

Directional Histogram Model for Three-Dimensional Shape Similarity

Xinguo Liu Robin Sun* Sing Bing Kang Heung-Yeung Shum
Microsoft Research Asia Zhejiang University Microsoft Research Microsoft Research Asia

Abstract

In this paper, we propose a novel shape representation we call Directional Histogram Model (DHM). It captures the shape variation of an object and is invariant to scaling and rigid transforms. The DHM is computed by first extracting a directional distribution of thickness histogram signatures, which are translation invariant. We show how the extraction of the thickness histogram distribution can be accelerated using conventional graphics hardware. Orientation invariance is achieved by computing the spherical harmonic transform of this distribution. Extensive experiments show that the DHM is capable of high discrimination power and is robust to noise.

1 Introduction

Most of the early object recognition techniques are applied on 2-D objects. Progress in 3-D object acquisition hardware and techniques such as laser range finders and stereo has led us to the need to compare and recognize 3-D objects. Object comparison is the key technique in applications such as shape similarity based 3-D objects retrieval, matching, recognition and categorization [3, 8, 19]. This technology will be key as 3-D models become more mainstream among consumers.

In order to compare 3-D objects, an object representation is required. Typically, a 3-D object is represented by a geometry model, appearance attributes, and optionally annotations. Geometry models include boundary and voxel representations, CSG trees, point clouds, range images and implicit functions. Appearance attributes include color, texture and BRDFs, which are of particular interest to the graphics community. Object annotation includes other attributes that can describe an object at a semantic level.

Annotations provide an efficient and effective way to retrieve objects from a 3-D database. For example, a car model can be easily retrieved using the keyword “car”, if such an annotation is provided *a priori*. Unfortunately, automatic annotation is very difficult, as it requires some form of generic 3-D recognition that involves not just form but function as well. In addition, manual labeling is very laborious if the

database is huge.

Most approaches to compare 3-D objects rely on shape comparison. However, the geometry model used for object representation is usually developed for specific tasks such as modeling, editing and rendering, and not necessarily for comparison. Because there are many different types of geometry models, comparing them is difficult without some form of conversion. In addition, these models are usually not invariant to scaling or rigid transformations. As a result, shape descriptors are usually extracted from the geometry model to be used directly for comparison. Ideally, these descriptors should be scale and rigid transform invariant, capable of good discriminability, robust to noise, and independent of specific geometric representations. Not many current descriptors fit all these criteria.

In this paper, we propose a novel shape representation we call the *Directional Histogram Model* (DHM). It is scale and rigid transform invariant (which include invariance to mirroring and origin-symmetric transforms), and based on our extensive experiments, have good discriminability and robustness to noise. The DHM is derived by first computing a directional distribution of thickness histogram signatures, which are translation invariant. The computation of the thickness histogram signatures can be dramatically speeded up using graphics acceleration. Orientation invariance is achieved by computing the spherical harmonic transform of this distribution.

The rest of this paper is organized as follows. After summarizing related approaches in Section 2, we describe our Directional Histogram Model in Section 3. We also derive a matrix descriptor for the shape similarity comparison. Section 4 analyzes the invariance properties in more detail. In Section 5, we address the similarity problem between 3-D objects using their matrix descriptors. In Section 6, we address some sampling issues for the histogram generation, and present some object comparison and retrieval results in Section 7. We provide concluding remarks in Section 8.

2 Related Work

Previous work related to shape similarity can be found mainly in three research areas: (1) object recognition and classification, (2) surface matching and alignment, and (3) 3-D shape comparison and shape similarity based object re-

*This work was done while Robin Sun interned at Microsoft Research Asia.

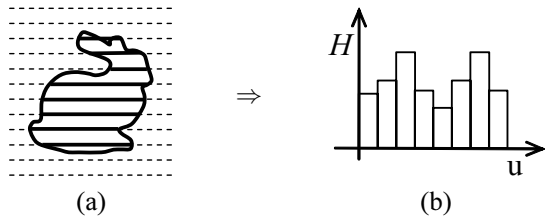


Figure 1: Constructing the Directional Histogram Model: (a) For each direction (θ, ϕ) , a group of parallel rays are used to sample the object “thickness” or extent, defined to be the distance between the nearest and farthest intersection points. (b) Histogram of thickness distribution.

trieval. The tasks of object recognition and classification are to determine whether a shape is a known object and find k representative objects in an object data set (see, for example, [4, 18]). Many object recognition approaches are view-based applied to 2-D objects.

The task of surface matching and alignment is to find overlapping regions between two 3-D objects, typically used for model reconstruction. Representative work include range-based approaches [6, 2], ICP-based approaches [5, 21], the spin-image [13, 14], geometric hashing [16], and structural indexing [23].

3-D shape comparison is related to surface matching, but focuses on comparing the object global shape, while surface matching compares only part of the object. By building a map from the 3-D shape onto a sphere, some approaches [15, 12, 22] generate some spherical representation for the shapes, and then compare them by the spherical representations. However, there are issues with occlusion, and these representations require orientation alignment.

