Three-dimensional object recognition using x-ray imaging

Sekwon Yeom, MEMBER SPIE
Bahram Javidi, FELLOW SPIE
University of Connecticut
Electrical and Computer Engineering
Department
371 Fairfield Road, Unit 1157
Storrs, CT 06269-1157
E-mail: Bahram@engr.uconn.edu

Young Jun Roh
Hyung Suck Cho
Korea Advanced Institute of Science and Technology
Department of Mechanical Engineering
373-1, Guseong-dong, Yuseong-gu
Daejeon, Korea, 305-701

Abstract. This paper presents a distortion-tolerant 3-D volume object recognition technique. Volumetric information on 3-D objects is reconstructed by x-ray imaging. We introduce 3-D feature extraction, volume matching, and statistical significance testing for the 3-D object recognition. The 3D Gabor-based wavelets extract salient features from 3-D volume objects and represent them in the 3-D spatial-frequency domain. Gabor coefficients constitute feature vectors that are invariant to translation, rotation, and distortion. Distortion-tolerant volume matching is performed by a modified 3-D dynamic link association (DLA). The DLA is composed of two stages: rigid motion of a 3-D graph, and elastic deformation of the graph. Our 3-D DLA presents a simple and straightforward solution for a 3-D volume matching task. Finally, significance testing decides the class of input objects in a statistical manner. Experiment and simulation results are presented for five classes of volume objects. We test three classes of synthetic data (pyramid, hemisphere, and cone) and two classes of experimental data (short screw and long screw). The recognition performance is analyzed in terms of the mean absolute error between references and input volume objects. We also confirm the robustness of the recognition algorithm by varying system parameters. © 2005 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1844532]

Subject terms: 3-D image processing; image reconstruction; image recognition; image classification; x-ray imaging systems.

Paper 040095 received Feb. 19, 2004; revised manuscript received Jul. 12, 2004; accepted for publication Aug. 11, 2004; published online Jan. 7, 2005.

1 Introduction

There has been much interest in object recognition in 3-D environments.1–19 In addition to conventional issues in 2-D object recognition, there have been new challenges facing us with 3-D information. One of the challenges concerning 3-D object recognition is the acquisition of the 3-D structure itself. Various reconstruction techniques have been derived to constitute a 3-D structure according to applications and environments. More accurate acquisition of 3-D information on the objects leads to more successful recognition. Another challenge regarding 3-D object recognition is that it generally places high demands on the computation and storage of data due to the huge amount of 3-D information. Despite those drawbacks of 3-D object recognition, a growing amount of research reveals potential advantages to studying 3-D space.

Holography and integral imaging (II) are popular 3-D visualization and recording techniques. Three-dimensional information is acquired at different imaging planes and perspective angles. We have utilized them for the 3-D object recognition.4–9 Advanced correlation filtering and recognition techniques have been applied to multiple views of 2-D images.10,11 Surface representation and matching techniques are approaches of considerable interest. Spin images are used for the surface matching in cluttered 3-D scenes.12

In Ref. 13, a 3-D face model is obtained after combining structure from motion (SfM) and a generic model. A 3-D frequency-domain representation is presented for pose-invariant face recognition in Ref. 14. Volume or cross-section images can be generated by various methods such as x-ray imaging, computed tomography (CT), digital tomosynthesis (DT), and magnetic resonance imaging (MRI). Volumetric information provides both the surface geometry and the inner structure of 3-D objects; we can overcome occlusion and surface reflection, which are common problems in conventional 2-D imaging. There have been numerous studies on the task of registration using volumetric information in the realm of medical imaging. In Ref. 15, two anatomical features are employed for brain registration. In Ref. 16, object localization and feature matching based on local geometric features are examined. A 3-D elastic volume matching technique using contour information is presented in Ref. 17. An affine transformation model using phase information from quadrature filters is adopted for volume registration in Ref. 18. Partial surface and volume matching techniques are presented and tested on various volume models in Ref. 19.

Object recognition is a comprehensive task of detecting and identifying objects in input scenes. The pose and distortion of the object are often estimated with object recognition. Numerous efforts have been applied to this task with 2-D and 3-D image data.20–24

In this paper, we present a distortion-tolerant method for 3-D volume object recognition. Volume information is reconstructed by an advanced x-ray imaging technique, called the uniform simultaneous algebraic reconstruction technique (USART). It was improved by employing spherical...
voxel elements for fast implementation and accurate estimation of voxel density.\textsuperscript{25,26}

The proposed object recognition system is composed of three stages: feature extraction, feature matching, and significance testing. For feature extraction, the conventional 2-D Gabor filtering\textsuperscript{27–29} is extended to 3-D space in order to analyze volume data. A Gabor feature is a multiresolution representation of object structure and energy in the spatial-frequency domain. Three-dimensional Gabor filtering extracts salient features according to 3-D location, spatial frequency, and bandwidth. We also achieve dimensionality reduction by sampling Gabor features on 3-D volume objects.

Dynamic link association (DLA) is a graph-matching technique for matching a reference and an unknown input object. Theoretically, the DLA scheme is tolerant to any distortion, rotation, or scaling of input objects. Since the main idea of the DLA was proposed in Ref. 30, many efforts have been made to apply the DLA to realistic problems.\textsuperscript{31–34} In Ref. 31, elastic graph matching (EGM) has been proposed to realize the DLA in a suboptimal way. In this paper, we extend the 2-D DLA technique to 3-D space and modify it in a simple and straightforward way. The modified DLA scheme is equipped with 3-D rotation tolerance and an efficient realization allowing distortion to some extent. Like the conventional model, the modified DLA scheme is composed of two stages: coarse matching and fine matching. However, during the coarse matching stage we search for the best-matched orientation as well as position of the 3-D rigid graph that is placed on the input scene. Rotation-invariant feature vectors are computed from selected Gabor jets. During the fine matching stage, nodes of the graph are elastically moved by searching for the best-matched positions. We also develop sequential and recursive realization for the fine matching stage. As the final step, we employ a statistical significance test to classify unknown input objects. It evaluates the statistical significance of each reference after the feature matching.

