles, we can obtain an observation s_k from each of the triangles, as described in Sect. 3.1. A natural choice for an importance weight w_k is the area of the k th triangle relative to the total mesh area. It is known that estimates in Eq. 1 are sensitive to the bandwidth parameters $\{H_k\}_{k=1}^K$ rather than the particular kernel used [3]. In our application, we have chosen the Gaussian kernel. The availability of a fast algorithm was the determining factor for this choice. We address the bandwidth selection issue in Sect. 3.2.

3.1 Feature Calculation

A shape descriptor can be estimated by using samples of shape features over the mesh triangles. Previous studies [1,4] have considered a single sample per triangle, namely the triangle barycenter. We claim this barycentric sampling is not the best option because of possible shape and size non-uniformities of triangles. Instead, we propose an estimate taking into consideration the multitude of

points uniformly distributed over the triangle geometry. In other words, we will replace s_k in Eq. 1 with the expectation of the local feature value over the k th triangle $E\{S|\textit{kth triangle}\}$ instead of the value just sampled at its barycenter [9]. This moment estimate can be obtained as follows. Let T be an arbitrary triangle in 3D space with vertices A , B , and C represented by \mathbf{p}_A , \mathbf{p}_B and \mathbf{p}_C respectively. By taking any one vertex as a pivot (say, pivot A), the relative coordinates of an arbitrary point P inside the triangle T can be expressed in terms of $\mathbf{e}_1 = \mathbf{p}_B - \mathbf{p}_A$ and $\mathbf{e}_2 = \mathbf{p}_C - \mathbf{p}_A$, as $\mathbf{p} = \mathbf{p}_A + x\mathbf{e}_1 + y\mathbf{e}_2$, where $x, y \geq 0$ and $x + y \leq 1$. Assuming that points $\{P\}$ are uniformly distributed inside the triangle T , each feature S can be expressed as a function of two variables (x, y) , i.e., $S(\mathbf{p}) = S(x, y)$. Thus, the expected local feature value over the triangle T reads as

$$E\{S|T\} = \iint_{\Omega} S_i(x, y) f(x, y) dx dy \quad (2)$$

where $f(x, y)$ is the bivariate uniform density of the pair (x, y) over the domain $\Omega = \{(x, y) : x, y \geq 0, x + y \leq 1\}$. To approximate Eq. 2, we apply Simpson's 1/3 numerical integration formula [10]. To remove the arbitrariness of the pivot, we compute the integral with respect to each pivot A, B , and C . Finally, we average the three integration results to obtain

$$\begin{aligned} E\{S|T\} \approx & (1/27)(S(\mathbf{p}_A) + S(\mathbf{p}_B) + S(\mathbf{p}_C)) \\ & + (4/27)(S((\mathbf{p}_A + \mathbf{p}_B)/2) + S((\mathbf{p}_A + \mathbf{p}_C)/2) + S((\mathbf{p}_B + \mathbf{p}_C)/2)) \\ & + (4/27)(S((2\mathbf{p}_A + \mathbf{p}_B + \mathbf{p}_C)/4) + S((\mathbf{p}_A + 2\mathbf{p}_B + \mathbf{p}_C)/4) + S((\mathbf{p}_A + \mathbf{p}_B + 2\mathbf{p}_C)/2)) \end{aligned} \quad (3)$$