View-based approaches have also been used for 3-D object recognition (e.g., [7]). However, such techniques tend to require large databases and memory footprints, and recognition rates tend to be slow.

There is a movement towards placing more emphasis on fast recognition rates, due to the potential of a 3-D search engine. This requires shape representations that not only fast to extract, but efficient to compare against other objects. Examples include multiresolutional Reeb graph (MRG) [11], shape distribution [20], shape histogram [1], ray-based descriptor [27, 26], groups of features [24], aspect graph [7], parameterized statistics [19], and 3-D FFT based descriptor [25].

The representation of MRG [11] provides a fully automatic similarity estimation of 3-D shapes by matching the topology. The topology information is analyzed based on the integrated geodesic distance, so the topology matching approach is pose invariant. However, the topology matching process is difficult to accelerate, which will be problem for large databases.

The shape distribution [20] is a very simple description for 3-D shape, which has advantages in 3-D object retrieval since it is easy to compute and efficient to compare. Osada *et al.* proposed the use of the $D2$ shape distribution [20], which is a histogram of distance between points on the shape surface. The directional histogram model presented in this paper generate histograms for all directions, and can be regarded as an extension to the $D2$ shape distribution.

Ankerst *et al.* [1] and Vranic *et al.* [27] proposed the use of feature vectors based on spherical harmonic analysis. While these are similar to our matrix descriptor, their spherical functions are sensitive to the shape centroid, which may change as a result of shape outliers or noise.

3 Directional Histogram Model

In this section, we introduce the Directional Histogram Model (DHM) for 3-D shapes. The DHM has the appropriate invariant properties and is expressive enough for 3-D shape similarity estimation.

To construct the DHM, we first choose a distribution of sampling directions. For each sampling direction, we compute a histogram of object extent or thickness using parallel rays. This procedure is illustrated in Figure 1. For each ray, the thickness is defined as the distance between the nearest and farthest points of intersection with the object surface. The DHM can be represented by a 3-D function $\mathcal{H}(\theta, \phi, \mu) : [0, \pi] \times [0, 2\pi] \times [0, 1] \mapsto R$, where θ, ϕ are the angular parameters for direction. For each (θ, ϕ) , the direction vector is $(\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)$, and

$$\mathcal{H}_{\theta, \phi}(\mu) \equiv \mathcal{H}(\theta, \phi, \mu)$$

is the thickness distribution of the object viewed from the direction (θ, ϕ) . Note that each thickness histogram is normalized with respect to the thickest value to ensure *scale invariance*. In our implementation, the sampling directions are computed as $\{ (\theta_i, \phi_j) \mid \theta_i = (i + 0.5)\frac{\pi}{N_s}, \phi_j = (j + 0.5)\frac{2\pi}{N_s}, 0 \leq i, j < N_s \}$, where $N_s \in Z^+$ is called the sampling rate.

The computation of the thickness histogram can be easily accelerated using commercially-available graphics hardware. For a given sampling direction, we render the front of the object, read its depth values, and repeat for the back of the object. The thickness is the difference between the front and back depth values.

For each sampling direction, the thickness distribution is represented using a histogram model. In our implementation, the thickness values μ is uniformly quantized into M integer values. $M = 64$ works well in our experiments. We

also normalize the the histogram such that

$$\int |\mathcal{H}_{\theta,\phi}(\mu)|^2 d\mu = 1, \quad \text{i.e.,} \quad \sum_{k=0}^{M-1} |\mathcal{H}_{\theta,\phi}(\frac{k}{M})|^2 = 1. \quad (1)$$

Properties 4 (Section 4) and 5 (Section 5) both rely on this normalization process.

Since two opposing sampling directions produce the same thickness values, the DHM is symmetric about the origin, i.e.,

$$\mathcal{H}_k(\theta, \phi) = \mathcal{H}_k(-\theta + \pi, \phi + \pi). \quad (2)$$

We refer to this property as being *origin-symmetric*.

Based on the model construction process, it is clear that the DHM is invariant to translation and scaling. However, at this point, it is orientation dependent.

To remove the dependence on orientation, we introduce a new additional component of the DHM we call the *matrix descriptor*. Thus, the DHM can now be regarded as M spherical functions $\mathcal{H}_k(\theta, \phi)$ defined on a unit sphere: $\mathcal{H}_k(\theta, \phi) \equiv \mathcal{H}(\theta, \phi, \frac{k}{M})$. Using spherical harmonic analysis, we have:

$$\mathcal{H}_k(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l h_{klm} Y_{lm}(\theta, \phi),$$

where $Y_{lm}(\theta, \phi)$ is the spherical harmonic and $h_{klm} = \langle \mathcal{H}_k(\theta, \phi), Y_{lm}(\theta, \phi) \rangle$. A useful property of $Y_{lm}(\theta, \phi)$ is $Y_{lm}(\theta + \alpha, \phi + \beta) = \sum_{m'=-l}^l D_{mm'}^l(\alpha) e^{im\beta} Y_{lm'}(\theta, \phi)$, where the coefficients satisfy: $\sum_{m'=-l}^l |D_{mm'}^l(\alpha)|^2 = 1$.