The combination of Gabor-based wavelets and DLA has been researched in 2-D space.\textsuperscript{31–33} However, the 3-D Gabor-based wavelets-on-volume data and the 3-D DLA technique with rotation tolerance have seldom been researched. Therefore, the main contributions of this paper can be summarized as follows:

1. We apply 3-D Gabor-based wavelets to volumetric information for feature extraction and dimensionality reduction.
2. We extend the 2-D DLA method to 3-D DLA with innovations, viz., robust rotation and distortion tolerance and efficient realization.
3. The proposed volume feature extraction-matching-recognition technique is not limited to presented volume data. We can apply this technique to any volumetric information.

In Sec. 2, we describe the framework of the recognition system. In Sec. 3, we briefly summarize the USART and demonstrate the reconstructed volume data. The 3-D
Gabor-based wavelets and the 3-D DLA are presented in Secs. 4 and 5, respectively. Statistical testing is explained in Sec. 6. In Sec. 7, experiments are reported. Results and performance analysis are found in Sec. 8. The conclusions follow in Sec. 9.

2 System Description

Figure 1 shows a block diagram for the 3-D volume recognition system. After 3-D volume reconstruction, an object-recognition task is performed via three subtasks: feature extraction, feature matching, and significance testing. At the first stage, we construct feature vectors using 3-D Gabor-based wavelets. The modified 3-D DLA is performed between reference objects and an input object. Finally, statistical testing decides the class of the unknown input object among multiple references.

In the following sections, we describe the various components of the 3-D volume reconstruction and recognition process, present experiment results, and analyze system performance.

3 Three-Dimensional Volume Reconstruction by X-Ray Imaging

In this paper, 3-D volume reconstruction by x-ray imaging is achieved by the USART.\cite{25,26} The USART estimates voxel density by combining several x-ray images, which are projected from different perspective angles. The USART has been improved by employing spherical voxel elements for fast implementation and accurate estimation.\cite{26}

The CT and DT build a layer of cross-section images from projected 2-D images.\cite{35} We may generate any arbitrary cross-section images and accumulate a series of them. However, inherent errors such as artifacts or blurring effects cause inaccurate estimation of 3-D volume information. The algebraic reconstruction technique (ART) method combined with the DT has proved to provide better performance.\cite{36} The ART is an iterative algorithm to compute voxel density using projected 2-D images. The USART is an advanced form of the ART. For the USART, we employ interpolated rays, which are optimally selected for each voxel, in order to avoid aliasing effects and achieve stable reconstruction performance. In this section, we briefly revisit the USART technique using spherical voxel elements and present several volume objects.

3.1 Overview of the USART

Figure 2 shows the x-ray imaging system.\cite{25,26} The system consists of a scanning x-ray source, a stage, and an x-ray digital imaging device; $X_s$, $s=1,...,S$, denotes the position that the x-ray source is electromagnetically scanned to; $I_s(u_x,v_y)$ is a grayscale image exposed on the x-ray imaging device corresponding to $X_s$; $u_x$ and $v_y$ are coordinates in the x and y directions from the reference point $P_x$. We image one stationary object from $S$ different views. Generally, the wavelength of the x-ray is about $10^{-2}$ to $10^{-6}$ μm. The x-ray intensity is reduced exponentially as it passes through materials. We can model the x-ray attenuation as

$$I_s=I_{s0} \exp \left( -\sum_{i=1}^{N} a_i^s t_i \right),$$

where $I_{s0}$ is initial intensity of the x-ray from source $s$; $a_i^s$ is the intersection length in element $i$ for the x-ray from source $s$; $t_i$ is the density of element $i$; and $N$ is the number of different elements that the x-ray passes through. The intersection length $a_i^s$ is a geometric parameter determined by the coordinates of the x-ray sources, the object, and the image plane.

A fast USART has been implemented by means of spherical voxels instead of conventional cubical voxels. Figure 3 shows the x-ray projection and reconstruction model of the spherical-voxel USART. The reconstruction process of the spherical-voxel USART is given by

$$\hat{f}_i(t+1)=\hat{f}_i(t)+\lambda_u \frac{1}{S} \sum_{i=1}^{S} \frac{1}{D^t_i} \left[ g_i^s(t) - h_i^s(t) \right],$$

where $\lambda_u$ is a density estimate of voxel $i$ after $t$ iterations, and $\lambda_u$ is a relaxation parameter that controls the convergence of the estimation. We assume an interpolated ray $\Omega^s_i$ is emitted from the x-ray source $s$ and passes through the center of voxel $i$; $g_i^s(t)$ is a measured value, and $h_i^s(t)$ is a modeled value of the $\Omega^s_i$ on the image plane; and $D^t_i$ is the total intersection length of the $\Omega^s_i$ in the whole reconstruction boundary. The projection of the $\Omega^s_i$ is modeled as

$$h_i^s(t)=d \sum_{p=1}^{P} \tilde{f}_p^{i_s}(t),$$

where $d$ is the diameter of spherical voxels; $\tilde{f}_p^{i_s}(t)$ is the density of the sphere $p$ on the $\Omega^s_i$ after $t$ iterations; $\tilde{f}_p^{i_s}(t)$ can be computed by the interpolation of density estimates of neighboring voxels of voxel $i$, and $P$ is the number of spheres on the $\Omega^s_i$ in the spherical reconstruction boundary.
It is noted that the computation of the geometric parameter $a_i$ is not required for the spherical-voxel USART. This results in great saving of computational time and memory.