3.2 Bandwidth Selection

The KDE formulation given in Eq. 1 gives us the liberty to set a different bandwidth matrix H_k for each triangle in a given mesh. With this richness of choice, no assumption needs to be made about the shape of the kernel function or implicitly about the shape of the k th triangle. However, the fast Gauss Transform (FGT) algorithm in [11] precludes the use of a bandwidth matrix H_k per triangle. The computation of the sum in Eq. 1 without resorting to a fast transform leads to prohibitive computational load. To give an idea, for example, on a Pentium 4 PC (2.4 GHz CPU, 2 GB RAM), for a mesh of 130,000 triangles, direct evaluation of the S_r -descriptor (1024-point pdf) takes 125 seconds against the 2.5-second computation time with FGT. Accordingly, we adopt a fixed form of $H_k = H$, i.e., the bandwidth matrix does not vary across the triangles. This can be done in two ways: either at the mesh level, in which case every mesh will be attributed its own bandwidth matrix, or at the database level, in which case a single H is valid for all meshes. At the mesh level, the bandwidth matrix for a given feature can be set by Scott's rule of thumb [3]: $H_{\text{Scott}} = (\sum_k w_k^2)^{1/d+4} C^{1/2}$, where d is the feature dimension and C is the estimate of the feature covariance matrix. At the database level, we consider the average Scott bandwidth matrix over the 3D meshes in the database. In our experiments, we have tested these two options one against the other by comparing their retrieval performances.

4 Experiments

We have tested our descriptors in a retrieval scenario on the Princeton Shape Benchmark (PSB) [2], which consists of 3D objects described as triangular meshes. PSB is a publicly available database containing a total of 1814 models, categorized into general classes and consisting of two equally sized sets: one for training and another for testing purposes. We present the retrieval performance of descriptors using precision-recall curves and discounted cumulative gain (DCG) values [2]. Recall that DCG is a statistic that weights correct results near the front of the list more than correct results appearing later in the ranked list.

We applied the following normalization to the meshes to secure invariance to translation, scale, and rotation. For translation invariance, the object’s center of mass was translated to the origin. For scale invariance, the area-weighted average distance of surface points to the origin was set to unity. Finally, to guarantee rotation and flipping invariance, we have used the continuous PCA approach [12]. We have taken 1024 target points within the domain of definition of S_r and S_t and 2560 for S_c . Finally, we observed that the Minkowski l_1 -metric yielded the best retrieval statistics among its alternatives such as l_2 , l_∞ , χ^2 , etc. We have also found it useful to normalize descriptors to unit l_1 -norm.

4.1 Bandwidth Selection Strategy

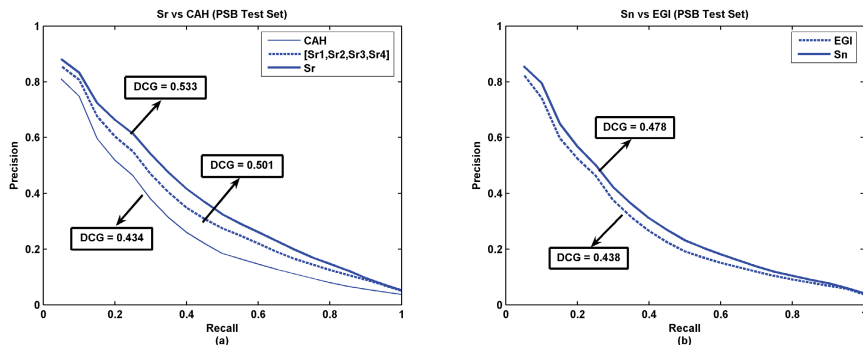
One of the core concerns in our algorithm was the judicious setting of the bandwidth parameters. Due to the FGT constraint, it was pointed out in Sect. 3.2 that it is necessary to operate on a database or mesh basis, but not on a triangle basis. We tested the mesh and database alternatives with our local features S_r , S_t , and S_c , always with Scott’s rule-of-thumb. Since we observed that the off-diagonal terms of the matrices in the Scott’s rule were negligible as compared to the diagonal terms, we decided to use only diagonal bandwidth matrices. Table 1 displays the comparison of DCG scores for S_r , S_t , and S_c on the PSB Training Set. It is clear that setting H at the database level is more advantageous compared to the mesh level setting. The results reported in the following experiments are therefore for the database level setting.

