A New MPEG-7 Standard: Perceptual 3-D Shape Descriptor

Duck Hoon Kim[1], In Kyu Park[2], Il Dong Yun[3], and Sang Uk Lee[1]

[1] School of Electrical Engineering and Computer Science, Seoul National University, SEOUL 151-742, KOREA
ducks@diehard.snu.ac.kr, sanguk@ipl.snu.ac.kr

[2] School of Information and Communication Engineering, INHA University, INCHEON 402-751, KOREA
pik@ieee.org

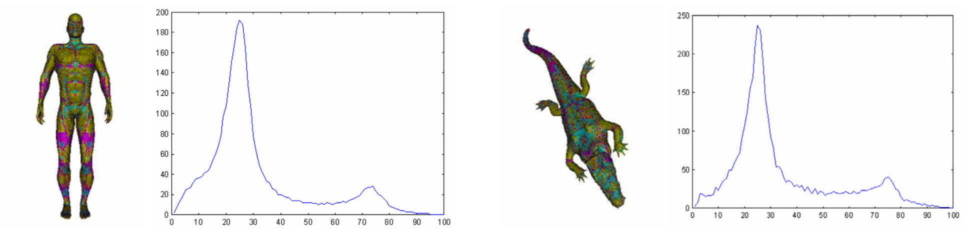
[3] School of Electronic and Information Engineering, Hankuk University of Foreign Studies, YONGIN 449-712, KOREA
yun@hufs.ac.kr

Abstract. In this paper, we introduce the perceptual 3-D shape descriptor (P3DS) for 3-D object, which has been proposed and adopted as an MPEG-7 standard recently. The descriptor is used as an abstract representation of 3-D object in content-based shape retrieval system. Unlike the conventional descriptors, the P3DS descriptor supports the functionalities like *Query by Sketch* and *Query by Editing* which are very useful in real retrieval system. Given a couple of the descriptors, a matching technique is also developed to measure the similarity between them. High retrieval score has been observed when the developed descriptor and the matching technique are used in the retrieval system with the MPEG test database.

1 Introduction

As the applications of 3-D computer graphics become more popular in human life, the amount of 3-D object we need to handle increases day by day. In order to reuse the existing 3-D objects, they should be stored in an archive and retrieved again for further use when necessary. Therefore, efficient methods for managing 3-D objects are quite necessary. In this context, multimedia industry has been developing an international standard for multimedia contents description and retrieval, known as MPEG-7. Recently, MPEG-7 community has witnessed the necessity of a description scheme for 3-D object. In this paper, we introduce the perceptual 3-D shape descriptor (P3DS) for 3-D object, which has been proposed and adopted as an MPEG-7 standard recently [1,2,3,4,5,6,7]. Note that an amendment of the P3DS descriptor has been in progress and currently in Working Draft stage [7]. The P3DS descriptor provides the abstract and compact representation of the conventional objects in polygonal mesh, B-spline surface, and voxel representation. The topological structure of 3-D object is preserved in the P3DS descriptor, providing an interactive query design in content-based retrieval system. Based on the perceptual description as a building block, an efficient attributed relational graph matching method, *i.e.* Double Earth Mover's Distance, is proposed and optimized in the shape retrieval system.

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3332, pp. 238–245, 2004.



(a) (b)

Fig. 1. The limitation of the discriminative ability in the Shape3D descriptor: (a) the android and its shape spectrum and (b) the crocodile and its shape spectrum.

2 Previous 3-D Shape Descriptors

In MPEG-7, the Shape3D descriptor has been developed and adopted as an international standard for description and browsing of 3-D VRML (Virtual Reality Markup Language) database [8]. The Shape3D descriptor implies the shape spectrum of 3-D mesh object, which is the histogram of the shape indices calculated over the entire mesh. It yields high retrieval scores for the ground truth data set [9]. However, the Shape3D descriptor has a few drawbacks. First, the shape spectrum, *i.e.* the histogram of the shape indices, represents only local geometries of the 3-D surfaces so loses the spatial information. Hence, the discriminative ability may be limited. As shown in Fig. 1, the android and the crocodile have different topological structures with different shape. However, two objects have very similar shape spectrums, *i.e.* the distance between two objects is only 0.025981. As a consequence, the crocodile is probably retrieved when the android is selected as query, which is an undesirable result. Next, the similarity measure of two shape spectrums can not evaluate correct distance since neighboring bins are not considered when there is no match between the exact corresponding bins.

