

Multiresolution Wavelet Analysis of Shape Orientation for 3D Shape Retrieval

Zhenbao Liu

liuzhenbao@npal.cs.tsukuba.ac.jp

Jun Mitani

Department of Computer Science, University of Tsukuba
Tennodai, Tsukuba, Ibaraki, 305-8573, Japan
mitani@cs.tsukuba.ac.jp

Yukio Fukui

fukui@cs.tsukuba.ac.jp

Seiichi Nishihara

nishihara@cs.tsukuba.ac.jp

ABSTRACT

In the present paper, we propose a novel 3D shape descriptor by performing multiresolution wavelet analysis on shape orientation. We consider the spatial orientation of the polygon surfaces of a shape as important information and characterize this information by setting view planes. We then analyze these view planes by multiresolution wavelet analysis, a powerful tool used in signal processing, and lower the high resolution to low frequency domains because the high resolution contains too much information, which must be reduced in order to capture the main components. We compare the proposed descriptor to two of the best-performing descriptors on the Princeton Shape Benchmark, Spherical Harmonics Descriptor and Light Field Descriptor, and analyze the performance of the proposed descriptor from several aspects. We also compare the proposed descriptor to the Spherical Wavelet Descriptor, which won the best paper award at SMI06, a near method to our descriptor. The proposed descriptor improves the retrieval performance.

Categories and Subject Descriptors

I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling – *Curve, surface, solid, and object representations*;
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *Retrieval models*; H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing – *Abstracting methods*.

General Terms: Algorithms, Performance, Experimentation, Measurement

Keywords: 3D shape retrieval, 3D shape descriptor, shape orientation, multiresolution wavelet analysis.

1. INTRODUCTION

In the past twenty years, the use of multimedia information has expanded quickly in a number of fields such as image, audio, video, 3D model. The availability of an enormous number of

multimedia data has forced researchers to consider how to supply users with good retrieval methods. The 3D model, as a relatively new form of multimedia, has higher dimension than other multimedia data, and so it is more difficult to find very efficient retrieval methods. Therefore the computer graphics community has shown considerable interest in 3D shape retrieval.

3D shape retrieval can be described as follows. A shape retrieval algorithm searches the similar shapes to a given query shape, which belong to the same class in a database. It is difficult to infer mathematically which algorithm of shape retrieval is better than others. Therefore, the only way of comparison is to adopt a shape benchmark data set to evaluate the performance of the algorithms on the same shape benchmark. The best known benchmark is the Princeton Shape Benchmark [1] which is widely adopted by researchers. This benchmark contains generic 3D shapes for testing.

We think the performance of a 3D shape descriptor should be judged on discriminative power and time-space cost requirements. The discriminative ability of an algorithm is usually judged in two ways. The first is to draw the recall precision curve according to a similarity measure on the database, and compare the retrieval ability by observing the relative up or down positions of the curves. This means has a shortage that if two curves of recall precision intersect at the approximately medium position of recall value, it is difficult to judge which method is relatively superior. The second is to use quantitative statistics on five common recommended tools: Nearest Neighbor, First Tier, Second Tier, E-Measure, and Discounted Cumulative Gain (DCG), to evaluate the retrieval results. It is easy to analyze retrieval precision by these quantitative tools and then the five quantitative parameters are often used to compare algorithms on the Princeton Shape Benchmark. The time-space costs represent time complexity and storage complexity and can not be ignored. We think the computation time is more important than storage cost because recent storage devices have sufficient memory, and users can not wait for the retrieval results in a long time. Thus the computation time should not be given little regard.

In the present paper, we consider that the spatial orientation of the polygon surfaces on a shape contain important information. We can imagine that if one person observes the shape orientation from six directions, front and back, left and right, and up and down, the features of this shape can be ascertained. We characterize this information by a computation algorithm. We first normalize the shape to obtain the invariance to affine transform, and then compute the bounding cube of the shape. Each plane of the cube representing one view plane, is partitioned into many view points

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MIR'08, October 30–31, 2008, Vancouver, British Columbia, Canada.
Copyright 2008 ACM 978-1-60558-312-9/08/10...\$5.00.

horizontally and vertically with a high resolution. Each view point on the view plane stores the orientation of the nearest face to this point. Here we only care about the nearest face and the orientation of this face relative to the eye of the view point. We analyze the six planes by multiresolution wavelet analysis, and lowered the high resolution to low frequency domains because the high resolution contains too much information, which must be reduced in order to capture the main components. We organized experiments on the Princeton Shape Benchmark of generic 3D shapes, and found that this approach leads to an improvement in retrieval performance for several common descriptors, such as the Spherical Harmonics Descriptor [2] and the Spherical Wavelet Descriptor [3], and also possesses fast computation time and a compact representation relative to the Light Field Descriptor [4]. The proposed descriptor can achieve better balance among the retrieval precision, time complexity, and space complexity.

