ONE-DIMENSIONAL SIGNATURE REPRESENTATION FOR THREE-DIMENSIONAL CONVEX OBJECT RECOGNITION

Ho-Jin Lee[1], Dong-Gyu Sim[2] and Rae-Hong Park[1]
[1]Department of Electronic Engineering, Sogang University, Seoul, Korea
Hojin.lee@lsgen.com, rhpark@sogang.ac.kr
[2]Department of Computer Engineering, Kwangwoon University, Seoul, Korea
dgsim@kw.ac.kr

ABSTRACT

A simple method to represent three-dimensional (3-D) convex objects is proposed, in which a one-dimensional signature based on the discrete Fourier transform is used to efficiently describe the shape of a convex object. It has position-, orientation-, and scale-invariant properties. Experimental results with synthesized 3-D simple convex objects are given to show the effectiveness of the proposed simple signature representation.

KEYWORDS

Convex Object, Invariance, Discrete Fourier Transform, Signature

1. INTRODUCTION

Representation of an object is a first step in three-dimensional (3-D) computer vision and object recognition [1], in which two-dimensional (2-D) shape features based on their complementary property are used. A good representation method makes it easy to retrieve, classify, and/or recognize input objects from database. A 3-D object in a 3-D space can be described using various features such as shape, position, and orientation, among which the shape information is useful for both retrieval and recognition of an object [2, 3]. Some approaches to content-based representation of a 2-D image have also been presented. A one-dimensional (1-D) signature to represent an image for content-based retrieval from image databases was proposed [4]. Fourier-Mellin transformation approximation was used to describe an image, in which Fourier power spectrum was computed for pattern recognition, reconstruction, and image database retrieval [5]. Both methods have desirable representation properties invariant to geometric transformations such as translation, rotation, and scaling.

In a 3-D convex object, points on the straight line that joins any two points are also in the object [6]. The centroid of a convex object always lies in the object. Convex objects include solid cube, convex polyhedral, quadrics, superquadrics, and hyperquadrics. For robotics applications, computing the distance between convex objects was presented by nonuniform rational B-spline curves or patches [7] and by interior point approach [8].

A new shape decomposition method of an object was proposed under some concavity constraints, giving a compact geometrical and topological representation [9]. The more extended Gaussian image model from range data was proposed for recognition of 3-D objects including convex and concave shapes [10]. Also, hierarchical extended Gaussian image (EGI) was used to describe nonconvex as well as convex objects [11].

DOI : 10.5121/ijcga.2013.3201

Efficient representation of an object is a fundamental and important step in 3-D computer vision and pattern recognition. A good representation method leads to effective retrieval and recognition of 3-D objects, thus a variety of 3-D object representation approaches have been proposed. Kang and Ikeuchi [12] proposed a complex extended Gaussian image (CEGI) to overcome the disadvantages of the conventional EGI. Hebert *et al*. [13] defined a spherical attribute image to represent 3-D objects. Also, a 3-D feature descriptor using concentric ring signature was used to represent local topologies of 3-D point clouds, which was applied to 3-D shape matching and retrieval [14].

This paper presents an efficient method to represent 3-D convex objects. It is an extended version of the previous paper [15], in which a 1-D signature is further proposed to effectively describe the shape of a 3-D convex object. Since it is position-, orientation-, and scale-invariant, application to 3-D object retrieval and recognition is straightforward. The series of operations to the input object gives the simple 1-D signature.

The rest of the paper is structured as follows. Section 2 presents the proposed 1-D signature representation method, which is invariant to position, orientation, and scaling. Each step of the proposed representation is explained in terms of invariant properties, which are desirable for 3-D object recognition. Experimental results with synthetic convex test images are shown in Section 3. Finally, Section 4 gives conclusion.

2. PROPOSED 1-D SIGNATURE REPRESENTATION

The proposed 1-D signature representation method presents three invariant properties of feature values: position-, rotation-, and scale-invariant. Figure 1 shows the block diagram of the proposed method. Each step for 3-D to 1-D representation is described in the following subsections.