In academia, a lot of research groups have developed various 3-D shape descriptors. Especially, people in the research group of Princeton University have been showing remarkable activities in 3-D shape retrieval [10,11,12,13]. More specifically, they have developed various descriptors [10,11,12], proposed the benchmark framework for 3-D shape retrieval, and performed the experiments for the comparison between existing 3-D shape descriptors [13]. However, their 3-D shape descriptors are mainly based on the distribution of vertices or faces so that it is difficult to provide the user-friendly querying interface such as *Query by Sketch* and *Query by Editing* which will be described in Section 3.3.

3 Perceptual 3-D Shape (P3DS) Descriptor

In order to overcome the drawbacks of the previous descriptors, we develop the perceptual description based on the part-based representation of a given object. The part-based representation is expressed by means of an attributed relational graph (ARG) which can be easily converted into the P3DS descriptor. More specifically, the P3DS descriptor is

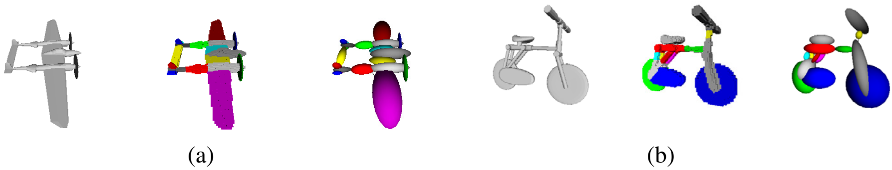


Fig. 2. Part-based and ARG representations of different 3-D objects. From left to right in (a) and (b), original mesh object, part-based representation, and ARG are shown sequentially.

designed to represent and identify 3-D objects based on the part-based simplified representation using ellipsoidal blobs. Actually, it is based on the assumption that the part-based representation and the actual shape are coherent with human visual perception. In this context, if volume-based decomposition is adopted as the part-based representation, the object is assumed to have its own volume, *e.g.* the object is composed of one or more closed mesh surface. On the other hand, in the case of mesh-based decomposition, the object is assumed to be manifold, *i.e.* an edge is shared by two triangles, unless it belongs to the boundary. For the fast processing and better result, it is recommended to use manifold mesh object with no hole. As expected, if the encoder does not produce the part-based representation properly, the retrieval performance would not be good. In this section, the P3DS descriptor is introduced with an example of its former self, *i.e.* ARG, and with the formal syntax in MPEG-7 standard. Moreover, new functionalities of the P3DS descriptor are addressed as well.

3.1 Node and Edge Features in the ARG

The perceptual description of a given 3-D object is generated from the part-based representation. It has the form of an ARG, composed of a few nodes and edges. A few examples of the part-based representation and ARG are shown in Fig. 2, where the morphological voxel-based decomposition [14] is adopted as the part-based representation.

A node represents a meaningful part of the object with unary attributes, while an edge implies binary relations between nodes. In the descriptor, there are 4 unary attributes and 3 binary relations which are derived from the geometric relation between the principal axes of the connected nodes. In detail, a node is represented by an ellipsoid parameterized by volume V, convexity C, and two eccentricity values E_1 and E_2. More specifically, the convexity is defined as the ratio of the volume in a node to that in its convex hull, and the eccentricity is composed of two coefficients, $E_1 = \sqrt{1 - c^2/a^2}$ and $E_2 = \sqrt{1 - c^2/b^2}$, where a, b, and c $(a \geq b \geq c)$ are the maximum ranges along 1st, 2nd, and 3rd principal axes, respectively. Edge features, *i.e.* binary relations between two nodes, are extracted from the geometric relation between two ellipsoids, in which the distance between centers of connected ellipsoids and two angles are used. The first angle is between first principal axes of connected ellipsoids and the second one is between second principal axes of them.

```
Perceptual3DShape {                                          Number of bits   Mnemonics
   numberOfNodes                                             8                uimsbf
   BitsPerAttribute                                          4                uimsbf
   for ( i = 0 ; i < (numberOfNodes^2 - numberOfNodes) / 2 ; i++ ) {
      IsAdjacent[i]                                          1                bslbf
   }
   for ( i = 0 ; i < numberOfNodes ; i++ ) {
      Volume[i]                                              BitsPerAttribute uimsbf
      Center_X[i]                                            BitsPerAttribute uimsbf
      Center_Y[i]                                            BitsPerAttribute uimsbf
      Center_Z[i]                                            BitsPerAttribute uimsbf
      Transform_1[i]                                         BitsPerAttribute uimsbf
      Transform_2[i]                                         BitsPerAttribute uimsbf
      Transform_3[i]                                         BitsPerAttribute uimsbf
      Transform_4[i]                                         BitsPerAttribute uimsbf
      Transform_5[i]                                         BitsPerAttribute uimsbf
      Transform_6[i]                                         BitsPerAttribute uimsbf
      Variance_X[i]                                          BitsPerAttribute uimsbf
      Variance_Y[i]                                          BitsPerAttribute uimsbf
      Variance_Z[i]                                          BitsPerAttribute uimsbf
      Convexity[i]                                           BitsPerAttribute uimsbf
   }
}
```

Fig. 3. The binary representation syntax of the P3DS descriptor.