2. PREVIOUS WORK

Recently, a number of 3D shape retrieval methods have been proposed. The reader can refer to [5] for a survey of methods before 2004, and to [6] for the latest methods till 2008. In the present paper, we discuss only some of the methods relevant to the present study.

Several methods have been used to characterize the intrinsic attributes, such as the distances to the center [7] [2] [8], and the curvature [9], of 3D shapes, and to project them onto a sphere to form spherical functions. Since the spherical function has 2.5 dimensions, processing is easier than that in 3D space. The spherical harmonics are first introduced in the 3D model retrieval by Vranic et al. in [7]. Kazhdan et al. [2] applied the invariance properties of spherical harmonics and presented an affine invariant descriptor based on spherical harmonics. Vranic et al. [8] improved this method by combining it with [7]. Novotni and Klein [10] presented a 3D Zernike descriptor by computing 3D Zernike descriptors from voxelized models as natural extensions of spherical harmonics based descriptors. However, for these methods, which are dependent on spherical functions, the small change in position of the sphere center can result in a significant noise in the feature descriptor.

Statistics on the global geometric property of a 3D shape has been applied to shape matching. Ankerst et al. [11] used shape histograms decomposing shells and sectors around a model's centroid. Osada et al. [12] matched 3D models with shape distributions on the Euclidean distance histograms of two arbitrary points of the surface. Liu et al. [13] presented a novel 3D shape descriptor for effective shape matching and analysis that utilized both local and global shape signatures, and term their descriptor "generalized shape distributions" because it is an extension of shape distributions [12]. These methods have a common limitation that they are capable of only capturing similar gross shape properties, and are powerless to capture the detailed shape properties.

After researchers found that histograms of Euclidean distances could not be used to pose-changing shapes such as bending or stretching, geodesic distances over the surface of the shape gained the attention. Tung and Schmitt [14] used the geodesic distances to construct an augmented multiresolution Reeb graph for 3D shape retrieval. Jain and Zhang [15] computed the spectral embeddings given by eigenvectors and eigenvalues of a geodesic distances matrix. These descriptors have an advantage in that they

are invariant to non-rigid transformations. However, the computation of geodesic distances brings a big burden because the time cost is very high. The more recent work is that Ben-Chen and Gotsman [16] proposed a descriptor for characterizing shape using conformal factors. This descriptor is also invariant to pose changes, and is easy to compute. However this descriptor is subject to the constraint of the manifold mesh. Experiments of these descriptors were conducted on the McGill database [17], which was designed to test methods invariant to pose changes.

Light Field Descriptor [4], which is a representative method for reducing a 3D shape to a 2D space, produced projections of a 3D shapes from many viewing angles, and then encoded these projections as features by Zernike moments and Fourier descriptors. The LFD represents a visual perception similar to that of humans, is considered to be the best-performing descriptor on the Princeton Shape Benchmark. However this descriptor must produce approximately one hundred projections, and has significant time cost. Therefore, the LFD is not applicable to the real-time retrieval.

Next, we introduce the research related to the present study. Relative to these descriptors characterizing the intrinsic attributes, such as the distances to the center [7] [2] [8] [3], the curvature [9], Euclidean distances [11] [12] [13] or geodesic distances [14] [15] between vertices, and conformal geometry [16] in addition to Gaussian curvature, in this paper we will introduce one new intrinsic attribute, shape orientation. And we utilize a tool, wavelet, to analyze the information carried by shape orientation. Since wavelet analysis was first performed to image and graphics slightly more than 10 years ago, the wavelet has widely been applied to image processing, computer vision, computer graphics and other areas. Wavelet based algorithms define the state-of-the-art for applications including coding, restoration, and segmentation. Wavelet was first introduced to 3D shape retrieval by Laga et al. [3]. They used spherical wavelets to analyze the spherical functions defined by the sampling of the distances between surface and the center of mass of an object. Since the spherical function has a shortage that it is sensitive to the choice of the spherical center, and from a mathematical viewpoint, spherical wavelet transform has not yet been well defined up to now, this descriptor has not achieved satisfactory results on Princeton Shape Benchmark.