### 3.2 Volume Objects Reconstructed by the Spherical-Voxel USART

We present five classes of volume objects reconstructed by the spherical-voxel USART: pyramid, hemisphere, cone, short screw (screw 1), and long screw (screw 2). X-ray images of the pyramid, hemisphere, and cone data are synthesized from their geometric models. For the two screws, experimental data were obtained. The geometric models of three classes of objects and real images of two types of screws are shown in Fig. 4.

Eight x-ray images are used for volume reconstruction, and the size of volume data is $60 \times 60 \times 60$ voxels. Figure 5 shows the outer surface and inner structure of the 3-D screw (screw 1) after 1, 10, 20, and 30 iterations of the

![Fig. 5 3-D volume object of screw 1 and sliced views at $x=30$, $y=30$, and $z=30$ after USART: (a) one iteration, (b) 10 iterations, (c) 20 iterations, (d) 30 iterations.](image-url)
USART reconstruction. Figure 6 visualizes other 3-D volume objects after 30 USART iterations. The voxel density is normalized so that the maximum value is 100. The outer surface represents the mean value of the voxel density in each object. As shown in Fig. 5, the artifact errors decrease as the USART iteration reconstruction process is continued. The reconstruction error is defined as the difference between the measured value $g_i^*(t)$ and the modeled value $h_i^*(t)$ of the x-ray projection. It was shown that the reconstruction errors approach steady state levels after 20 to 30 USART iterations.26

4 Three-Dimensional Gabor Filtering and Feature Vector Extraction

In this section, we discuss 3-D Gabor-based wavelets and present Gabor jets of volume objects. We employ 3-D Gabor wavelets for feature extraction and data reduction of volume objects.
4.1 Three-Dimensional Gabor-Based Wavelets

The 2-D Gabor filter acts as a bandpass filter with special selection of the passband according to the Gaussian envelope and the carrier frequency of the complex plane wave.\(^{27-29}\) The 2-D Gabor wavelets can be easily extended to three dimensions. The 3-D impulse response (or kernel) of the Gabor-based wavelets is

\[
g(x) = \frac{|k|^3}{\sigma^3} \exp\left(-\frac{|k|^2|x|^2}{2\sigma^2}\right) \left[\exp(\frac{\alpha x^2}{2}) - \exp\left(-\frac{\sigma^2}{2}\right)\right],
\]

where \(x\) is a position vector, \(k\) is a wave number vector, and \(\sigma\) is the standard deviation of the 3-D Gaussian envelope. The size of the Gaussian envelope is the same in the \(x\), \(y\), and \(z\) directions, and is proportional to \(\sqrt{2}\sigma/|k|\). The second term in the square brackets, \(\exp(-\sigma^2/2)\), subtracts the dc value so that one has zero mean response.\(^{29}\) The frequency response \(G(k')\) of \(g(x)\) is given by

\[
G(k') = (2\pi)^{3/2} \left[\exp\left(-\frac{\sigma^2}{2|k|^2}|k' - |k|^2|\right)
- \exp\left(-\frac{\sigma^2}{2|k|^2}(|k'|^2 + |k|^2)\right)\right].
\]

The sampling of \(k\) is done by \(k_{lmn} = k_{0l}[\sin \theta_2 \cos \phi_m \sin \theta_1 \sin \phi_m \cos \theta_1]\), where \(\theta_2 = [(l - 1)/L](\pi\phi_m = [(m - 1)/M](\pi)\), and \(k_{0l} = k_0/\delta^l\) for \(l = 1, \ldots, L, m = 1, \ldots, M\), and \(n = 1, \ldots, N\); here \(k_{0l}\) is the magnitude of the wave number vector; \(\phi_m\) is the azimuth angle; \(\theta_1\) is the elevation; \(\delta\) is the spacing factor in the frequency domain; \(l, m,\) and \(n\) are the indices of the Gabor kernels; \(L, M,\) and \(N\) are the total numbers of decompositions along two tangential axes and a radial axis, respectively; and \(t\) denotes a matrix transpose. The carrier frequency of the bandpass filter is determined by \(k\). The Gabor-based wavelets are sensitive to the direction of edges. The response is strong if the direction of \(k\) is perpendicular to an edge.

Figure 7 shows the intensity of an impulse response for the 3-D Gabor wavelets. Selected frequency responses for the 3-D Gabor wavelets are shown in Fig. 8. The outer surface of volume is fixed at \(|x| = \sqrt{2}\sigma/|k_{lmn}|\) in Fig. 7 and \(|k' - k_{lmn}| = \sqrt{2}|k'_{lmn}|/\sigma\) in Fig. 8.

4.2 Feature Vector Extraction

We can define \(g_{lmn}(x)\) by sampling \(k\) as \(k_{lmn}\). Let \(h_{lmn}(x)\) be the output of the filtered input \(V(x)\) after convolution with \(g_{lmn}(x)\):

\[
h_{lmn}(x,y,z) = \sum_{x' = 1}^{L_x} \sum_{y' = 1}^{L_y} \sum_{z' = 1}^{L_z} g_{lmn}(x-x',y-y',z-z')
\times V(x',y',z'),
\]

where \(L_x, L_y,\) and \(L_z\) are the sizes of the volume data in the \(x\), \(y\), and \(z\) directions, respectively; \(h_{lmn}(x)\) is also called the \textit{Gabor coefficient}, and the magnitude of the Gabor coefficient is called the \textit{Gabor jet}. One Gabor jet vector is composed of a set of the Gabor jets: \(v(x) = [h_{lmn}(x)]; l = 1, \ldots, L, m = 1, \ldots, M, n = 1, \ldots, N\). Figure 9 shows selected Gabor jets of the reconstructed pyramid volume in Fig. 6(a). The parameters are set at \(\sigma = \pi, k_0 = \pi/2, \delta = 2, L = 4, M = 4, N = 3\).