3.2 Binary Representation of the P3DS Descriptor

In order to preserve the topological shape of the object and to adopt the user-friendly querying interface in the retrieval system, the actual descriptor utilizes slightly different attributes which can provide the shape and relation of ellipsoidal blobs more intuitively. In detail, the P3DS descriptor contains three node attributes, such as Volume, Variance, and Convexity, which can be converted easily into the 4 unary attributes. Next, it contains two edge attributes, such as Center and Transform, from which the 3 binary relations can be computed as well. Fig. 3 shows the Binary Representation Syntax of the P3DS descriptor. Actually, the size of the descriptor is very compact. For an example, when an ARG has 5 nodes, the P3DS descriptor, of which the attributes are quantized in 8-bit (default), can be represented by only 582 bits. Note that Volume, Center, Variance and Convexity are normalized in the interval [0, 1], while Transform is normalized in the interval [-1,1].

3.3 New Functionalities for 3-D Shape Retrieval

Using the P3DS descriptor, we believe that it would produce a few new functionalities of 3-D shape retrieval, which cannot be provided by the previous descriptors. Note that the most important advantage of the P3DS descriptor is that there is the exact coincidence between the perceptual description in the P3DS descriptor and human cognitive description. Therefore, when we see the descriptor, we can understand the topological shape of

3-D object and also expect the proper retrieval result. For example, if the descriptor has some information of 5 parts in which four of them are connected to a main body, then we can perceive that the descriptor represents animal-like shape and we would expect the retrieval results of such animals. The human readability of the descriptor enables us to design new kinds of querying strategy as follows: Since the part-based representation is very easy to build and edit, users can make a simple shape consisting of ellipsoidal blobs and their connections. The built object is easily translated to the P3DS descriptor and put into the search engine, *i.e. Query by Sketch*. Similarly, users can modify the existing descriptor in an interactive way and then they can try a new retrieval to get different results, *i.e. Query by Editing*.

4 Shape Matching Using the P3DS Descriptor

The similarity measurement of two P3DS descriptors is the most basic procedure since the input query is compared with the model in the database one by one using the P3DS descriptor and then the models with high similarity are browsed. The one-to-one comparison of two P3DS descriptors consists of three steps: (1) Forming ARG's from the descriptors, (2) Defining the Volume as weight in each node, (3) Comparing the ARG's using the Double EMD (Earth Mover's Distance [15]) method. During the Double EMD procedure, a distance matrix is first generated between query nodes of a query ARG and model nodes of a model ARG (Inner EMD), followed by measuring the similarity by calculating the amount of work required to move the weights from the query nodes to the model nodes using the conventional EMD method based on the distance matrix (Outer EMD).

The Inner EMD is defined as follows. First, the unary-distance matrix is constructed by computing the distance between the unary attribute vectors of the query and model ARG's. Next, to define the difference of binary relations between nodes of query and model ARG's, *i.e.* N_q and N_m in Fig. 4(a), as one element of the binary-distance matrix, the binary relations are utilized to form the axes of the vector space. More specifically, the coordinate system in Fig. 4(a) consists of three axes, such as one distance and two angles defined in Section 3.1. Then N_q and its connected nodes form a set of points in the vector space, while N_m and its connected nodes form another set of points. In this context, N_q and N_m are located at the origin of the vector space. Finally, an imaginary point, *i.e.* empty circles in Fig. 4(a), of which the distance from any point is equal to d in Fig. 4(b), is provided to make the sum of the weights of N_q and its connected nodes and that of N_m and its connected nodes be equal. Note that the imaginary point can penalize the weight transition to other unconnected nodes. As shown in Fig. 4(b), a distance matrix with respect to N_q and N_m is constructed from the Euclidean distances between nodes of query and model ARG's. Then the conventional EMD based on this distance matrix yields one element of the binary-distance matrix. In this way, each element of the binary-distance matrix can be computed. Finally, the final distance matrix is computed by adding the unary-distance matrix and the binary-distance matrix.