In the present paper, we will compare the proposed descriptor to two of the best-performing descriptors on the Princeton Shape Benchmark, namely, the Spherical Harmonics Descriptor, the Light Field Descriptor, and will analyze the performance of the proposed descriptor from several aspects. In addition, we will compare our descriptor to the Spherical Wavelet Descriptor, which received the best paper award at SMI06, a near method to our descriptor.

3. SHAPE DESCRIPTOR

We normalize the 3D shape into a canonical coordinate frame and characterize the shape orientation by setting view planes, and then perform one important analysis, multiresolution wavelet analysis on view planes. The final wavelet coefficients of low scales are used to be the feature vector.

3.1 Normalization

In the first step, shapes are aligned into a canonical coordinate frame by Principal Component Analysis (PCA) [18] [19] [20] to

determine the invariant measure with respect to translation, rotation, reflection and scaling of the original shape. The covariance matrix C is approximated as follows

$$C = \frac{1}{n} \sum_{i=1}^n S_i (g_i - m)(g_i - m)^T \quad (1)$$

where S_i and g_i are the area of a triangle of a shape and the center of gravity, respectively, m is the center of mass of a shape, and n is the number of triangles of the shape. The first eigenvector of this covariance matrix corresponding to the largest eigenvalue points to the direction of the largest variance along which the rotation is applied.

3.2 Sampling shape orientation from view planes

We place six view planes to the six faces of the bounding cube, and then decompose each view plane into several view points by a single resolution $N \times N$ horizontally and vertically. The decomposition is uniform in the horizontal and vertical directions.

The orientation of a face on the surface of a 3D shape can be described with the normal vector from inside to outside. We sample the face orientation by casting a perpendicular ray representing the view direction from a view point. The sampled face is the first one which the ray hits.

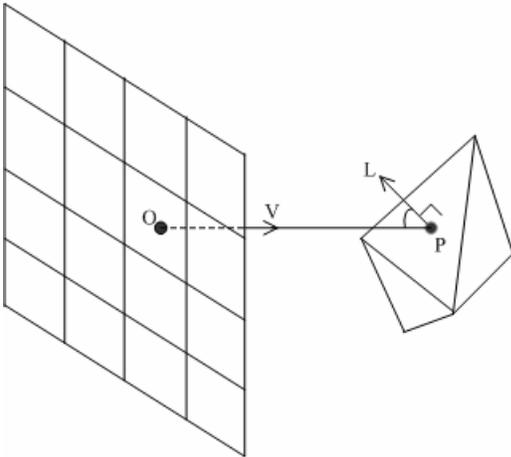


Figure 1. Sampling the orientation of one face

Figure 1 illustrates the sampling process. The point O represents a view point, which is the center of each small decomposed view plane between the horizontal grid x and the vertical grid y on a view plane. A ray V representing the view direction casts from the view point O and is perpendicular to the view plane. It hits the first face with the intersection P . The vector L and V illustrate the orientation of the face normal and the view direction respectively. The value of inner product $(L, -V)$ is assigned to the view point O as the sampling value $o_{x,y} = (L, -V)$. The reason that a minus sign exists before the vector V is that we sample only the acute angle between the directions L and V . Therefore, the sampling value is in the range of $[0, 1]$. The orientation matrix O from a view plane is as follows.

$$O = \begin{bmatrix} o_{0,0} & \cdots & o_{0,N-1} \\ \vdots & o_{x,y} & \vdots \\ o_{N-1,0} & \cdots & o_{N-1,N-1} \end{bmatrix} \quad (2)$$

We use one function form to denote the matrix, and this function is called the orientation function, and is given as follows.

$$O(x,y) = o_{x,y}, x,y \in [0, N-1] \quad (3)$$

Here, we adopt the resolution $N=64$ to depict this function as in the following Figure 2 which shows one human shape, and six orientation sampling pictures from front and back, left and right, and up and bottom, corresponding to six view planes. In each of orientation sampling pictures, the orientation value is in $[0, 1]$, and the gray value of the corresponding pixel belongs to $[0, 255]$. That is, the degree of whiteness represents the orientation value, the whiter the pixel, the larger the value.

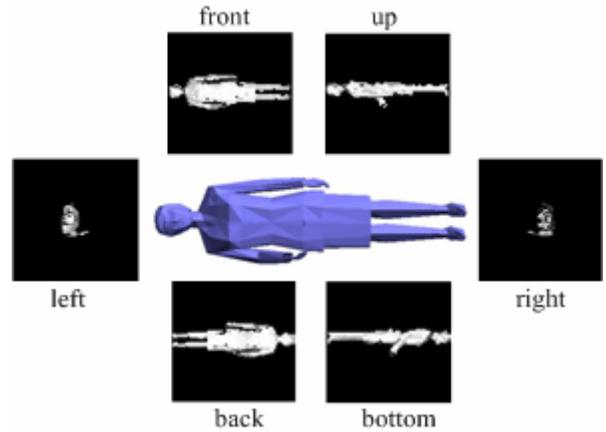


Figure 2. Sampling orientation of all surfaces on the 3D shape. The six figures are from six viewing planes including up and bottom, front and back, and left and right, respectively. The orientation value is in $[0, 1]$ and the gray value of the corresponding pixel belongs to $[0, 255]$.