Since $\sum_{l=0}^L \sum_{m=-l}^l h_{klm} Y_{lm}(\theta, \phi)$ converges to $\mathcal{H}_k(\theta, \phi)$ as $L \rightarrow \infty$, we can assume that $\mathcal{H}_k(\theta, \phi)$ is band width limited for simplicity. Assuming that the bandwidth of $\mathcal{H}_k(\theta, \phi)$ is less than N , we have

$$\mathcal{H}_k(\theta, \phi) = \sum_{l=0}^N \sum_{m=-l}^l h_{klm} Y_{lm}(\theta, \phi). \quad (4)$$

Based on the spherical harmonic coefficients h_{klm} , we define the matrix descriptor \mathcal{M} as $\mathcal{M} = (a_{lk})_{M \times N}$, where

$$a_{lk} = \sqrt{\sum_{m=-l}^l |h_{klm}|^2}. \quad (5)$$

By the above definition, a_{lk} represents the energy sum of $\mathcal{H}_k(\theta, \phi)$ at the band of l . Therefore, the matrix descriptor gives the energy distribution of DHM over each (discrete) thickness value and each band index.

4 Invariance Properties

In this section, we analyze the matrix descriptor and show some of its invariance properties. *For the purposes of this*

analysis, the matrix descriptor is considered to be separate from the DHM, even though by our definition, the matrix descriptor is part of the DHM.

Property 1 *The matrix descriptor \mathcal{M} of a 3-D object is invariant to rotation, translation, and scaling.*

Since the DHM is invariant to translation and scaling, the matrix descriptor is invariant to translation and scaling.

Suppose the object is rotated by (α, β) . The DHM of the rotated object is then $\mathcal{H}'(\theta, \phi, \mu) = \mathcal{H}(\theta + \alpha, \phi + \beta, \mu)$. Therefore,

$$\begin{aligned} h'_{klm} &= \langle \mathcal{H}'_k(\theta + \alpha, \phi + \beta), Y_{lm}(\cdot) \rangle \\ &= \langle \sum_{s=0}^N \sum_{t=-s}^s h_{kst} Y_{st}(\theta + \alpha, \phi + \beta), Y_{lm}(\cdot) \rangle \\ &= \langle \sum_{s=0}^N \sum_{t=-s}^s h_{kst} \sum_{r=-s}^s D_{tr}^s(\alpha) e^{it\beta} Y_{sr}(\cdot), Y_{lm}(\cdot) \rangle \\ &= \sum_{s=0}^N \sum_{t=-s}^s \sum_{r=-s}^s h_{kst} D_{tr}^s(\alpha) e^{it\beta} \langle Y_{sr}(\cdot), Y_{lm}(\cdot) \rangle \\ &= \sum_{t=-l}^l h_{klt} D_{tm}^l(\alpha) e^{it\beta}, \end{aligned}$$

where “ \cdot ” denotes θ, ϕ . Using the orthogonality property of $e^{it\beta}$, we have

$$\begin{aligned} |h'_{klm}|^2 &= \sum_{t=-l}^l |h_{klt}|^2 |D_{tm}^l(\alpha)|^2 \\ \sum_{m=-l}^l |h'_{klm}|^2 &= \sum_{m=-l}^l \sum_{t=-l}^l |h_{klt}|^2 |D_{tm}^l(\alpha)|^2 \\ &= \sum_{t=-l}^l |h_{klt}|^2 \sum_{m=-l}^l |D_{tm}^l(\alpha)|^2 \\ &= \sum_{t=-l}^l |h_{klt}|^2 = \sum_{m=-l}^l |h_{klm}|^2. \end{aligned}$$

Since $a'_{lk} = a_{lk}$ and $\mathcal{M} = \mathcal{M}'$, the matrix descriptor is thus rotation invariant.

Let \mathcal{M}_i be the matrix descriptor derived from a DHM \mathcal{H}_i , $i = 0, 1$. If $\mathcal{M}_0 = \mathcal{M}_1$, then \mathcal{H}_0 and \mathcal{H}_1 are equivalent DHMs, denoted by

$$\mathcal{H}_0 \sim \mathcal{H}_1.$$

Property 2 *The matrix descriptor \mathcal{M} of a 3-D object is invariant to origin-symmetric transform and mirror transform.*

Since the DHM of a 3-D object is origin-symmetric, $\mathcal{H}(\theta, \phi, \mu) = \mathcal{H}(-\theta + \pi, \phi + \pi, \mu)$. Then

$$\mathcal{H}(\theta, \phi, \mu) \sim \mathcal{H}(-\theta + \pi, \phi + \pi, \mu),$$

i.e., the matrix descriptor \mathcal{M} is invariant to origin-symmetric transform.