4.2 Density-Based Versus Histogram-Based Descriptors

In this section, we demonstrate the performance advantage of the proposed KDE-based approach compared to the histogram-based analogues in the literature. A test case is Cord and Angle Histograms (CAH) [4]. The features in CAH are identical to our S_r -feature up to a parameterization. The CAH-descriptor consists of the concatenation of cord length and angle histograms. We first applied our framework in Eq. 1 to the individual components of S_r . The resulting descriptor, denoted as $[S_{r,1}, S_{r,2}, S_{r,3}, S_{r,4}]$, consists of the concatenation of univariate densities. In Figure 1(a), we provide the precision-recall curve corresponding to CAH and $[S_{r,1}, S_{r,2}, S_{r,3}, S_{r,4}]$ on PSB Test Set. The respective DCG values are 0.434 and 0.501, indicating the superior performance of our framework under identical

Table 1. DCG values for *Mesh-level* and *Database-level* setting of the bandwidth matrix on PSB Training set

	S_r	S_t	S_c
Mesh-level DCG	0.511	0.514	0.499
Database-level DCG	0.541	0.567	0.543
Performance Gain (%)	6	11	11

**Fig. 1.** (a) Precision-recall curves comparing S_r and $[S_{r,1}, S_{r,2}, S_{r,3}, S_{r,4}]$ to CAH-descriptor, (b) Precision-recall curves comparing S_n to EGI-descriptor

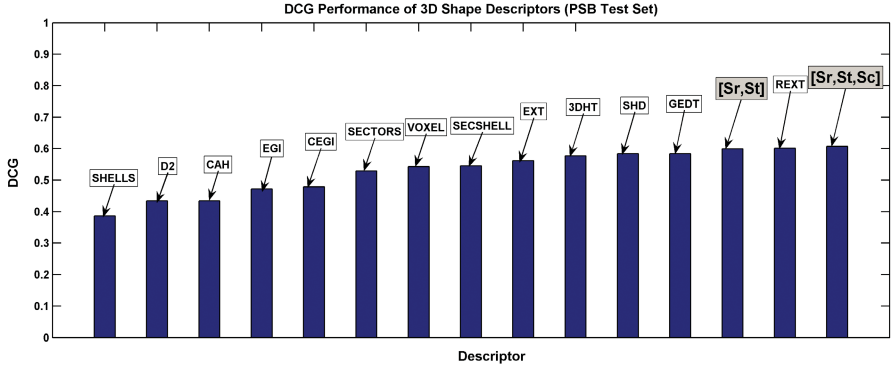
feature sets. An additional improvement can be gained by estimating the joint density of S_r . That is, we directly use the joint density of S_r as a descriptor. The DCG value of the S_r -descriptor is 0.533, one more step of improvement to the univariate case (DCG = 0.501). A second instance of our framework outperforming its competitor is the EGI-descriptor [1,2,6,7], which consists of binning the surface normals. The density of the direction component \mathbf{n}_P of our S_t -feature is equivalent to the EGI-descriptor. The S_n -descriptor ($S_n(\mathbf{p}) \triangleq \mathbf{n}_P$) achieves a DCG of 0.478 compared to the DCG score of 0.438 for EGI (see Figure 1(b)).

4.3 General Performance of Density-Based Descriptors

In this section, we discuss individual performances of the three proposed descriptors S_r , S_t , and S_c , and explore their fusion alternatives. As shown in Table 2, the proposed local features yield similar DCG performance scores on the PSB Test Set. We can observe in the same table that their pair-wise concatenations $[S_r, S_t]$, $[S_r, S_c]$, and $[S_c, S_t]$ increase the DCG scores significantly. Furthermore, the triple-wise concatenation boosts the DCG performance further. In fact, based on the scores reported in [2], the $[S_r, S_t, S_c]$ -descriptor has the highest DCG score among all other well-known 3D shape descriptors, as shown in Figure 2. Except for the 3D Hough Transform Descriptor (3DHT) [8] and CAH [4], all the descriptor scores shown in Figure 2 are taken from [2]. Due to space limitations, we refer the reader to [2] for brief descriptions and acronyms of these descriptors. The

Table 2. DCG Performance of Density-based Descriptors on PSB Test Set

	S_r	S_t	S_c	$[S_r, S_t]$	$[S_r, S_c]$	$[S_t, S_c]$	$[S_r, S_t, S_c]$
DCG	0.533	0.543	0.533	0.599	0.579	0.585	0.607
Size	1024	1024	2560	2048	3584	3584	4608

**Fig. 2.** Comparison of 3D shape descriptors on PSB Test Set (Except CAH, 3DHT, and our descriptors, DCG values are taken from [2].)