Since this descriptor samples the orientation of the surfaces, the number of sampling could be less than other descriptors [2] [3] [7] [8] based on sampling the distances from the gravity center to the surfaces directly or indirectly. The proposed descriptor samples only the orientation of faces and requires fewer samples, because only one sample point could represent one face and additional samples are not needed. However, when sampling the distances from the face to the center, different points in the face can generate many different distances. Therefore, the face must be sampled several times. Please see Figure 3 in which (a) shows O view point samples orientation of the triangle once, (b) shows when sampling the distances between one triangle and the gravity center C , it must sample many different values d_1, d_2, d_3, d_4 , and so on for ensuring that the sampling rate is sufficient. We can see that this is one advantage of the proposed descriptor, which can reduce the sampling time and memory storage of sampling. In a later section we will present the results of experiments on the influence of sampling number on retrieval performance.

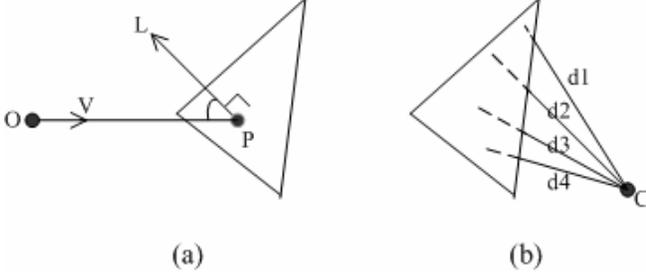


Figure 3. Difference between (a) sampling orientation and (b) sampling distances.

3.3 Multiresolution Wavelet Analysis

The wavelet is a useful mathematical tool for signal processing. The wavelet has two properties. The first is the unique ability to capture a local significant value in a small region of the space; on the other hand, it also has the frequency characteristics of Fourier transform.

Wavelet transform [21] and decomposition must be realized by virtue of one mother wavelet. We adopt the Daubechies function as the mother wavelet. The function bases are constructed by a linear combination of different scaling and translations of this wavelet. The scale function φ and wavelet function ψ are also defined according to the Daubechies wavelet, as follows.

$$\varphi_{s,m,n}(x,y) = 2^{s/2} \varphi(2^s x - m, 2^s y - n) \quad (4)$$

$$\psi_{s,m,n}^i(x,y) = 2^{s/2} \psi^i(2^s x - m, 2^s y - n), i=\{H,V,D\} \quad (5)$$

where s is the scale of the wavelet decomposition, m and n are the translations on the x -axis and y -axis, respectively, and ψ^H , ψ^V , and ψ^D are wavelets in the horizontal, vertical, and diagonal directions, respectively.

The orientation function $O(x,y)$ described in the above equation (3) is decomposed from the high scale $s+1$ to s by the Daubechies wavelet, and multiresolution analysis can be realized. The initial scale is the orientation function $O(x,y)$ with an initial resolution $N \times N$, $N=2^s$, which is decomposed iteratively by the following equations:

$$O_\varphi(s,m,n) = \frac{1}{2^s} \sum_{x=0}^{2^s-1} \sum_{y=0}^{2^s-1} O(s+1,x,y) \varphi_{s,m,n}(x,y) \quad (6)$$

$$O_\psi^i(s,m,n) = \frac{1}{2^s} \sum_{x=0}^{2^s-1} \sum_{y=0}^{2^s-1} O(s+1,x,y) \psi_{s,m,n}^i(x,y), i=\{H,V,D\} \quad (7)$$

where O_φ defines the approximation of the $s+1$ scale function O at the scale s by the scale function φ , and O_ψ^i ($i=\{H,V,D\}$) are the details in the horizontal, vertical, and diagonal directions, respectively.