To show the invariance to mirror transform, we can assume the mirror is the $X - Y$ plane without loss of generality according to Property 1. Let $\mathcal{H}'(\theta, \phi, \mu)$ be the DHM of the mirrored object. Then

$$\begin{aligned}\mathcal{H}'(\theta, \phi, \mu) &= \mathcal{H}(-\theta + \pi, \phi, \mu) \\ &\sim \mathcal{H}(-\theta + \pi, \phi + \pi, \mu) = \mathcal{H}(\theta, \phi, \mu).\end{aligned}$$

Therefore, the matrix descriptor is invariant to mirroring.

Property 3 Let $\mathcal{M} = (a_{lk})_{M \times N}$ be a matrix descriptor. Then

$$a_{lk} = 0, \text{ if } l \text{ is odd.} \quad (6)$$

If l is odd, then $Y_{lm}(-\theta + \pi, \phi + \pi) = -Y_{lm}(\theta, \phi)$. Since $\mathcal{H}_k(\theta, \phi) = \mathcal{H}_k(-\theta + \pi, \phi + \pi)$, we have

$$\begin{aligned}h_{klm} &= \langle \mathcal{H}_k(\theta, \phi), Y_{lm}(\theta, \phi) \rangle \\ &= \langle \mathcal{H}_k(-\theta + \pi, \phi + \pi), -Y_{lm}(-\theta + \pi, \phi + \pi) \rangle \\ &= -h_{klm}.\end{aligned}$$

Therefore $h_{klm} = 0$, and $a_{lk} = \sqrt{\sum_{m=-l}^l |h_{klm}|^2} = 0$.

Property 4 The squared sum of the matrix descriptor elements is 1.

$$\begin{aligned}\sum_{l=0}^{N-1} \sum_{k=0}^{M-1} a_{lk}^2 &= \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \sum_{m=-l}^l ||h_{klm}||^2 \\ &= \sum_{k=0}^{M-1} \oint_s ||\mathcal{H}_k(\theta, \phi)||^2 ds = \oint_s \sum_{k=0}^{M-1} ||\mathcal{H}_k(\theta, \phi)||^2 ds \\ &= \oint_s \sum_{k=0}^{M-1} ||\mathcal{H}_{\theta, \phi}(\frac{k}{M})||^2 ds = \oint_s ds = 1,\end{aligned}$$

where s denote the unit sphere, and $\oint_s ds = 1$ is assumed in the spherical harmonic analysis.

5 Shape Similarity

Let O_1, O_2 be two 3-D objects. The similarity between O_1 and O_2 can be measured using the norm of their matrix descriptors' difference $\mathcal{M}(O_1) - \mathcal{M}(O_2)$:

$$d(O_1, O_2) = ||\mathcal{M}(O_1) - \mathcal{M}(O_2)||.$$

The matrix norm we take in this paper is the L^p norm, with $p = 2$. Let v_l be the l -th row vector in a shape matrix. Note that the v_l represents the energy of the directional histogram model at l -th frequency, it can be weighted when calculating

the object distance. When using the L^p norm, the weighted form of distance function can be represented as

$$d_p(O_1, O_2) = \left(\sum_{l=0}^M \omega_l ||v_{1l} - v_{2l}||^p \right)^{1/p}, \quad (7)$$

where v_{jl} is the l -th row vector in the shape matrix of object O_j , $\omega_l > 0$ are the weights, and $|| \cdot ||$ is the L^p norm of a vector. By adjusting the weights, we can emphasize the importance of objects at some frequency for special purpose when evaluating the shape similarity. In our experiments, we chose $p = 2$ and all weights $\omega_l = 1$. With this choice, the following property on the d_2 distance function holds.

Property 5 The d_2 distance between any two objects is between 0 and $\sqrt{2}$.

Since the elements in the matrix descriptor are all positive, we have $d_2^2(O_1, O_2) = \sum_{l,k} (a_{1lk} - a_{2lk})^2 < \sum_{l,k} (a_{1lk}^2 + a_{2lk}^2)$. Then $d_2^2(O_1, O_2) < 2$, according to Property 4. Therefore $d_2(O_1, O_2) < \sqrt{2}$. In Table 2, the maximum distance value is about 1.

2-D Shape Similarity The directional histogram model and matrix descriptor can be adapted for 2-D objects. For a 2-D object, the DHM is a 2-D function $\mathcal{H}(\phi, \mu) : [0, 2\pi] \times [0, 1] \mapsto R$, with ϕ being the orientation angle. For each ϕ , $\mathcal{H}_\phi(\mu) \equiv \mathcal{H}(\phi, \mu)$ gives the thickness distribution of the object viewed from the direction ϕ . By applying Fourier transform to the 2-D directional histogram model, we have: $\mathcal{H}_k(\phi) = \mathcal{H}(\phi, \frac{k}{M}) = \sum_{l=0}^{\infty} h_{kl} e^{-i\phi}$. The matrix descriptor for the 2-D object is $\mathcal{M} = (a_{lk})_{M \times N}$, $a_{lk} = \sqrt{|h_{kl}|^2}$. It can be shown in a similar way that the 2-D object's matrix descriptor is also invariant to translation, scaling, rotation, origin-symmetric and mirroring. As before, $a_{lk} = 0$ if l is odd. The squared sum of all the matrix descriptor elements is 1. Based on the 2-D matrix descriptor, the distance function d_p is also well-defined for 2-D object, and ranges from 0 to $\sqrt{2}$.