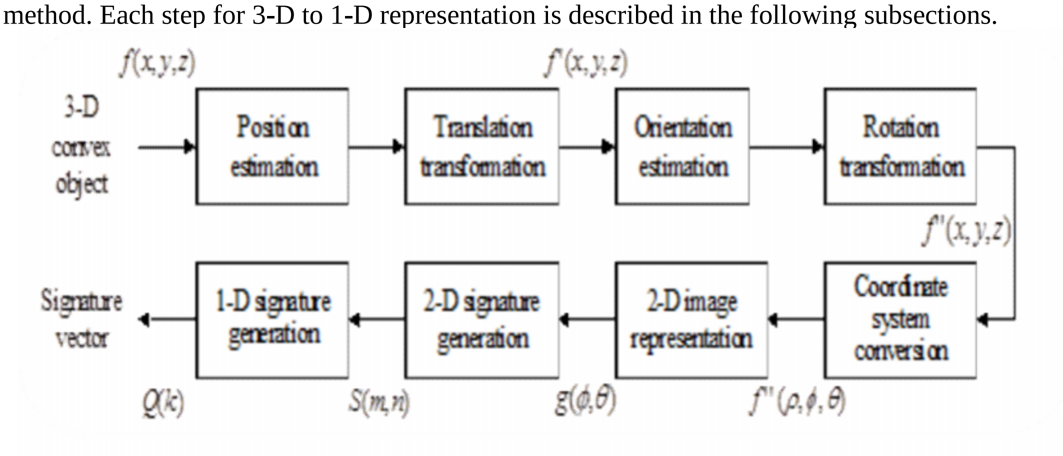

Figure 1. System block diagram.

2.1. Input Object

A 3-D object is represented as a set of cells in a 3-D coordinate system. Thus, a 3-D signal can be defined to represent a 3-D object. A simple voxel-based spatial occupancy representation is used to represent a 3-D input object [16, 17]. Voxels are arranged and indexed in the 3-D Cartesian coordinate system. Information or property of an object such as shape, position, and orientation can be defined using the distribution of the voxels that are marked as inside the object. Especially, position and orientation can be parameterized with some vectors and angles [15].

Figure 2 shows each step of the proposed 1-D signature generation procedure. Figure 2(a) shows the definition of the parameters used for 3-D object representation, where for easy understanding of the whole shape of the object, surface based representation is used, with $f(x,y,z)$ denoting the input object, a 3-D signal that has only two values 0 or 1. If position (x_1,y_1,z_1) is occupied by the object, in voxel-based spatial occupancy representation (x_1,y_1,z_1)=1, and 0, otherwise. In this paper, this binary 3-D signal representation is used to represent an object.

The position and orientation of an object can be described with two vectors and an angle. Vectors $\boldsymbol{v}_c$, $\boldsymbol{v}_a$, and angle ζ denote the centroid, the principal axis of the object, and the rotation angle about $\boldsymbol{v}_a$, respectively. The vector $\boldsymbol{v}_c$ signifies the parameterized position of the object centroid and can be obtained by averaging the 3-D coordinates of the object voxels whereas $\boldsymbol{v}_a$ and ζ represent the parameterized orientation of the object. The vector $\boldsymbol{v}_a$ can be estimated as a dominant eigen vector of the 3-D coordinate distribution of the object voxels.

2.2. Position-Invariant Property

Position invariance is obtained at the stage of coordinate conversion. After obtaining the centroid (x_c, y_c, z_c) of the input object $f(x,y,z)$, new coordinates whose origin is at (x_c, y_c, z_c) is defined to represent the volume elements of an input object. Translation of $f(x,y,z)$ to the centroid (x_c, y_c, z_c) expressed as

$$f' = T_{x_c, y_c, z_c}[f]$$

is the first transformation performed on the input object $f(x,y,z)$ for position-invariant object shape description, where T denotes the transformation describing the translation process. Figure 2(b) shows the translated object $f'(x,y,z)$. Note that $f'(x,y,z)$ has only the shape and orientation information of the object.