Next we state the detailed implementation. Like Fourier transform, wavelet transform is separable from the two dimensional version of transform to a number of one dimensional transforms. This can simplify the complexity of transform from a high dimension to low dimension. For one 2D matrix, each of the rows can be transformed by one dimensional wavelet transform followed by the transformations of each of the columns. These two procedures

can therefore be used to perform wavelet transform on the orientation matrix. Unlike 2D Fourier transform, the 2D wavelet transform is composed of only simple addition operation and multiplication operation, no exponential operation. And the time complexity of 2D Fourier transform and wavelet transform is $O(n^2 \log n)$ and $O(n^2)$ respectively and therefore wavelet behaves well in computation speed.

For every viewing plane defined, we decompose the orientation function from the scale of $s=6$ to $s=2$. The scale of $s=6$ represents the initial state, that is, the orientation function described in equation (3). In addition, the scale of $s=2$ is the lowest resolution containing the main components at the lowest frequency. The following Table 1 shows the decomposition from the initial state to the two lowest scales, $s=3$, and $s=2$. Here, we omit the medium scales, $s=5$, and $s=4$. This table shows the wavelet decompositions of the front viewing plane. The leftmost column of this table is a front viewing orientation function, i.e., a 64×64 matrix, in which each entry is an orientation sampling value. And in the right columns of this table, there are two images of wavelet coefficients at two scales, and the image on the left is the second lowest scale $s=3$ which has the dimensions of 8×8 , and the image on the right is the lowest scale $s=2$ which has the dimensions of 4×4 . The whiteness represents the magnitudes of wavelet coefficients. Therefore, the whiter the pixel, the larger the wavelet coefficients.

Table 1. Wavelet decomposition under the two lowest scales. The leftmost is the image of the front view orientation and two right images are wavelet coefficients of the two scales.

Scale = 6	Scale = 3	Scale = 2

3.4 Dissimilarity metric

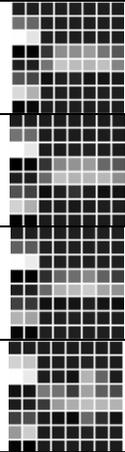
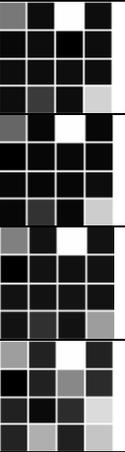
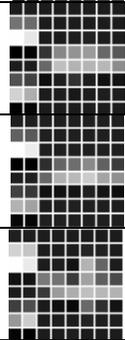
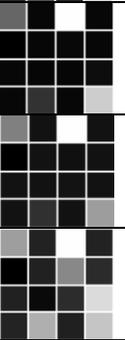
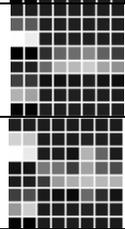
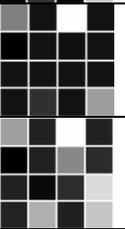
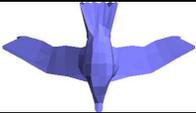
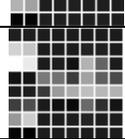
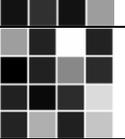
We use the wavelet coefficients of the final two scales as the feature vector V . Note that for the six view planes, there are six groups of wavelet coefficients and these groups of coefficients compose the feature vector V . Since the wavelet coefficients are close to the visual perception of human, we adopt the L_1 norm as the dissimilarity metric.

$$D = |V_1 - V_2| \quad (8)$$

Where V_1 is the feature vector of shape 1, and V_2 is the feature vector of shape 2.

We show several examples for three human shapes and one bird shape in the following Table 2. In addition, the corresponding final two-scale wavelet coefficients of the front viewing plane are also listed. In the left column, there are four shapes including three human shapes and one bird shape. The right column is one group of wavelet coefficients corresponding to the relative shape. The coefficients are obtained from multiresolution wavelet analysis of the front view plane, from which we can see that the coefficients of the three human shapes can be discriminated from the coefficients of the bird shape.

Table 2. Shapes and features. The left column is composed of shapes. On the right column, a part of the feature is shown as wavelet coefficients of front view plane at two lowest scales.

Shapes	Features (front view plane)	
	Scale = 3	Scale = 2
		
		
		
		

4. EXPERIMENTS

In this section, we conduct experiments on the Princeton Shape Benchmark, analyze the parameters, and choose the best parameters for retrieval. We name this descriptor MWA (Multiresolution Wavelet Analysis) descriptor. We also compare our MWA descriptor to other descriptors and discuss the advantages and disadvantages of this descriptor.

4.1 Retrieval results

We show some examples of retrieval results on the Princeton Shape Benchmark. We first chose several sets of 3D shapes randomly, including birds, sports cars, head models, humans and swords. From each set, given the first shape as a query, Figure 5 (located at the end of this paper) shows ten retrieved objects orderly that are most similar to query.