6 Sampling Issues

In this section, we address the sampling issues for the DHM.

Computational Cost: First, we examine the computational cost to generate the DHM. We use a squared window to render the object in the hardware accelerated approach, and let N_w be the window size (i.e., window width). Recall that N_s is the number of angles θ and ϕ (i.e., sampling rate). For each direction θ_i, ϕ_j , the object is rendered twice, and the depth buffer is read twice as well.

Fortunately, only half of all the sampling directions is needed, since opposing directions produce the same thickness histograms as shown by Property 2. As a result, the bulk of time cost is $T = N_s^2(T_b(N_w) + T_r(N_w))$, where

$T_b(N_w)$ is the time cost to read the depth buffer values from a $N_w \times N_w$ window and $T_r(N_w)$ is the render time in the same window. Usually, $T_r(N_w)$ is roughly proportional to the object's face number N_f , i.e., $T_r(N_w) \approx \lambda_{N_w} N_f$, where λ_{N_w} is a constant. Therefore, $T \approx N_s^2 (T_b(N_w) + \lambda_{N_w} N_f)$. This is verified by the performance results shown in Figure 2.

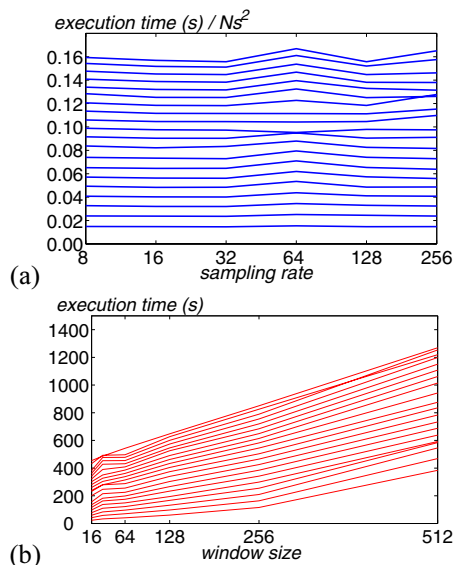


Figure 2: Performance of DHM generation. Many objects are used for testing, each of which corresponds to a curve. For curves from bottom to top, the vertex number of the corresponding object ranges uniformly from 5,000 to 100,000. (a) Execution time vs. sample rate at window size $N_w = 128$. (b) Execution time vs. window size at sampling rate $N_s = 64$. The curves in (a) and (b) show that the execution time is approximately proportional to the squared sampling rate and the window size. This performance is tested on PIII 667 MHz PC with 512 MB memory and a 3D commercial graphics card, and the time is measured in seconds.

Window Size: For a given sampling direction, the number of intersecting rays used to sample the object thickness distribution is proportional to the squared window size N_w^2 . It is reasonable to expect that a higher ray density should produce more accurate thickness distribution. If the ray density is too coarse, the resulting thickness distribution and the final

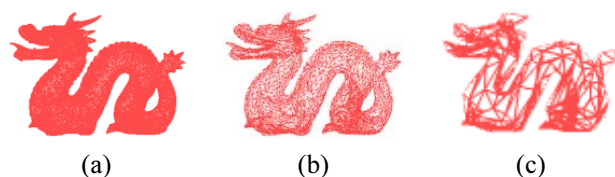


Figure 3: Three (out of nine) dragons used in Figure 4.

matrix descriptor will be strongly dependent on the location of the object as it is being sampled.

We examined the relationship between the error of the matrix descriptor introduced by perturbing the object and the window size N_w through extensive experimentation. In our experiments, a 3-D dragon model is first simplified to generate objects of varying complexity. Figure 3 shows three (out of nine) simplified models. We measure the complexity of 3-D object by its average edge length (normalized with respect to the diameter of the object bounding sphere). For each window size $N_w = 32, 64, 128, 256$ and for each dragon object, the matrix descriptor is computed 50 times, each time with the model randomly perturbed. These matrix descriptors were then compared with the reference matrix descriptor computed without disturbing the model, and perturbation error is analyzed with the usual statistics of mean and standard deviation. (The matrix is compared using the L^2 norm throughout this paper.)

The experiment results are summarized in Figure 4. Figure 4(a) shows that while the perturbation error is quite constant over different object complexity, Figure 4(b) shows an interesting trend of the perturbation error being roughly proportional to the ray interval (i.e., inversely proportional to the window size). While this shows that a larger window size is better, it would decrease the rendering speed and depth buffer extraction. We found that $N_w = 128$ is a good trade-off.