2.3. Orientation-Invariant Property (1)

Orientation- and scale-invariant properties are obtained after defining feature values. Orientation of an input object can be parameterized using two parameters $\boldsymbol{v}_a$ and ζ. The direction of the principal axis is estimated by the eigenvector of the matrix corresponding to the largest eigenvalue, where the matrix is constructed by a set of coordinates of a number of volume elements in the object. The orientation information of the principal axis can be eliminated by the rotation transformation, which is defined by

$$f'' = R_{v_n, -\tau}[f'],$$

where R denotes the rotation operation with τ denoting the inclination angle of the principal axis from the z-axis as described in Figure 2(b). Rotation is performed about the axis $\boldsymbol{v}_n$ that is perpendicular to the surface defined by $\boldsymbol{v}_a$ and the z-axis. Figure 2(c) shows the result of the rotation transformation of the translated object $f'(x,y,z)$ shown in Figure 2(b).

2.4. Coordinate System Conversion

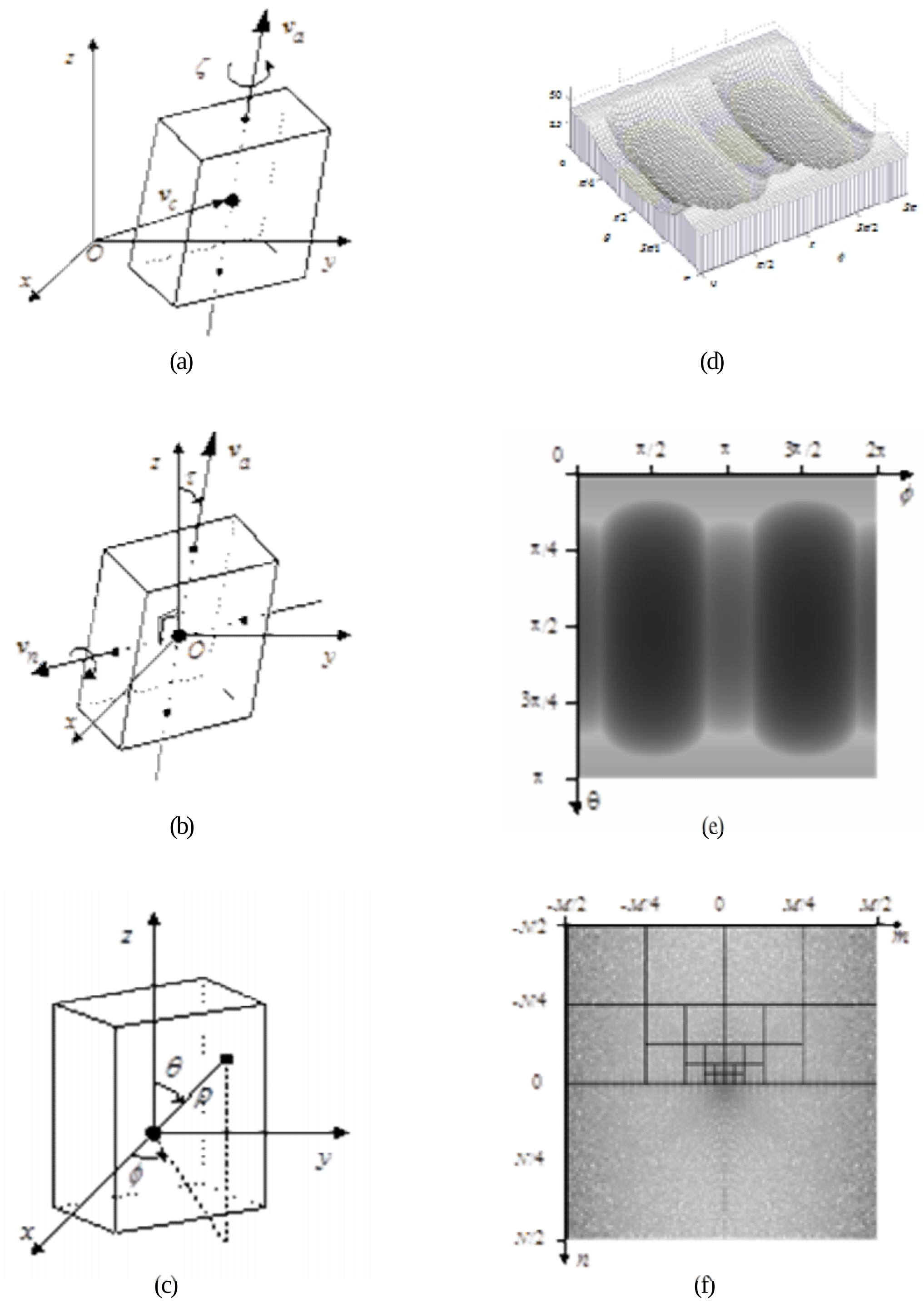

Figure 2. Signature generation procedure. (a) $f(x, y, z)$, (b) $f'(x, y, z)$, (c) $f''(x, y, z)$, (d) $f''(\rho, \phi, \theta)$, (e) $g(\phi, \theta)$, (f) $S(m,n)$.