4.2 Parameter analysis

We chose three targets to analyze the retrieval performance for adjusting experiments parameters, the recall precision curves, the average recall precision, and the average computation time.

Recall precision curves have been used extensively in 3D retrieval methods. The precision is defined as the fraction of objects relevant to the input query, and the recall is given by the fraction of retrieved objects from the test database. The average recall precision is the average of the values of recall precision on all the shapes.

We tested the sampling influence on the retrieval performance. The sampling rate is defined as $N \times N$, and we investigated the following cases.

- 1) $N = 128$,
- 2) $N = 64$,
- 3) $N = 32$.

We then computed the recall precision curves, average recall precision and average computation time.

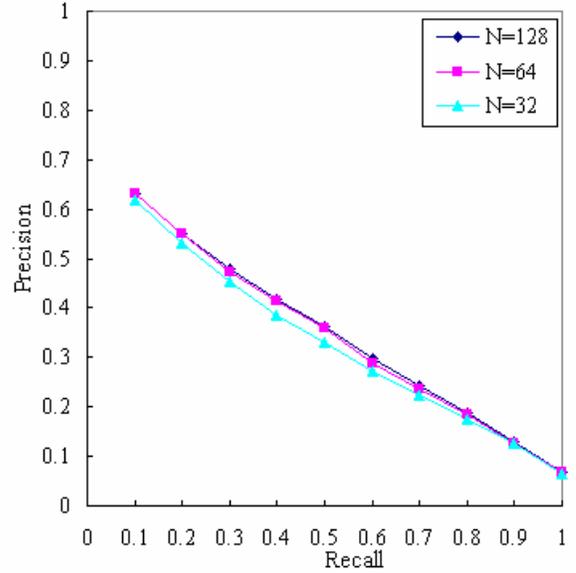


Figure 4. Recall precision curves

Table 3. Average recall precision and computation time

N	Average Recall Precision	Average Computation Time
128	33.65%	1,468ms
64	33.33%	839 ms
32	31.80%	521ms

We compared the retrieval effectiveness between the tests with a high sampling rate and two low rates, and a comparison of performance is shown in Figure 4 and Table 3. We found that even if the sampling rate is reduced from $N = 128$ to $N = 64$, the retrieval precision changed little. Furthermore, there is not remarkable decline in retrieval precision in spite of reducing the sampling to a great extent from $N = 128$ to $N = 32$. And also, the computation time can be economized in the low sampling rate.

Therefore, we decide to adopt the $N \times N$ ($N = 64$) as the recommended sampling rate for the balance on retrieval precision and sampling time.

4.3 Evaluation on the Princeton Shape Benchmark

We evaluated the proposed retrieval method, MWA descriptor, on the Princeton Shape Benchmark, which contains a collection of generic 3D models, and has been distributed via website.

We computed the quantitative statistics on seven recommended parameters, namely, Computation Time, Storage Size, and five tools for evaluating retrieval precision, Nearest Neighbor, First Tier, Second Tier, E-Measure, Discounted Cumulative Gain

(DCG), in order to evaluate the retrieval results. The statistics are summarized by averaging these five tools over all shapes in the data set. See Table 4 and Table 5.

Table 4. Computation time and storage size

Computation Time (ms)	Storage Size (byte)
839	1,920

Table 5. Retrieval precision

Nearest Neighbor	First Tier	Second Tier	E-Measure	DCG
58.0%	31.7%	41.3%	24.5%	58.4%

The storage size of a feature vector is measured by bytes. The average computation time, which is shown in the above Table 4, is obtained on a PC with a Pentium 2.0 G processor and 512 M of memory running Windows XP, and averaging the computation time over all 3D shapes. This condition is the same as that on which the Princeton Shape Benchmark runs.

We compared this MWA descriptor to the following methods on the same Princeton Shape Benchmark, the Spherical Wavelet Descriptor, which won the best paper award at SMI06, a near method to our descriptor, and two best-performing descriptors, Spherical Harmonics Descriptor, Light Field Descriptor, and analyzed the performance of this descriptor from several aspects.

1) Spherical Wavelet Descriptor (SWD)

SWD is based on the spherical function of sampling the surface distances to the gravity center, and the usage of spherical wavelet transform is proposed as a tool for the analysis of 3D shapes represented by functions on the unit sphere. And 512 spherical wavelet coefficients are extracted for the new descriptor. And in SWD paper, the storage length is given according to the dimension of the feature space, that is, the dimension of feature vector is 512. Here in the Table 6 we give the actual storage requirement of 512 float coefficients, that is, 2048 bytes.