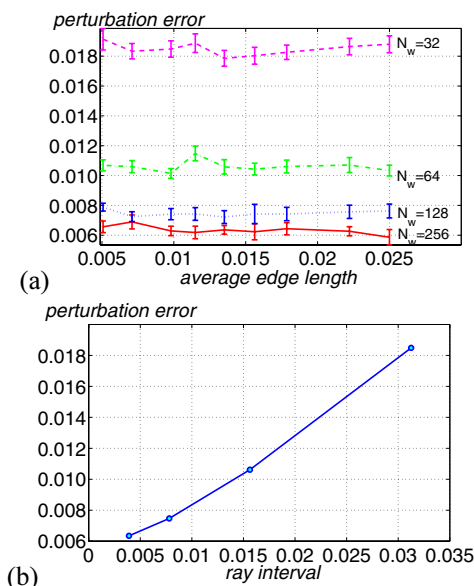


Figure 4: Relationship between the matrix descriptor and window size. (a) Perturbation error (with standard deviation bars) vs. average edge length for different window sizes (N_w), (b) Perturbation error vs. ray interval ($1/N_w$).

Directional Sampling Rate: For a typical object with 20K



Figure 5: Five of the objects used in our experiments.

vertices in our database and a 128×128 window, we can render the object and read the depth buffer at about 30 fps. Therefore, the total time cost is about $\frac{1}{30} N_s^2$ seconds. In addition, we impose a strict limitation on the sampling rate N_t for efficiency. Recall that the bandwidth of $\mathcal{H}_k(\theta, \phi)$ is assumed as N in Section 3, then at least $2N \times 2N$ samples are needed for $\mathcal{H}_k(\theta, \phi)$ according to spherical harmonic analysis [10]. However, in practice, the bandwidth of $\mathcal{H}_k(\theta, \phi)$ is not necessarily limited. Using a finite number of samples would then result in loss of information (power).

Because the power distribution of $\mathcal{H}_k(\theta, \phi)$ depends on the object shape, the number of samples needed for arbitrary objects remains unknown. We analyze this problem by performing a lot of experiments. First, we calculate the matrix descriptor M_{N_s} for many objects (five of which are shown in Figure 5) at different sampling rates N_s . For each object, the quality of the approximation due to the sampling rate is calculated by comparing them against M_{256} (here M_{256} is used as the ground truth M_∞ , which is not possible to obtain in practice).

The results of this experiment (Figure 6(a)) show that the approximation error drops very quickly as the sampling rate is increased. A sampling rate at least $N_s = 128$ is needed for a reasonable approximation. However, it will take about 8 minutes to generate the DHM for a typical object in our database when $N_s = 128$. While this is not unreasonable, it is impractical for time-critical applications such as for 3-D search engines for web applications. As a result, we have to compromise by reducing the sampling rate at the expense of fidelity of representation.

In another set of experiments, we calculate distances between pairs of different objects under different sampling rates ($N_s = 8, 16, 32, 64, 128, 256$). Basically, the larger the distance, the better the discrimination power. Results indicate that most of the distances increase monotonically as N_s is increased. To enable comparisons between different objects, we normalized the distance with respect to the distance at $N_s = 256$ for each object pair. The graph of the average (with standard deviation bars) is shown in Figure 6(b). This graph clearly shows that the sampling rate $N_s = 16$ is sufficient for shape similarity estimation, and increasing the sampling rate to more than 16 would produce only marginal improvements in accuracy at the expense of speed. Note that it takes about only 8 seconds when $N_s = 16$.

While the matrix descriptor of an object is theoretically

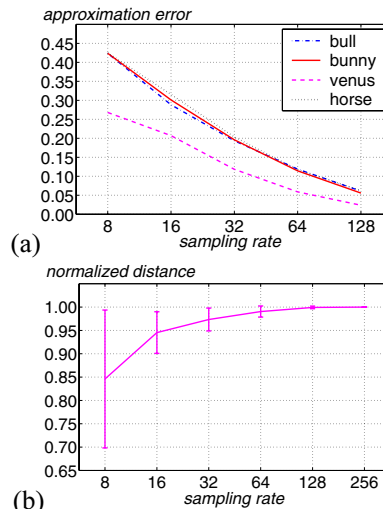


Figure 6: Results of experiments on sampling rate N_s . (a) Approximation errors vs. sampling rate, (b) Normalized distance vs. sampling rate.

invariant to rigid transform and scaling, it will change to some degree when the object undergoes these transformations because of the finite N_s^2 directional sampling. In our experiments, we found that the variances introduced by scaling and translation are very small. The variances introduced by rotation are somewhat larger, but they are not significant with a sampling rate $N_s \geq 16$, as shown in Table 1.

$N_s = 16$	10°	20°	30°	40°	50°
ball	0.001	0.002	0.004	0.003	0.003
venus	0.002	0.015	0.013	0.013	0.013
bunny	0.002	0.003	0.004	0.005	0.007
horse	0.031	0.043	0.043	0.039	0.035
bull	0.007	0.006	0.012	0.019	0.018
$N_s = 64$	10°	20°	30°	40°	50°
ball	0.001	0.001	0.002	0.001	0.001
venus	0.004	0.005	0.005	0.004	0.003
bunny	0.011	0.015	0.014	0.007	0.007
horse	0.016	0.021	0.021	0.013	0.013
bull	0.011	0.013	0.014	0.009	0.009

Table 1: Changes of matrix descriptor due to rotation. The matrix descriptor of the rotated object is compared against with that of the original object. Results under two sampling rate are listed.