In the Outer EMD, the dissimilarity between the query and model ARG's is measured by calculating the amount of work required to move the weights from the query nodes

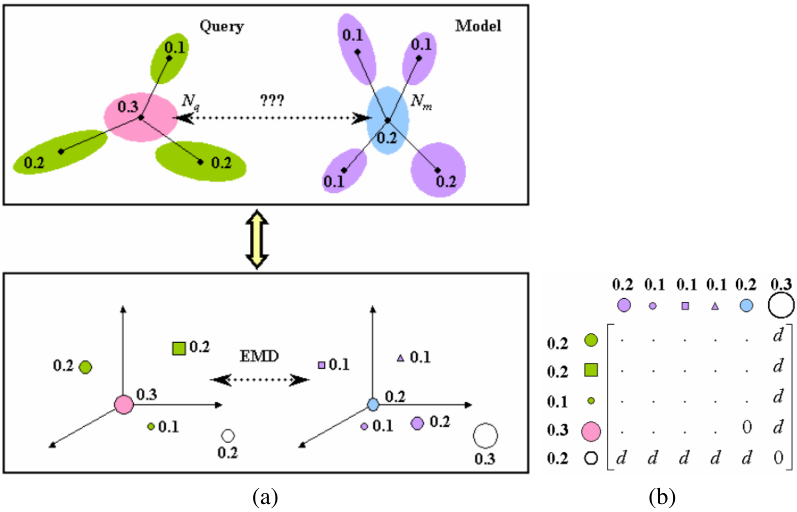


Fig. 4. An example of Inner EMD: (a) vector space representation for computing the Inner EMD and (b) a distance matrix in the procedure of calculating one element of the binary-distance matrix considering N_q and N_m.

to the model nodes based on the final distance matrix. In other words, a total amount of work for all of the nodes refers to the dissimilarity between the two ARG's.

5 Experimental Results

In order to evaluate the performance of the developed 3-D shape retrieval system using the P3DS descriptor, intensive experiments are carried out on the MPEG-7 official 3-D VRML database including 3,900 objects. The objects in the database are classified into 102 categories in a hierarchical structure. For the query dataset, we also use the MPEG-7 query dataset composed of 336 objects in 10 leaf categories from 8 broad categories. Note that MPEG's basic rule to select the leaf categories is that the objects should have perceptually meaningful ARG structure. In other words, the object itself and its ARG should be similar enough based on the human visual ability in object recognition. In the experiment, we show the retrieval performance and the comparison with the Shape3D descriptor. Note that d in Section 4 is set to 1 experimentally. As shown in Table 1, it is observed that the performance of the P3DS descriptor is quite good for both Bull's Eye Performance (BEP) and Average Normalized Modified Retrieval Rate (ANMRR) [8], also significantly better than the Shape3D descriptor. Note that higher BEP score and lower ANMRR score imply better performance. Moreover, the complexity of the retrieval is quite endurable. The retrieval time depends on the complexity of the object structure. For most cases, it would take less than 5 seconds on a PC with Pentium IV CPU. Currently, we use simple one-to-one matching for the whole objects in the database. We believe it is necessary to develop an algorithm for fast indexing which fully utilizes the hierarchical database structure.

Table 1. Retrieval performance and comparison with the Shape3D descriptor (Higher BEP and lower ANMRR scores show better performance)

Category		P3DS descriptor		Shape3D descriptor	
		BEP	ANMRR	BEP	ANMRR
Aircraft / multi_fuselages / 3_bodies	30	0.63667	0.27326	0.32111	0.51363
Animal / arthropod / with_wings / bee	30	1.00000	0.00000	0.55556	0.31395
Animal / humanoid / sitting	60	0.80972	0.17468	0.55556	0.35791
Automobile / tank / equipvaried	30	1.00000	0.00007	0.53222	0.32587
Furniture / chair / 4_legged	30	0.58556	0.32306	0.34333	0.57812
Furniture / chair / with_a_post	30	0.68333	0.24401	0.25000	0.60753
Letter / O	15	0.58667	0.29977	0.32000	0.51549
Plant / flower / 20_petaled	30	0.66111	0.23979	0.49222	0.38677
Ship / single_mast / romanship	30	0.93333	0.05263	0.33778	0.49117
Simplex cellular_phone	51	0.85621	0.11558	0.36140	0.47758
Total	336	0.79182	0.16326	0.42122	0.44665

Table 2. Influence of the quantization steps in the P3DS descriptor for BEP score (Higher BEP score shows better performance)