$[S_r, S_t, S_c]$ -descriptor developed in this work has a DCG value of 0.607, while the next best descriptor REXT (Radialized Extent Function) [12] has a DCG value of 0.601 [2]. Note also that the $[S_r, S_t]$ -descriptor (DCG = 0.599) comes third in the competition. The average REXT-descriptor size reported in [2] is 17.5 kilobytes, while for our $[S_r, S_t, S_c]$ -descriptor this figure is 22 kilobytes. The average generation time for the REXT-descriptor is 2.2 seconds [2], while our $[S_r, S_t, S_c]$ -descriptor can be computed in 0.9 second on the average.

5 Discussion and Conclusion

We have analyzed and experimented with a new 3D object description and retrieval method. In the analysis framework we developed, we have limited ourselves to the first order local shape features. The features are local in the sense that they measure a property of the surface point by point, without taking into consideration the information about their neighbors. The three feature sets are fairly representative of such first order feature varieties.

We have shown first that probability distribution-based shape descriptors benefits significantly from kernel based estimation in contrast to the histogram-based shape descriptors. Second, the kernel estimates become more informative if a numerical-analytical approach is used in contrast to pure barycentric sampling. Third, the retrieval performance significantly improves using descriptor fusion. We have shown that with all these enhancements, our scheme has climbed on

the competition ladder to the top position in its category, i.e., the one of purely 3D descriptors. Two pieces of wisdom gathered by these experiments are as follows: (i) Features involving surface normals are more informative; (ii) bandwidth parameter per database is more useful as compared to per-mesh setting.

Future research will concentrate on potential improvements of decision fusion. A second natural avenue of research is in the direction of second- and higher-order features, that is, features using the neighborhood of a given triangle. Finally, we plan to test the triangle-based bandwidth selection strategy.

References

1. Tangelder, J.W.H., Veltkamp, R.C.: A survey of content based 3D shape retrieval methods. In: Proc. of the Shape Modeling International 2004 (SMI '04), Genoa, Italy (2004) 145–156
2. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The Princeton shape benchmark. In: Proc. of the Shape Modeling International 2004 (SMI '04), Genoa, Italy (2004) 167–178
3. Härdle, W., Müller, M., Sperlich, S., Werwatz, A.: Nonparametric and Semiparametric Models. Springer Series in Statistics. Springer (2004)
4. Paquet, E., Rioux, M.: Nefertiti: a query by content software for three-dimensional models databases management. In: Proc. of the International Conference on Recent Advances in 3-D Digital Imaging and Modeling (NRC '97), Washington, DC, USA, IEEE Computer Society (1997) 345
5. Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D.: Shape distributions. ACM Trans. Graph. **21** (2002) 807–832
6. Horn, B.K.P.: Extended Gaussian images. Proc. of the IEEE **72** (1984) 1671–1686
7. Kang, S.B., Ikeuchi, K.: The complex EGI: A new representation for 3D pose determination. IEEE Trans. Pattern Anal. and Mach. Intell. **15** (1993) 707–721
8. Zaharia, T., Prêteux, F.: Indexation de maillages 3D par descripteurs de forme. In: Actes 13ème Congrès Francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA'2002), Angers, France. (2002) 48–57
9. Akgül, C.B., Sankur, B., Yemez, Y., Schmitt, F.: A framework for histogram-induced 3D descriptors. In: European Signal Processing Conference (EUSIPCO '06), Florence, Italy (2006)
10. Press, W.H., Flannery, B.P., Teukolsky, S.A.: Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press (1992)
11. Yang, C., Duraiswami, R., Gumerov, N.A., Davis, L.: Improved fast Gauss transform and efficient kernel density estimation. ICCV **1** (2003) 464
12. Vranić, D.V.: 3D Model Retrieval. PhD thesis, University of Leipzig (2004)