Coordinate system conversion is needed to define rotation-invariant features in two steps: 1) estimation of the center coordinates and the principal axis direction of the input object, 2) new object representation using the estimated center coordinates and the principal axis, and the spherical coordinate conversion. The invariant properties to position and the principal axis orientation can be obtained by a series of geometric transformations (i.e., translation and rotation) as mentioned above. The other orientation angle ζ, however, still remains in $f'(x,y,z)$. To be able to eliminate this factor, coordinate system conversion from Cartesian to spherical is performed to define rotation-invariant features, with the newly defined origin and the principal axis of the input object. In this new coordinate system, an input object is located at the origin of the coordinate system. The relationship between the variables of the Cartesian coordinate (x, y, z) and the spherical coordinate (ρ, ϕ, θ) [15] is shown in Figure 2(c), in which rotation parameter ζ in the Cartesian coordinate system is converted to the translation parameter in the spherical coordinate system. After the coordinate conversion, the 3-D signal that represents an object is expressed as $f'(\rho, \phi, \theta)$.

At the next step, this property is combined with a property of the discrete Fourier transform (DFT) [18] to construct a 2-D signature image that is invariant to object orientation. Figure 2(d) shows the volumetric representation of the object voxels in the spherical coordinate system.

2.5. Equivalent 2-D Representation

If the input object $f(x,y,z)$ satisfies the convexity, the volumetric representation $f'(x,y,z)$ transformed thus far has no hole in it, so that 2-D equivalent representation of $f'(x,y,z)$ can be expressed as [15]

$$g(\phi,\theta) = \max_{f''(\rho,\phi,\theta)\neq 0}(\rho) \ ,$$

where $g(\phi, \theta)$ represents the 2-D image signal of which independent variables are ϕ and θ, as shown in Figure 2(e). This process changes object representation from volume-based to surface-based, reducing the data complexity. The process of the data complexity reduction does not cause any loss of information. The 2-D representation $g(\phi, \theta)$ has the surface information of an input object. Note that data complexity reduction is achieved by representing an object from 3-D to 2-D. In Figure 2(e), brighter gray level values are used to represent larger signal values.

2.6. Orientation-Invariant Property (2)

The orientation factor in $g(\phi, \theta)$ denoted by the angle ζ can be eliminated by using the property of the DFT [12]: ζ in the Cartesian coordinate is converted into the translation parameter. If the signature image is defined by the power spectra of $g(\phi, \theta)$, the orientation angle ζ is eliminated. The 2-D signature image $P(m,n)$ is defined by

$$P(m,n) = \sqrt{\mathrm{Re}[G(m,n)]^2 + \mathrm{Im}[G(m,n)]^2} \ ,$$

where $-M/2 \leq m \leq M/2-1$, $-N/2 \leq n \leq N/2-1$, with $M \times N$ denoting the image size (M, N: even numbers). G is the DFT of $g(\phi, \theta)$ and Re[•] and Im[•] signify the real and imaginary parts, respectively. Note that $P(m,n)$ contains only the shape information of the input object $f(x,y,z)$, i.e., position and orientation information is eliminated.

2.7. Scale-Invariant Property

In some cases, two objects with the same shape are different in their sizes. If size is not an important factor to discriminate an object from the others in a specific application, a scale-invariant signature is needed. The scale-invariant signature image $S(m,n)$, i.e., normalized absolute DFT coefficients, can be obtained by normalizing $P(m,n)$ with respect to the DC component:

$$S(m,n) = \frac{P(m,n)}{P(0,0)},$$

where $P(0,0)$ is the DC component of $g(\phi, \theta)$. Note that $S(m,n)$ is position-, orientation-, and scale-invariant with $S(m,n)$ retaining only the shape information. Figure 2(f) shows an example of the 2-D signature image $S(m,n)$.