Object Simplification: Computational cost analysis shows that the time cost is proportional the object complexity in terms of face/vertex number. State-of-art simplification algorithms (e.g., [9, 17]) are capable of simplifying large objects quickly (typically, in only a few seconds). Using an

efficient model simplification algorithm before sampling the thickness distributions would clearly be advantageous.

Generally, model simplification may introduce some error into the matrix descriptors. We call this error the *simplification error*. How much of a simplification error depends on the simplification level. To study this effect, we ran many trials involving many objects, and summarized the results in Figure 7. The simplified object is characterized by its normalized average edge length with respect to the object bounding sphere’s diameter. The results show that within the range of simplification we used, only small simplification errors (< 0.06) are obtained. Note that the most simplified versions of the “dinosaur”, “wolf”, “man” and “bunny” models consist of only 1300, 1500, 982, 1556 vertices respectively. In Figure 7, the window size for rendering is $N_w = 128$. It is curious to note that the curves increase more dramatically after average vertex distance of 0.0125 (shown by a vertical line in Figure 7), which corresponds to about 1.5 times the ray interval.

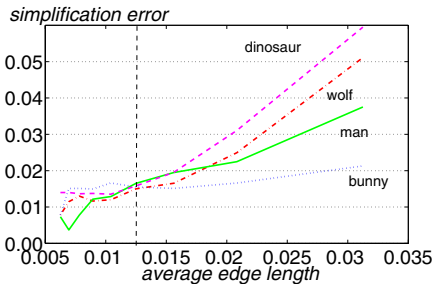


Figure 7: Error in the matrix descriptor introduced by simplification. The simplification is characterized by the normalized average edge length of the simplified model.

7 Object Comparison and Retrieval

For shape similarity comparison, we found that $N_s = 16$, $N_w = 128$ is a set of good sampling parameters in terms of accuracy and efficiency, based on our experimental results in Section 6. Using these sampling parameters, we obtain typical shape comparison results shown Table 2. In Figure 8, we compare the shape similarity between interpolated objects. It is interesting to note that as we morph the object to another, the distance actually increases monotonically as we expected.

Based on the shape similarity measurements in Section 5, we build a simple example shape based 3-D object retrieval prototype system. In this prototype, a sample shape is specified by the user, then the most k similar objects are retrieved from a small 3-D database in a local machine. There are a total of 500 different 3-D objects in our 3-D database.

In our prototype system, the matrix descriptors of all the objects in the database are pre-computed and stored. The

	0.000	0.353	0.780	1.029	1.049
	0.353	0.000	0.525	0.831	0.846
	0.780	0.525	0.000	0.434	0.449
	1.0289	0.8311	0.4341	0.0000	0.352
	1.049	0.846	0.449	0.352	0.000

Table 2: Object comparison. For each object, the nearest distance value, except that to itself, is shown in bold.

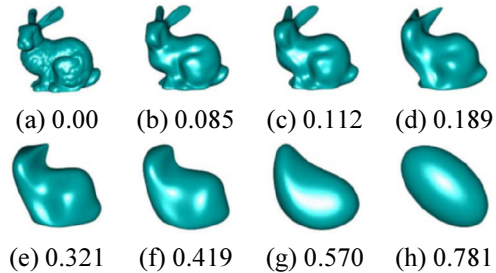


Figure 8: Shape comparison between interpolated objects. (a) The original object. (b)-(g) interpolated objects between (a) and (h). The number under each object is the distance to the original object (a).

sampling rate for the directions is $N_s = 16$. The average time spent to compute one matrix descriptor is about 8 seconds for each object. (The very dense models are simplified first.) Some input and output examples of our prototype system are shown in Figure 9. To further reduce the matrix variance due to rotation at sampling rate $N_s = 16$ as shown in Table 1, a simple pre-orientation procedure was applied. We used Principle Component Analysis (PCA) to compute the major axis for automatic alignment.

8 Concluding Remarks

We have presented the Directional Histogram Model (DHM) to study the shape similarity problem of 3-D objects. This novel representation is based on the thickness variations with different viewing directions. In each viewing direction, a histogram model for the thickness distribution is built. We also presented a new shape descriptor in matrix form which we call matrix descriptor.

A major advantage of the matrix descriptor is its invariance to the transforms of scaling, translation, rotation, origin-symmetric and mirroring. Just as important, the matrix descriptor is easy and fast to compute using commercially-available graphics hardware, stable against noise, and expressive enough to distinguish between different global shapes. The invariance properties of the matrix descriptor makes it is highly suitable for shape similarity


















































i	o_1	o_2	o_3	o_4	o_5	o_6
						
						
						
						
						
						
						

Figure 9: Examples of 3-D object retrieval. In each row, the column “ i ” is the input sample object, the column “ o_j ” is the j -th most similar object retrieved.

measurement, and it produces reasonable shape comparison results as experiments have shown.