5 3-D Modified Dynamic Link Association

In this section, we extend the 2-D DLA technique to 3-D space for comparison of two volume objects. Our proposed 3-D DLA has two novel properties: (1) a robust volumematching technique, which is tolerant to translation, rotation, and distortion in 3-D space, and (2) sequential and recursive realization for the fine matching. The proposed system is composed of two stages: \textit{rigid graph matching} (RGM) and \textit{elastic graph matching} (EGM). The EGM is often taken in the literature to mean the entire suboptimal
system for the DLA. In this paper, we use “EGM” to refer only to the fine matching stage, and we adopt another term, “RGM,” for the coarse matching stage.

5.1 Rigid Graph Matching with Rotation Tolerance
We employ regular hexahedron graphs as shown in Fig. 10(a); however, any arbitrary graph can be used for the 3D DLA technique. Let \( R \) and \( S \) be two identical and rigid 3-D graphs placed on the reference and unknown input data, respectively. During the RGM, we search for the best-matched position and orientation of the graph \( S \) with respect to the graph \( R \).

We can describe any rigid motion of the graph \( S \) by a translation vector and a rotation matrix. Let \( \mathbf{p} \) be a 3-D translation vector \( \mathbf{p} = [p_x, p_y, p_z]^T \), and \( \mathbf{e} \) be a vector of three Euler angles \( \mathbf{e} = [\phi, \theta, \psi]^T \). Any rigid motion of the graph \( S \) can be modeled as

![Fig. 9 Gabor jets of the pyramid and sliced views at x=30, y=30, and z=30 of the pyramid after 30 USART iterations [Fig. 6(a)] when (a) n=1, m=1, l=1; (b) n=2, m=2, l=2; (c) n=3, m=3, l=3.](image-url)
\[ x'_i = A(e)(x'_i - x''_i) + p, \quad i = 1, \ldots, K, \]  

where \( x'_i \) and \( x''_i \) are the positions of node \( i \) in the graphs \( R \) and \( S \), respectively; \( x''_i \) is a node position at the center of the graph \( R \); \( K \) is the total number of nodes in the graph; and \( A \) is a rotation matrix, which is determined by \( e \). Figure 11 shows the rotation theorem for the Euler angles and corresponding rotation matrices. Any 3-D rotation can be defined by a general rotation matrix \( A = BCD \). The first rotation \( (D) \) is by an angle \( \varphi \) about the \( z \) axis, the second one \( (C) \) is by an angle \( \theta \in [0, \pi] \) about the \( x \) axis, and the third one \( (B) \) is by an angle \( \psi \) about the \( z \) axis again. All the
rotations are counterclockwise. It is noted that $\mathbf{p}$ corresponds to the position vector of the central node in the graph $S$.

We search for the best-matched position and orientation of the graph $S$ by maximizing a cost function $C_{\text{RGM}}$:

$$C_{\text{RGM}} = \sum_{i=1}^{K} c_i(x_i', x_i),$$  

(10)

where $\hat{\mathbf{p}}$ and $\hat{\mathbf{e}}$ are the estimates of the location and orientation of the graph $S$. A node cost $c_i$ is the cross-correlation coefficient of two function vectors:

$$\{\hat{\mathbf{p}}, \hat{\mathbf{e}}\} = \arg \max_{\mathbf{p}, \mathbf{e}} C_{\text{RGM}}.$$  

(9)
Euler angles in the 3-D real domain. In the experiment, we possible voxel displacements in the 3-D integer domain and angle \( D \) angles is likewise restricted by a predetermined search volume data in the \( x \) invariant feature vector.

An arbitrary rotation can convert the integer image data. An arbitrary rotation can convert the integer \( L \) and \( D \) only defined in the integer domain, which is natural for putational load.

Another consideration is that the Gabor jet vectors are \( L \) and \( D \), respectivey, \( \tan 2 \) \( 1/2 \) for the search interval \( d \) for all the saved nodes \( i \). In the experiments, we set \( D \) and \( L \), to arbitrary values with consideration of the computational load.

Another consideration is that the Gabor jet vectors are only defined in the integer domain, which is natural for image data. An arbitrary rotation can convert the integer vector to real noninteger values. In that case, we simply apply the 3-D nearest neighbor interpolation method.

### 5.2 Elastic Graph Matching with Sequential and Recursive Realization

The graph \( S \) has elastic property during the EGM. We change nodes’ positions independently by maximizing a cost function \( C_{\text{EGM}} \):

\[
\{ \hat{x}_1^*, \ldots, \hat{x}_K^* \} = \arg \max_{\tilde{x}_1^*, \ldots, \tilde{x}_K^*} C_{\text{EGM}}(\tilde{x}_1^*, \ldots, \tilde{x}_K^*; \hat{p}, \hat{e}).
\]