2.8. 1-D Signature

Although $S(m,n)$ is an invariant feature of an input object $f(x,y,z)$, it is not a good descriptor of an object. Low data complexity is needed for efficient retrieval or recognition of 3-D objects. For this goal, a 1-D signature $Q(k)$ is defined. The weighted average of the 2-D signature image $S(m,n)$ is used to construct the 1-D signature $Q(k)$.

Figure 2(f) shows the blocking scheme used to average the signature values. The overlaid black lines on the 2-D signature image $S(m,n)$ denote the boundary of the blocks, in which 4-layered subdivision in each quadrant is shown. Quad-divisions to low-frequency area are performed iteratively until 1×1 block of the DC component of S is defined. The 1-D signature $Q(k)$ consists of the average values of the blocks, where the averaging blocks are defined only in the upper two quadrants ($n \leq 0$). By virtue of the even symmetry property of the 2-D signature $P(m,n)$, only half of the coefficients are required. Then, the proposed 1-D signature $Q(k)$ is used for recognition of 3-D convex objects.

In summary, the proposed method has three invariant properties of feature values. Position invariance is obtained by coordinate conversion whereas rotation and scale invariances are achieved by defining feature values using DFT coefficients and weighted average operations, respectively.

3. Experimental Results and Discussions

To show the effectiveness of the proposed 1-D signature representation $Q(k)$ for 3-D convex object recognition, three simple 3-D convex objects $f(x,y,z)$ are synthesized and their 1-D representations are obtained. Figures 3(a), 3(b), and 3(c) show the specifications of the test polyhedral objects in the Cartesian coordinates, where the number denotes the line length in voxel element unit. The voxel array used in representing 3-D objects has the dimension of 256×256×256, with the voxel value equal to 0 or 1 depending on whether the element is occupied by an object or not.

As explained in Figure 2, 2-D representations $S(m,n)$, which are defined based on the DFT coefficients, of 3-D test objects are obtained. Then, 1-D signatures $Q(k)$ are generated for each object, in which 44 blocks are defined. As explained in Figure 2(f), 7-layered subdivision of a 256×256 quadrant of $S(m,n)$ down to 1×1 gives 22 blocks and note that two quadrants ($n \leq 0$) are used. The average values of the 44 blocks formulate the 1-D signature $Q(k)$.

Invariant characteristics to rotation and scale can be tested by changing the rotation angle and object size [15]. Rotation- and scale-invariant characteristics can be used to define invariant feature values of some objects that have the same shape if it is assumed that two objects have the same shape with different size and object pose.

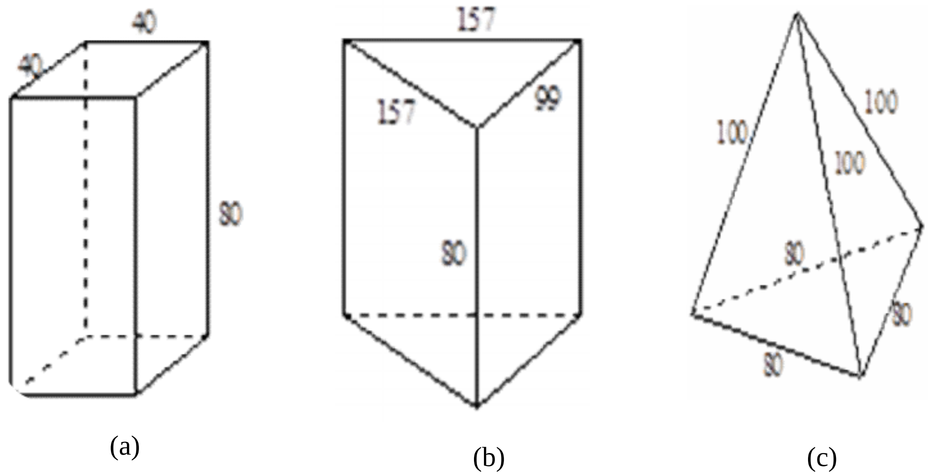

Figure 3. Synthesized 3-D convex objects. (a) object 1, (b) object 2, (c) object 3

To test the invariant and discriminating properties of the proposed 1-D signature $Q(k)$, Euclidean distances between the signature vectors of the synthesized objects at different positions, orientations, and scales are computed. Table 1 shows the average distance between the objects, which is used as a dissimilarity or distance measure. For the position, orientation, and scale parameters of an input object, arbitrary values are randomly generated for test objects. For the translation parameters, the values from –75 to 75 are used whereas for the scaling parameters, the values from 0.5 to 1.5 are used. The rotation parameters used have the range from 0 to 2π.