Category		BEP				
		8 bits	7 bits	6 bits	5 bits	4 bits
Aircraft / multi_fuselages / 3_bodies	30	0.63667	0.63222	0.64444	0.66444	0.55556
Animal / arthropod / with_wings / bee	30	1.00000	1.00000	1.00000	1.00000	1.00000
Animal / humanoid / sitting	60	0.80833	0.80861	0.80000	0.74861	0.60528
Automobile / tank / equipvaried	30	1.00000	1.00000	1.00000	1.00000	0.98778
Furniture / chair / 4_legged	30	0.59000	0.59556	0.59556	0.60667	0.49667
Furniture / chair / with_a_post	30	0.68000	0.67444	0.68222	0.67111	0.68444
Letter / O	15	0.59111	0.59111	0.58667	0.52889	0.43556
Plant / flower / 20_petaled	30	0.66000	0.66778	0.67333	0.62333	0.56667
Ship / single_mast / romanship	30	0.93222	0.93111	0.93111	0.93222	0.92333
Simplex cellular_phone	51	0.85659	0.85429	0.85275	0.83045	0.80238
Total	336	0.79132	0.79162	0.79194	0.77422	0.71490

Actually, the P3DS descriptor consists of an array of floating point (32bits) attributes. In this context, the quantization of the attributes with lower number of bits should be considered in order to achieve more compact representation. In our implementation, we perform an uniform quantization with a number of steps 2^b, where b denotes the number of quantization bits. Table 2 presents the influence of the quantization bits to the overall retrieval performance. Note that 8 bits uniform quantization produces very little degradation of the performance compared with the results without quantization. Moreover, lower bits representation up to 5 bits still produces quite little degradation, while the performance goes down much at 4 bits.

6 Conclusion

In this paper, we introduced the Perceptual 3-D Shape (P3DS) descriptor and presented the various experimental results not only in shape retrieval but also in the effect of the attribute quantization. In MPEG-7 community, the P3DS descriptor has been proposed and adopted as a new standard for 3-D shape description and currently in Working Draft stage. Our future work includes the development of practical retrieval system which can be used in real application. The combination of context-based (keyword-based) and content-based retrieval system should be taken into account. Furthermore, some scheme like relevance feedback is quite necessary in order to increase the subjective performance to real user.

References

1. Kim, D.H., Yun, I.D., Park, I.K., Kim, D.K.: Perceptual description for 3D object indexing and retrieval. ISO/IEC JTC1/SC29/WG11 M8980 (2002)
2. Park, I.K., Kim, D.K., Kim, D.H., Yun, I.D., Lee, S.U.: 3D perceptual shape descriptor: Result of exploration experiments and proposal for core experiments. ISO/IEC JTC1/SC29/WG11 M9210 (2002)
3. Park, I.K., Kim, D.K., Kim, D.H., Yun, I.D., Lee, S.U.: Perceptual 3D shape descriptor: Result of core experiment. ISO/IEC JTC1/SC29/WG11 M9481 (2003)
4. Park, I.K., Kim, D.H., Yun, I.D., Lee, S.U.: Perceptual 3D shape descriptor: Result of core experiment. ISO/IEC JTC1/SC29/WG11 M9809 (2003)
5. Park, I.K., Kim, D.H., Yun, I.D.: Perceptual 3D shape descriptor: Result of core experiment. ISO/IEC JTC1/SC29/WG11 M10093 (2003)
6. Park, I.K., Kim, D.H., Yun, I.D.: Perceptual 3D shape descriptor: Result of core experiment. ISO/IEC JTC1/SC29/WG11 M10324 (2003)
7. Yamada, A., Kim, S.K.: WD 1.0 of MPEG-7 new visual extensions. ISO/IEC JTC1/SC29/WG11 N6367 (2004)
8. Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7. John Wiley & Sons (2002)
9. Zaharia, T., Preteux, F.: Results of 3D shape core experiment. ISO/IEC JTC1/SC29/WG11 M6315 (2000)
10. Osada, R., Funkhouser, T., Chazelle, B., Dobkin, D.: Shape distributions. ACM Transactions on Graphics **21** (2002) 807–832
11. Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., Jacobs, D.: A search engine for 3D models. ACM Transactions on Graphics **22** (2003) 83–105
12. Kazhdan, M., Chazelle, B., Dobkin, D., Funkhouser, T., Rusinkiewicz, S.: A reflective symmetry descriptor for 3D models. Algorithmica **38** (2003)
13. Shilane, P., Min, P., Kazhdan, M., Funkhouser, T.: The princeton shape benchmark. In: Proceedings of Shape Modeling International 2004, IEEE Computer Society (2004) 167–178
14. Kim, D.H., Yun, I.D., Lee, S.U.: A new shape decomposition scheme for graph-based representation. (to appear in Pattern Recognition)
15. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision **40** (2000) 99–121