---

**Table 1** Procedures for the EGM.

<table>
<thead>
<tr>
<th>Step</th>
<th>Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Set ( 0 &lt; d &lt; d_e ) (e.g., ( d = d_e/2 ), where ( d_e ) is edge size) Set maximum iteration number ( I_{\text{max}} ) Let ( i = 0 )</td>
</tr>
<tr>
<td>2</td>
<td>Let ( i = i+1 ) Let node index ( i = 0 ) Compute ( C_0 = C_{\text{EGM}}(x_1^<em>, \ldots, x_K^</em>; \hat{p}, \hat{e}) )</td>
</tr>
<tr>
<td>3</td>
<td>Let ( i = i+1 ) and ( x_i^* = x_i^* + d ) Let ( i = i+1 ) and ( x_i^* = x_i^* + d ) Recompute the EGM cost: ( C_i = C_{\text{EGM}}(\tilde{x}_1^<em>, \ldots, \tilde{x}_K^</em>; \hat{p}, \hat{e}) ) Record ( i ) and ( x_i^* ) if ( C_i &gt; C_0 ) Go to step 3 until ( i = K )</td>
</tr>
<tr>
<td>4</td>
<td>Change ( x_i^* = x_i^* + d ) for all the saved nodes ( i )</td>
</tr>
<tr>
<td>5</td>
<td>Terminate if there are no recorded nodes in step 4 or ( I = I_{\text{max}} ) Otherwise, go to step 2</td>
</tr>
</tbody>
</table>
\[ C_{E_{\text{GM}}} = \sum_{i=1}^{K} c_i (\mathbf{x}_i', \mathbf{x}_i' ; \hat{\mathbf{p}}, \hat{\mathbf{e}}) - \lambda \sum_{(i,j) \in E^R, E^T} \| \Delta_{ij}' - \Delta_{ij} \|^2 \]

\[ = C_{E_{\text{GM}}} - \lambda \sum_{(i,j) \in E^R, E^T} \| \Delta_{ij}' - \Delta_{ij} \|^2, \]  

where \( \mathbf{x}_i' \) is the position vector of node \( i \) in the graph \( R \); \( \mathbf{x}_i' \) is the position vector of node \( i \) in the graph \( S \); \( K \) is the total number of nodes in the graph; and \( \hat{\mathbf{p}} \) and \( \hat{\mathbf{e}} \) are the estimates of position and orientation during the RGM, respectively. We define \( \Delta_{ij}' = \mathbf{x}_i' - \mathbf{x}_i' \) and \( \Delta_{ij}' = \mathbf{x}_i' - \mathbf{x}_i' \); \( (i,j) \) indicates an edge from node \( i \) to node \( j \); node \( j \) can be one of the six nearest neighbors of the node \( i \) in the regular hexahedron graph. Let \( E^R \) and \( E^S \) be sets of node pairs in the graph \( R \) and the graph \( S \), respectively, composed of one-to-one corresponding nodes and edges. Note that we have already...
Table 2 The overall performance of the distortion-tolerant object recognition.

<table>
<thead>
<tr>
<th>Test set</th>
<th>Shape</th>
<th>$N_D$</th>
<th>$N_D$</th>
<th>$N_F$</th>
<th>$P_D$ (%)</th>
<th>$P_F$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pyramid</td>
<td>16</td>
<td>16</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Hemisphere</td>
<td>16</td>
<td>15</td>
<td>1</td>
<td>93.33</td>
<td>6.67</td>
</tr>
<tr>
<td>3</td>
<td>Cone</td>
<td>16</td>
<td>14</td>
<td>2</td>
<td>86.67</td>
<td>13.33</td>
</tr>
<tr>
<td>4</td>
<td>Screw 1</td>
<td>16</td>
<td>15</td>
<td>0</td>
<td>93.33</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Screw 2</td>
<td>16</td>
<td>15</td>
<td>1</td>
<td>93.33</td>
<td>6.67</td>
</tr>
</tbody>
</table>

Table 3 Five rotation angle sets for rotated objects.

<table>
<thead>
<tr>
<th>Set</th>
<th>Rotation angle (deg)</th>
<th>$\psi$</th>
<th>$\theta$</th>
<th>$\psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>$-30$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>$-45$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>$-60$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
estimated $\hat{p}$ and $\hat{e}$ during the RGM. The initial location and orientation of the graph $S$ for the EGM are computed by $\hat{p}$ and $\hat{e}$. During the EGM, we relocate all nodes' positions in the elastic graph $S$.

The parameter $\lambda$ controls the flexibility of deformation in the graph $S$. During the RGM, the rigidity of the graph implies infinite penalty for any deformation of the graph, but during the EGM we reshape the graph with less constraint. The first part of $C_{\text{EGM}}$ is the same as $C_{\text{RGM}}$. The larger the value of $\lambda$, the higher the deformation penalty of the graph $S$. If $\lambda$ is infinite, $C_{\text{EGM}}$ can be maximized when $\Delta_{ij}^r = \Delta_{ij}^s$ for all $i$ and $j$.

We develop a sequential and recursive method to imple-
ment the EGM in a fast and effective manner. Table 1 demonstrates the overall procedures of the EGM. All steps are presented in one positive dimension; however, they can be easily extended to 2-D and 3-D space. We simplify the EGM by assuming all Euler angle estimates are equal to zero during the RGM. Let $x_i$ be the initial position of node $i$ for the EGM. We assume that there exists a global maximum in Eq. (13) and it is located in the small region around the initial value for the EGM.

At the first step, we set $d$ and $l_{\text{max}}$, where $d$ is a fixed displacement for nodes in one dimension and $l_{\text{max}}$ is a maximum iteration number. We then compute the initial

---

Fig. 14 Results of the experiment in Sec. 8.2. The reference is a hemisphere after 30 USART iterations without rotation, and the input is screw 1 after 30 USART iterations with rotation. The input object is rotated with rotation angle set 1 ($\varphi=30$ deg, $\theta=0$, $\psi=0$). (a) Reference object, (b) input object at the initial state of RGM, (c) input object after RGM, (d) input object after EGM. (a) to (d) each show an object from four different perspectives.
At step 3, we recompute $C_{\text{EGM}}$ with a new node position $x_i^+$ and save the node’s index and position if the new position provides a larger $C_{\text{EGM}}$. This task is performed sequentially for all nodes in the graph $S$. At step 4, we place the recorded nodes in new positions. Finally, we terminate the procedures according to a termination criterion or iterate steps 2 through 5. In this paper, “sequential” procedures refer to step 3, and “iteration” refers to procedures from step 2 to 5.