Object recognition is performed by comparing the distance defined based on the 1-D signatures $Q(k)$, in which the distance is used as a dissimilarity measure for object matching or recognition. As shown in Table 1, the distance (intra-distance) is small between the same objects (diagonal elements in Table 1) whereas the distance (inter-distance) is large between different objects (off-diagonal elements). The inter-distance of the pair objects 1 and 2 yield relatively small distance values. Other pair of objects produces high distance values. About 0.02-0.1 can be used as the threshold to distinguish whether a pair of objects is the same or not. Table 1 shows good invariant and discriminating properties of the proposed 1-D signature representation. The ratio of inter- to intra-distance is large enough (from 76 to 330), which is desirable for reliable object recognition.

Table 1. Intra- and inter-distances between object pairs.

Distance	Object 1	Object 2	Object 3
Object 1	0.0012	0.1305	0.2316
Object 2		0.0014	0.1072
Object 3			0.0007

4. CONCLUSIONS

For effective representation of 3-D convex objects, a 1-D signature invariant to position, orientation, and scale is proposed. The invariant and discriminating properties of the proposed 1-D representation are supported by experiments. The proposed 1-D signature representation can be applied to effective shape-based 3-D object retrieval and recognition. Future work will be the application of the proposed 1-D signature to more complex and real object dataset and the extension of the proposed algorithm to recognition of non-convex objects.

REFERENCES

[1] Pang, B., & Ma, H. (2011) "An efficient way of 3D model representation system in recognition system". In Proceedings of 2011 International Conference on Multimedia and Signal Processing, pp. 107-111.

[2] Khatun, A., Chai, W. Y., Iskandar, D. A., & Islam, M. R. (2011) "The effectiveness of ellipsoidal shape representation technique for 3D object recognition system". In Proceedings of 2011 7th International Conference on Information technology in Asia, pp. 1-6.

[3] Zhang, Q., Jia, J., & Li, H. (2010) "A GPU based 3D object retrieval approach using spatial shape information". In Proceedings of 2010 International Symposium on Multimedia, pp. 212-219.

[4] Milanese, R. & Cherbuliez, M. (1999) "A rotation, translation, and scale-invariant approach to content-based image retrieval", Journal of Visual Communication and Image Representation, Vol. 10, No. 2, pp. 186-196.

[5] Derrode, S. & Ghorbel, F. (2001) "Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description", Computer Vision and Image Understanding, Vol. 83, No. 1, pp. 57-78.

[6] Parvin, B, & Viswanathan, S. (1995) "Tracking of convex objects". In Proceedings of International Symposium on Computer Vision, pp. 295-298.

[7] Turnbull, C., & Cameron, S. (1998) "Computing distances between NURBS-defined convex objects". In Proceedings of 1998 IEEE International Conference on Robotics and Automation, Vol. 4, pp. 3685-3690.

[8] Chakraborty, N. Jufeng, P., Akella, S., & Mitchell, J. (2006) "Proximity queries between convex objects: An interior point approach for implicit surfaces". In Proceedings of 2006 IEEE International Conference on Robotics and Automation, pp. 1910-1916.

[9] Liu, H., Liu, W., & Latecki, L. J. (2010) "Convex shape decomposition". In Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition, pp. 97-104.

[10] Matsuo, H., & Iwata, A. (1994) "3-D object recognition using MEGI model from range data". In Proceedings of International Conference on Pattern Recognition, Vol. 1, pp. 843-846.

[11] Xu, J. Z., Suk, M., & Ranka, S. (1996) "Hierarchical EGI: A new method for object representation". In Proceedings of 3rd International Conference on Signal Processing, Vol. 2, pp. 926-929.