2) Spherical Harmonics Descriptor (SHD)

In this method, the 3D shape is first voxelized into a grid. The spherical function $f_r(\theta, \phi)$ can be described by the intersection of the mesh with the respective voxel. This function is analyzed by spherical harmonics transform to obtain the rotational invariant descriptor as follows:

$$f_r(\theta, \phi) = \sum_{l \geq 0} \sum_{|m| \leq l} f_{l,m} Y_l^m(\theta, \phi) \quad (9)$$

3) Light Field Descriptor (LFD)

This is representative of reducing one 3D shape to a 2D space, generating projections of a 3D shapes from 100 viewing angles, and then encoding each projection as features by 35 Zernike moments and 10 Fourier coefficients. LFD represents the visual perception similar to that of humans, and is thought to be the best-performing descriptor on the Princeton Shape Benchmark.

Here, we investigated the retrieval performance of the three methods on the Princeton Shape Benchmark. The statistics on

SWD is taken from its paper, and the data of other two methods are from a summary of the survey on the Princeton Shape Benchmark. See Table 6 and Table 7.

Table 6. Computation time and storage size

Method	Computation Time (ms)	Storage Size (byte)
SWD	Unknown	2,048
SHD	1,690	2,184
LFD	3,250	4,700

Table 7. Retrieval precision

Method	Nearest Neighbor	First Tier	Second Tier	E-Measure	DCG
SWD	46.9%	31.4%	39.7%	20.5%	65.4%
SHD	55.6%	30.9%	41.1%	24.1%	58.4%
LFD	65.7%	38.0%	48.7%	28.0%	64.3%

4.4 Discussion

4.4.1 Storage analysis

We implement the proposed method, MWA descriptor based on sampling the shape orientation and the multiresolution wavelet analysis on the sampling result. As for each viewing plane, we adopt wavelet coefficients on the final two scales as the feature vector, and the storage of wavelet coefficients is $8 \times 8 + 4 \times 4$. Thus, the total storage size for the six viewing planes is 480 floating coefficients, or 1,920 bytes in total. The storage is slightly smaller than the SWD and the SHD. Compared with the LFD, MWA descriptor saves approximately 60% more storage space than the LFD.

4.4.2 Time analysis

The time consumption is concentrated on the sampling and multiresolution wavelet analysis. MWA descriptor samples the orientation of faces on the shape surface and need not sample many sampling points. As such, it is unlike SWD, based on sampling distances from surface to the gravity center. MWA adopt wavelet analysis needing smaller time complexity $O(n^2)$, however SHD adopt the spherical harmonics analysis with high computational complexity, which is determined by that of the associated Legendre transform, and the direct computation requires time of $O(n^3)$. And therefore, it is not strange that MWA descriptor is twice as fast as the SHD. The LFD must produce a large number of views for guaranteeing the retrieval precision, and therefore, it requires too much computation time to implement in real time retrieval.

4.4.3 Analysis of the retrieval precision

We listed the retrieval precision on five parameters. Our MWA descriptor performed better for the four anterior parameters, namely, Nearest Neighbor, First Tier, Second Tier, E-Measure, than the SWD, but the SWD performed better for the DCG parameter. In addition, MWA descriptor provides slightly better

discrimination than the SHD, but is inferior to the LFD, which supplies the better retrieval precision than MWA descriptor.

4.4.4 Merits and demerits of this descriptor

Compared to other methods, MWA descriptor provides better retrieval performance and also realizes a balance between space-time costs and retrieval precision. MWA descriptor is also robust with respect to various types of meshes. MWA descriptor has a few drawbacks. The most important of which may be that the descriptor does not provide the best retrieval performance on Princeton Shape Benchmark for generic 3D models. Another disadvantage is that MWA descriptor is dependent on shape normalization and increases the time cost on the shape alignment.

5. CONCLUSION AND FUTURE WORK

A new 3D shape descriptor, MWA descriptor, is proposed in the present paper. The proposed descriptor characterizes the orientations of shape faces, and performs multiresolution wavelet analysis on orientation signatures. The wavelet coefficients in low scales are used as the feature vector. We tested the MWA descriptor on Princeton Shape Benchmark, a data set designed for 3D generic models, and evaluated the retrieval performance about time-space complexities, and retrieval precision measurements by Nearest Neighbor, First Tier, Second Tier, E-Measure, and DCG. MWA descriptor achieves better retrieval precision and also reduces storage space and time costs. We believe that the proposed descriptor may be a good choice for the real-time retrieval applications.