There are two major limitations of the DHM: (1) it can only describe rigid shapes, and (2) it only captures the global shape. As a result of (1), non-rigid transformed versions of the object are deemed different. As a result of (2), local features cannot be represented and compared.

Future work include investigating means of extending the current formulation to handle the DHM’s inability to handle non-rigid shapes and local features. In addition, it would be interesting to analytically predict how many directional samples would be sufficient for any given 3-D object. We are currently in the process of making our prototype system into a full-fledged, practical 3-D search engine. All the user-related issues would need to be addressed.

References

[1] M. Ankerst, G. Kastenmüller, H.-P. Kriegel, and T. Seidl. 3D shape histograms for similarity search and classification in spatial databases. In *Advances in Spatial Databases, 6th Intl. Symp., SSD '99*, volume 1651, pages 207–228, 1999.

[2] F. Arman and J.K. Aggarwal. Model-based object recognition in dense-range images - a review. *ACM Computing Surveys*, 25(1):5–43, 1993.

[3] S. Berchtold and H.-P. Kriegel. S3: Similarity search in CAD database systems. In *ACM SIGMOD*, pages 564–567, 1997.

[4] P.J. Besl and R.C. Jain. Three-dimensional object recognition. *ACM Computing Surveys*, 17(1):75–145, 1985.

[5] P.J. Besl and H.D. McKay. A method for registration of 3-D shapes. *PAMI*, 14(2):239–256, 1992.

[6] G. Blais and M.D. Levine. Registering multiview range data to create 3D computer objects. *PAMI*, 17(8):820–824, 1995.

[7] C.M. Cyr and B.B. Kimia. 3D object recognition using shape

similarity-based aspect graph. In *ICCV*, pages I: 254–261, 2001.

[8] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs. A search engine for 3D models. *to appear in ACM TOG*, 2003.

[9] M. Garland and P.S. Heckbert. Surface simplification using quadric error metrics. In *SIGGRAPH*, pages 209–216, 1997.

[10] D.M. Healy, D.N. Rockmore, and S.S.B. Moore. FFTs for the 2-sphere - improvements and variations. Technical Report PCS-TR96-292, 1996.

[11] M. Hilaga, Y. Shinagawa, T. Kohmura, and T.L. Kunii. Topology matching for fully automatic similarity estimation of 3D shapes. In *SIGGRAPH*, pages 203–212, 2001.

[12] K. Ikeuchi and M. Hebert. Spherical representations: From EGI to SAI. Technical Report CMU-CS-95-197, Computer Science Department, CMU, 1995.

[13] A. Johnson and M. Hebert. Surface registration by matching oriented points. In *Intl. Conf. on Recent Advances in 3-D Digital Imaging and Modeling*, pages 121–128, 1997.

[14] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *PAMI*, 21(5):433–449, 1999.

[15] S.B. Kang and K. Ikeuchi. The complex EGI: A new representation for 3-D pose determination. *PAMI*, 15(7):707–721, 1993.

[16] Y. Lamdan and H.J. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. In *ICCV*, pages 238–249, 1988.

[17] P. Lindstrom and G. Turk. Fast and memory efficient polygonal simplification. In *IEEE Visualization*, pages 279–286, 1998.

[18] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, 1998.

[19] R. Ohbuchi, T. Otagiri, M. Ibato, and T. Takei. Shape-similarity search of three-dimensional models using parameterized statistics. In *Pacific Graphics*, pages 265–274, 2002.

[20] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Matching 3D models with shape distributions. In *Shape Modeling Intl.*, pages 154–166, 2001.

[21] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3D model acquisition. In *SIGGRAPH*, pages 438–446, 2002.

[22] H.-Y. Shum, M. Hebert, and K. Ikeuchi. On 3D shape similarity. In *CVPR*, pages 526–531, 1996.

[23] F. Stein and G. Medioni. Structural indexing: Efficient 3-D object recognition. *PAMI*, 14(2):125–145, 1992.

[24] M.T. Suzuki. A web-based retrieval system for 3D polygonal models. In *Joint 9th IFSA World Congress and 20th NAFIPS Intl. Conf.*, pages 2271–2276, 2001.

[25] D.V. Vranic and D. Saupe. 3D shape descriptor based on 3D fourier transform. In *EURASIP Conf. on Digital Signal Processing for Multimedia Communications and Services*, pages 271–274, 2001.

[26] D.V. Vranic and D. Saupe. Description of 3d-shape using a complex function on the sphere. In *IEEE Intl. Conf. on Multimedia and Expo*, pages 177–180, 2002.

[27] D.V. Vranic, D. Saupe, and J. Richter. Tools for 3D-object retrieval: Karhunen-loeve transform and spherical harmonics. In *IEEE 2001 Workshop Multimedia Signal Processing*, pages 293–298, 2001.