We described only positive displacement ($d>0$) in one dimension. It is noted that $x_i^+ = x_i^{-} - d$ should be considered for negative displacement. Thus, in 3-D space, one position vector has six possible transitions, which are $x_i^+ = x_i^- + d$, $y_i^+ = y_i^- + d$, and $z_i^+ = z_i^- + d$.

**Fig. 15** Results of the experiment in Sec. 8.2. The reference is a hemisphere after 30 USART iterations without rotation as shown in Fig. 14(a), and the input is a hemisphere after 30 USART iterations with rotation. The input object is rotated with rotation angle set 5 ($\varphi=30$ deg, $\theta=-60$ deg, $\psi=0$). (a) Input object at the initial stage of RGM, (b) input object after RGM, (c) input object after EGM. (a) to (c) each show an object from four different perspectives.
Recursively, we repeat the whole procedure of Table 1 while reducing $d$ gradually. We use the term “recursion” to represent the repetition of the whole procedure in Table 1 while reducing $d$, and “iteration” to represent procedures from the steps 2 through step 5 with a fixed $d$. Different EGM recursions are demonstrated in Sec. 8.

6 Statistical Significance Testing

At the final stage, we decide the class of input objects by statistical significance testing as used in Refs. 31, 32. Let $C_{sr}$ denote $C_{EGM}$ as computed with a reference object $r_i$ and an unknown input object $s$. We order $C_{sr}$ in descend-
where \( \mu \) is the mean and \( \sigma \) is the standard deviation of the set \( \{ C_{sr_i} \}_{i=1}^{N_r-1} \), and the thresholds \( \tau_1 \) and \( \tau_2 \) are determined heuristically in the experiment.

We also use two parameters for the performance evaluation:

\[
P_D = \frac{\text{number of correct decisions (} N_D \text{)} }{\text{total number of input objects (} N_O \text{)}} \times 100 \ (\%),
\]

\[
P_F = \frac{\text{number of wrong decisions (} N_F \text{)} }{\text{total number of input objects (} N_O \text{)}} \times 100 \ (\%),
\]

where \( P_D \) is the correct decision rate, \( N_O \) is the number of input data tested, \( N_D \) is the number of correct decisions accepted by the statistical significance test, \( P_F \) is the false alarm probability, and \( N_F \) is the number of wrong decisions (the reference is not in the same class as the input object) that are accepted.

**7 Experiments**

We present two experiments on the 3-D object recognition task: one involves the distortion of input objects, and the other their rotation. The former is an experiment based on input objects that are reconstructed at all USART iterations. The latter is based on input objects that are reconstructed from the rotated objects at the 30th USART iteration.

The design parameters of the 3-D Gabor-based wavelets are the same throughout this paper \( (\sigma=\pi, \ k_0=\pi/2, \ \delta=2, \ L=4, \ M=3, \ N=4) \). Therefore, one Gabor jet vector at one node is composed of 48 Gabor jets, and the dimension of the feature vector is 4 in Eq. (12). The 3-D graph of a 7\( \times \)7\( \times \)7 grid is used for the RGM, and that of a 5\( \times \)5\( \times \)5 grid for the EGM; the edge size \( d_e \) is set at 8 voxels. We chose the thresholds \( \tau_1 = 0.1 \) and \( \tau_2 = 0.8 \) for the first experiment and \( \tau_1 = 0.05 \) and \( \tau_2 = 0.7 \) for the second experiment. The choice was made heuristically during the experiments. The control parameter \( \lambda \) was set at \( 10^{-5} \); \( d \) was set at 4 voxels for the first EGM recursion, 2 for the second, and 1 for the third.

**8 Result and Performance Analysis**

We experimented with five classes of volume objects: the pyramid, the hemisphere, the cone, and screws 1 and 2. The performance is analyzed in terms of the mean absolute error (MAE), and experiments with different EGM recursions and the control parameter \( \lambda \) are also described.

**8.1 Distortion-Tolerant Object Recognition**

In the first experiment, we perform five different tests according to five different classes of input object sets. Each test has five classes of references and 16 input volume data. For the reference objects, we choose the reconstructed volume objects after 30 iterations of the USART. The input data are composed of volume data after the reconstruction of an odd number of iterations and the reference itself, that is, volume objects after 1, 3,...,29 and 30 USART iterations.

The center of the graph \( R \) is placed at a fixed position \([30\ 30\ 030]'\). The center of the graph \( S \) is initially placed at \([15\ 15\ 15]'\). The search interval \( \Delta_p \) is 5 voxels for all three coordinates in 3-D space.

Figures 10 and 12 show examples of distortion-tolerant object recognition. The RGM process turns out to be a robust detection and aligning process for distorted input objects. We have a finer matching process during the EGM. In Fig. 10, the reference is a pyramid after 30 USART iterations, and the input is a pyramid at the first iteration. As shown in Figs. 10(c) to 10(h), the input object is severely distorted because of artifact errors. However, we successfully recognize the input object to be in the same class as the reference object. Figure 12 shows the classification result for reference and input objects in different classes. The reference is screw 1 after 30 USART iterations, and the input is screw 2 after 30 USART iterations. They are proved to be of different classes. Table 2 shows the overall result of the first experiment.