In the future, we will consider improving this descriptor in characterizing shape orientation more efficiently. An idea is to get a rotational invariant descriptor which could reduce the normalization process, and enhance the retrieval precision of MWA descriptor. We plan to set the view points on a sphere, sample the orientation of surfaces, store the orientation signal into the spherical grid, and finally analyze the spherical signal by spherical tools such as spherical harmonics. The coefficients of spherical harmonics can be used as a rotational invariant descriptor. We think this idea can improve the descriptor in enhancing the robustness to affine transform, advancing the retrieval precision. However it is also a difficult problem how to avoid the influence that the small change in position of the sphere center can result in a significant noise in the feature descriptor.

6. ACKNOWLEDGEMENTS

This research is supported in part by Grants-in-Aid for Scientific Research, MEXT, Japan (Scientific Research (B), No. 19300238).

7. REFERENCES

- [1] Shilane, P., Min, P., Kazhdan, M., and Funkhouser, T.: The princeton shape benchmark. *IEEE International Conference on Shape Modeling and Applications*, (2004).
- [2] Kazhdan, M., Funkhouser, T., Rusinkiewicz, S.: Rotation invariant spherical harmonics representation of 3D shape descriptors. *Eurographics Symposium on Geometry Processing*, (2003).
- [3] Laga, H., Takahashi, H., and Nakajima, M.: Spherical wavelet descriptors for content-based 3D model retrieval. *IEEE International Conference on Shape Modeling and Applications*, (2006).
- [4] D.-Y. Chen, M. Ouhyoung, X.-P. Tian, and Y.-T. Shen: On visual similarity based on 3D model retrieval. *Computer Graphics Forum (EUROGRAPHICS 03)*, Vol. 22, No. 3, pp. 223-232, (2003).
- [5] Tangelder, J.-W.H. and Veltkamp, R.-C.: A Survey of content based 3D shape retrieval methods. *IEEE International Conference on Shape Modeling and Applications*, (2004).
- [6] Tangelder, J.-W.H. and Veltkamp, R.-C. A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications*, DOI 10.1007/s11042-007-0181-0, (2008).
- [7] Vranic, D.V., Saupé, D., and Richter, J.: Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics. *IEEE Workshop on Multimedia Signal Processing*, (2001).
- [8] Vranic, D.V.: An improvement of rotation invariant 3D shape descriptor based on functions on concentric spheres. *IEEE International Conference on Image Processing*, Vol.3, pp.757-760. (2003).
- [9] Shum H.: On 3D shape similarity. *International Conference on computer vision and pattern recognition*, (1996).
- [10] Novotni, M. and Klein, R.: 3D Zernike descriptors and content based shape retrieval. *8th ACM Symposium on Solid Modeling and Applications*, (2003).
- [11] Ankerst, M., Kastenmüller, G., Kriegel, H.P., and Seidl, T.: 3D shape histograms for similarity search and classification in spatial databases, *6th International Symposium on Spatial Databases (SSD 99)*, (1999).
- [12] Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D.: Shape distributions, *ACM Transactions on Graphics*, Vol.21, No.4, (2002).
- [13] Liu, Y., Zha, H. and Qin, H.: The generalized shape distributions for shape matching and analysis, *IEEE International Conference on Shape Modeling and Applications*, (2006).
- [14] Tung, T., and Schmitt, F.: The augmented multiresolution Reeb graph approach for content-based retrieval of 3D shapes. *International Journal of Shape Modeling*, Vol.11, No.1, (2005).
- [15] Jain V., and Zhang H.: A spectral approach to shape-based retrieval of articulated 3D models. *Computer Aided Design*, 39, 398-407, (2007).
- [16] Ben-Chen M., and Gotsman C.: Characterizing shape using conformal factors. *Eurographics Workshop on 3D Object Retrieval*, (2008).
- [17] McGill 3D shape benchmark
<http://www.cim.mcgill.ca/~shape/benchMark/>
- [18] Paquet, E., Rioux, M., Murching, A., Naveen, T., and Tabatabai, A.: Description of shape information for 2-D and 3-D objects. *Signal Processing: Image Communication*, Vol.16, No.1-2, (2000).
- [19] Chen, D.Y. and Ouhyoung, M.: A 3D model alignment and retrieval system. *Proc. of International Computer Symposium*, (2002).
- [20] Jolliffe, I.T.: Principal component analysis. *Springer*, (1986).
- [21] Gonzalez, R.C., Woods, R.E.: Digital image processing, third edition. *Prentice Hall*, (2008).



Figure 5. Retrieval examples. The leftmost column is the query column and the 10 columns to the right are the retrieval results ordering by similarity.