[12] Kang, S. B., & Ikeuchi, K. (1993) "The complex EGI: A new representation for 3-D pose determination", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 7, pp. 707-721.

[13] Hebert, M, Ikeuchi, K., & Delingette, H. (1995) "A spherical representation of recognition of free-form surfaces", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 7, pp. 681-689.

[14] Nguyen, H. V., & Porikli, F. (2011) "Concentric ring signature descriptor for 3D objects". In Proceedings of International Conference on Image Processing, pp. 2893-2896.

[15] Lee, H.-J., Sim, D.-G., & Park, R.-H. (2001) "A 2-D representation for 3-D convex object recognition". In Proceedings of FCV 2001: 7th Korea-Japan Joint Workshop Computer Vision, pp. 8-14.

[16] Wu, X., Liu, W., & Wang, T. (2003) "A new method on converting polygonal meshes to volumetric datasets". In Proceedings of 2003 IEEE International Conference on Robotics, Intelligent Systems and Signal Processing, DOI: 10.1109/CAD-CG.2005.86.

[17] Patil, S., & Ravi, B. (2005) "Voxel-based representation, display and thickness analysis of intricate shapes". In Proceedings of Ninth International Conference on Computer Aided Design and Computer Graphics (CAD/DG 2005), Vol. 1, pp. 116-120.

[18] Proakis, J. G., & Manolakis, D. K. (2007) *Digital Signal Processing*, 4th edition, Pearson.

Authors

Ho-Jin Lee received the B.S. and M.S. degrees in electronic engineering from Sogang University, Seoul, Korea, in 1997 and 1999, respectively. He is working toward the Ph.D. degree in electronic engineering at Sogang University. Currently, he is with Lightspeed Genomics Inc., Santa Clara, CA, USA, as Systems & Software Engineer. His current research interests are 3-D object recognition and face detection.

Rae-Hong Park was born in Seoul, Korea, in 1954. He received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1981 and 1984, respectively. In 1984, he joined the faculty of the Department of Electronic Engineering, School of Engineering, Sogang University, Seoul, Korea, where he is currently a Professor. In 1990, he spent his sabbatical year as a Visiting Associate Professor with the Computer Vision Laboratory, Center for Automation Research, University of Maryland at College Park. In 2001 and 2004, he spent sabbatical semesters at Digital Media Research and Development Center, Samsung Electronics Co., Ltd. (DTV image/video enhancement), Suwon, Korea. In 2012, he spent a sabbatical year in Digital Imaging Business (R&D Team) and Visual Display Business (R&D Office), Samsung Electronics Co., Ltd., Suwon, Korea. His current research interests are computer vision, pattern recognition, and video communication. He served as Editor for the Korea Institute of Telematics and Electronics (KITE) *Journal of Electronics Engineering* from 1995 to 1996. Dr. Park was the recipient of a 1990 Post-Doctoral Fellowship presented by the Korea Science and Engineering Foundation (KOSEF), the 1987 Academic Award presented by the KITE, and the 2000 Haedong Paper Award presented by the Institute of Electronics Engineers of Korea (IEEK), the 1997 First Sogang Academic Award, and the 1999 Professor Achievement Excellence Award presented by Sogang University. He is a co-recipient of the Best Student Paper Award of the IEEE International Symposium on Multimedia (ISM 2006) and IEEE International Symposium on Consumer Electronics (ISCE 2011).

Dong-Gyu Sim was born in Chungchung Province, Korea, in 1970. He received the B.S. and M.S. degrees in Electronic Engineering from Sogang University, Seoul, Korea, in 1993 and 1995, respectively. He also received Ph.D. degree at the same University in 1999. He was with the Hyundai Electronics Co., Ltd. from 1999 to 2000, where was involved in MPEG-7 standardization. He was a senior research engineer at Varo Vision Co., Ltd., working on MPEG-4 wireless applications from 2000 to 2002. He worked for the Image Computing Systems Lab. (ICSL) at the University of Washington as a senior research engineer from 2002 to 2005. He researched on the ultrasound image analysis and parametric video coding. He joined the Department of Computer Engineering at Kwangwoon University, Seoul, Korea, in 2005 as an Associate Professor. He was elevated to an IEEE Senior Member in 2004. His current research interests are image processing, computer vision, and video communication.