For the input data set of screw 1, \( N_D \) is 15; but \( N_F \) is zero because the reference of the maximum \( C_{EGM} \) is rejected by the statistical significance test.

**8.2 Rotation-Tolerant Object Recognition**

In the second experiment, we test rotation-tolerant object recognition. We perform five different tests according to the five different classes of input object sets. Each test has five references and five input volume objects. The references are the same as in the previous experiment in Sec. 8.1. The input data are composed of the reconstructed volumes from rotated 3-D objects. All input objects are reconstructed with 30 USART iterations. Five different angle sets for rotated input objects are shown in Table 3. For screws 1 and 2, we rotate reconstructed volume objects computationally using the 3-D cubic interpolation according to each rotation angle set.

The graph \( R \) and the initial graph \( S \) are placed at the same locations as in the previous experiment. We search for the best-matched orientation angle as well as the best-matched location during the RGM. We set the search angle \( \Delta_x \) for the rotation at 15 deg for all three Euler angles.

<table>
<thead>
<tr>
<th>Test set</th>
<th>Shape</th>
<th>( N_O )</th>
<th>( N_D )</th>
<th>( N_F )</th>
<th>( P_D ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pyramid</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>Hemisphere</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>80</td>
</tr>
<tr>
<td>3</td>
<td>Cone</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>Screw 1</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>Screw 2</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>
Figures 13 to 16 show examples of rotation-tolerant object recognition. In Fig. 13, the reference is a cone after 30 USART iterations. The input object is a cone rotated according to rotation angle set \( \gamma=30 \) deg, \( \beta=-60 \) deg, \( \psi=0 \) and reconstructed at the 30th iteration. The system successfully classifies the input object with correct angle estimates. In Fig. 14, the reference is a hemisphere after 30 USART iterations and the input object is screw 1 after 30 USART iterations with the rotation \( \phi=30 \) deg, \( \theta=0 \), \( \psi=0 \). The objects prove to be of different classes. In Fig. 15, the reference is the hemisphere after 30 USART iterations as shown in Fig. 14(a). The input object is also a hemisphere after 30 USART iterations with rotation angle set \( \phi=30 \) deg, \( \theta=-60 \) deg, \( \psi=0 \). As shown in Fig. 15, we estimate...
wrong Euler angles \( \hat{\mathbf{e}} = [30\, \text{deg}, -75\, \text{deg}, 0\, \text{deg}] \) but the correct classification is made during the significance testing. Figure 16 shows the result when the reference is screw 1 after 30 USART iterations as shown in Figs. 12(a) and 12(b), and the input object is screw 1 after 30 USART iterations with the rotation \( \varphi = 30\, \text{deg}, \theta = -60\, \text{deg}, \psi = 0\). They are proved to be of the same classes with correct Euler angle estimates.

Table 4 shows the overall performance of the rotation-tolerant object recognition. The recognition is performed successfully for most of the input data. All estimated angles are correct except for rotation angle sets 3 to 5 of the hemisphere.

### 8.3 Performance Analysis

There are many factors affecting the performance of the 3-D volume object recognition. We analyze the performance according to the similarity between references and
input objects. The analysis is only concentrated on distortion-tolerant object recognition. The MAE is employed for our similarity measure. It is a matching criterion often used for motion estimation in image compression.\textsuperscript{37}

We define the MAE between the reference and the input volume with the location estimate $\hat{p} = [\hat{p}_x, \hat{p}_y, \hat{p}_z]$:

$$\text{MAE} = \frac{1}{L_x L_y L_z} \sum_{x=1}^{L_x} \sum_{y=1}^{L_y} \sum_{z=1}^{L_z} |V_R(x - \hat{p}_x, y - \hat{p}_y, z - \hat{p}_z) - V_s(x, y, z)|,$$  \hspace{1cm} (17)

where $V_R$ is the reference and $V_s$ is the input volume; $L_x$, $L_y$, and $L_z$ are the dimensions of the volumes.

Fig. 19 Performance analyses of different EGM recursions; the input is the cone after 1, 3, ..., 29 and 30 USART iterations; the reference image is composed of five objects after 30 USART iterations. The horizontal axis shows the USART iteration number of the input object; Fig. 14(b) corresponds to the EMG with one recursion. (a) EGM recursion is 2; (b) EGM recursion is 3.
$L_x$, $L_y$, and $L_z$ are the sizes of the volume data in $x$, $y$, and $z$ directions, respectively.

Figure 17(a) shows the MAE when the class of input objects is the cone. The MAE is computed for five reference objects. Figure 17(b) shows $C_{EGM}$. The $x$ axis shows the number of USART iterations of input objects. Figures 18(a) and 18(b) show MAE and $C_{EGM}$ when the input class is screw 1. As shown in Figs. 17 and 18, the MAE is smaller when the reference is identified as being of the same class as the input object. It also decreases as the number of the USART iterations increases. When the MAE is similar among different reference objects, the recognition process has difficulty obtaining correct results. Also, when the MAE is too large (i.e., the similarity is too low between the input and the true reference), the recognition can fail, yielding the highest $C_{EGM}$ for wrong reference object.

![Figure 17(a)](image1.png)

![Figure 17(b)](image2.png)

**Fig. 20** Performance analysis of the control parameter $\lambda$; the input is the cone after 1, 3, ..., 29 and 30 USART iterations; the USART reference image is composed of five objects after 30 USART iterations. The horizontal axis shows the USART iteration number of the input object; Fig. 14(b) corresponds to $\lambda = 10^{-5}$. (a) $\lambda = 10^{-4}$, (b) $\lambda = 10^{-6}$. 
We investigate the effects of the recursive EGM process. There was one EGM recursion (d = 4) in Fig. 17(b). Figure 19(a) shows the results of two EGM recursions (d = 4 and 2). Figure 19(b) shows the results of three EGM recursions (d = 4, 2, and 1). Figures 17(b), 19(a), and 19(b) show that C_{EGM} becomes larger with more recursions although the overall shapes of C_{EGM} are similar.

Figures 20(a) and 20(b) show the effect of different λ's. Figure 17(b) shows the results for \( \lambda = 10^{-5} \). Figures 20(a) and 20(b) are the results for \( \lambda = 10^{-4} \) and \( \lambda = 10^{-6} \), respectively. Large λ implies a large penalty for the graph deformation. A large λ may provide better results with less graph deformation, but the total cost is proven to become smaller, as shown in Fig. 20(a).

9 Conclusions and Future Work

In this section, we have presented 3-D distortion-tolerant volume recognition using a 3-D modified DLA technique based on 3-D Gabor feature vectors. Rotation-invariant features are extracted and constructed by the 3-D Gabor wavelets. The wavelets extract localized features of objects according to 3-D spatial frequency and bandwidth, as well as location. The modified 3-D DLA proves to be a reliable recognition technique, which is tolerant to rotation and distortion. The performance is analyzed in terms of the MAE. The effects of different EGM recursions and the control parameter λ are also presented.

The scope of applications of presented recognition technique is very broad. It can be applied to any 3-D volumetric information for alignment, registration, classification, and identification task.

Some tasks are left for future work. The consideration of noninteger positioning of nodes would be desirable for certain applications such as rotated nonrigid objects or facial expressions. We can utilize advanced statistical classification methods such as linear discriminant analysis with a pool of training data. More testing with various data and different parameters would be helpful for constructing more powerful recognition systems.

References

Sekwon Yeom received his BS degree in physics from Inha University, Inchon, Korea, in 1995. He received his BS and MS degrees from the Department of Electronics and Computer Engineering, Korea University, Seoul, Korea, in 1997 and 1999, respectively. He also received his MS degree from the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, in 2003. He is currently a PhD student at the same university. His research interests are target tracking and optical information processing.

Bahram Javidi is Board of Trustees Distinguished Professor at the University of Connecticut, the highest honor given to faculty members at the University. He has received numerous other awards from UConn, including the Provost Research Excellence Award, the Alumni Association Excellence in Research Award, the School of Engineering Distinguished Professor Award, and the Department of Electrical and Computer Engineering Outstanding Faculty Research Award. Dr. Javidi is a fellow of the Institute of Electrical and Electronics Engineers (IEEE), a fellow of the Optical Society of America (OSA), and a fellow of the International Society for Optical Engineering (SPIE). In 1990, the National Science Foundation named him a Presidential Young Investigator. He received the IEEE Lasers and Electro-optics Society Distinguished Lecturer Award in 2003 and 2004. He was the recipient of the IEEE Best Journal Paper Award from IEEE Transactions on Vehicular Technology in 2002. Dr. Javidi has published several books, including Optical and Digital Techniques for Information Security, Springer-Verlag, 2004; Image Recognition: Algorithms, Systems, and Applications, Marcel Dekker, New York, 2002; Three Dimensional Television, Video, and Display Technologies, Springer-Verlag, 2002; Smart Imaging Systems, SPIE Press, 2001; Real-time Optical Information Processing, Academic Press, 1994, and Optical Pattern Recognition, SPIE Press, 1994. Dr. Javidi has published more than 170 technical articles in major journals. He has made more than 220 conference contributions, including more than 70 invited conference papers, and 60 invited presentations. His papers have been cited more than 2200 times according to the Citation Index of WEB of Science. His papers have appeared in Physics Today and Nature, and his research has been cited in the National Academy of Engineering’s Frontiers in Engineering Newsletter, the IEEE Spectrum, Science, New Scientist, and National Science Foundation publications. Dr. Javidi is currently the editor in chief of the Springer-Verlag series on Advanced Science and Technologies for Security Applications.

Young Jun Roh received his PhD degree from the Mechanical Engineering Department at Korea Advanced Institute of Science and Technology. He received his MS degree in the same institute, and his BS in mechanical engineering at Yonsei University, Seoul, Korea. His research interests include visual inspection, pattern recognition, and 3-D x-ray imaging systems. Since 2003, he has been developing vision system for Mirae Corporation, Korea.

Hyung Suck Cho received his BS degree from Seoul National University, Korea, in 1971, his MS degree from Northwestern University, Evanston, IL, USA, in 1973, and his PhD degree from the University of California at Berkeley, CA, USA, in 1977. After that, he was a postdoctoral fellow in the Department of Mechanical Engineering, University of California at Berkeley, he joined Korea Advanced Institute of Science and Technology (KAIST) in 1978. After that, he was a professor in the Department of Production Engineering and the Department of Design and Automation. Since 1995, he has been in the Department of Mechanical Engineering at the same institute. From 1984 to 1985, he was a visiting professor at the Institute für Produktionstechnik und Automatisierung (IPA), Germany, where he carried out research on robot-based assembly. Dr. Cho’s research interests are focused on optomechatronics. In particular, his research highlights environment perception and recognition for mobile robots, machine vision and pattern classification, and application of artificial intelligence and machine intelligence. He has edited one handbook and three conference proceedings, and has published 10 book chapters and more than 425 research papers (354 papers in international journals and conferences, and 71 papers in Korean journals). Now, he serves on the editorial boards of six international journals: IEEE Transactions on Industrial Electronics, Journal of Robotic Systems, Robotics, Control Engineering Practice (IFAC), Journal of Advanced Robotics, and Journal of Engineering Manufacture (PIME).