Multi-resolution recognition of 3D objects based on visual resolution limits

Huimin Ma *, Tiantian Huang, Yanzhi Wang

Department of Electrical Engineering, 3D Image Simulation Laboratory, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 11 March 2008
Received in revised form 23 May 2009
Available online 3 September 2009

Communicated by J.A. Robinson

Keywords:
3D object recognition
Visual resolution limit (VRL)
Mesh simplification
Multi-resolution aspect graph

ABSTRACT

This paper presents a multi-resolution recognition method for 3D objects, based on the human visual model. In the first part of this paper, we propose a new visual resolution limit (VRL) calculation method that considers lens size, the scale of imaging cells and the distance, orientation and velocity of the object. In addition, we simplify 3D models with a novel mesh simplification method based on edge collapse, which controls the simplification degree with VRL. After applying viewpoint space partitioning to the simplified models at different resolutions, we develop a multi-resolution aspect graph library indexed by observation distance. Finally, we propose a 3D object recognition method based on multi-resolution aspect graphs and implement a real-time gradual multi-resolution recognition system that imitates human vision. We design and execute a set of experiments based on plane, car and ship models. Our results demonstrate that our recognition method is effective.

1. Introduction

Les and Les (2008) investigated visual recognition and understanding in non-visual understanding problems. One specific kind of visual recognition, 3D object recognition (Luo and Scherp, 1990; Nian et al., 2005; Kobayashi et al., 2000), is now widely used in the areas of robotics, automatic navigation and medical image analysis. It is becoming more popular to introduce human visual mechanisms and characteristics into 3D recognition systems.

There are two main 3D object recognition methods: one is based on the matching result between 3D original images and 3D library models, and the other is based on matching results between 2D projective images and 3D library models. Considering the high acquisition cost of real-time 3D metrical data, the second method is more practical. In this paper, we focus on the second 3D object recognition method.

There is some existing research on multi-resolution recognition of 3D objects. Without proposing any specific recognition method, Eggert et al. (1993) originated multi-resolution aspect graph representations. Zhang et al. (2006) presented a multi-resolution recognition method for 3D moving targets using sub-sampling, distortion, and fuzzy operations on a set of original aspect graphs to simulate the target’s motion.

Most multi-resolution representation methods, such as the ones we just mentioned[e1], are based on 2D projective images. In addition, most contemporary 3D recognition methods (Cyr and Kimia, 2001) based on aspect graphs are hard to apply due to the ambiguity of the scale factor. In this paper, we introduce a new multi-resolution representation method based on original 3D models to express morphologic features under different observation conditions more clearly[e2]. In addition, we present a scale factor related to actual distance and based on the visual resolution limit (VRL). We propose a new VRL calculation method that considers object distance, orientation and velocity. This calculation method lays the foundation for a practical 3D object recognition system.

The paper is organized as follows. Section 2 describes the structure of our 3D recognition system. Section 3 explains how the visual resolution limit controls 3D model simplification. Section 4[e3] presents the specific multi-resolution recognition methods used in our system. Section 5 gives experimental results. Finally, Section 6 concludes.

2. Structure of the 3D recognition system

The recognition system illustrated in Fig. 1 has two main parts: 3D model library foundation and target object recognition. The 3D model library simplifies the original 3D model (Hoppe et al., 1993; Ronfard and Rossignac, 1996) according to the VRL. Then the library applies viewpoint space partitioning (Luo et al., 2006) to the simplified models with different resolution levels. Then the library could do additional clustering on the partitioned results to obtain multi-resolution aspect graph representations of the 3D model (Cyr and Kimia, 2001). In target object recognition, the target object is segmented from the real 2D scene. Then the recognizer uses the target object to match the aspect graphs in the model library (Sebastian et al., 2001; Huttenlocher et al., 1993) after a series of pre-processing phases including Canny edge...
detection, translation, rotation, and mirror transformations. The recognizer constructs the final recognition result from the results that match.

We go on to discuss model simplification based on VRL and multi-resolution object matching algorithms, which are shown in the dashed rectangles in Fig. 1.

3. 3D Model simplification

3.1. Model simplification overview

The data size of an original 3D mesh model is enormous. For example, a traditional airplane model usually contains over 10,000 vertices and over 100,000 triangle patches. Without any simplification to reduce the amount of redundant information, this huge data size will make subsequent processing extremely computationally complex.

Current model simplification methods include vertex clustering (Rossignac, 1993), envelope control (Cohen et al., 1996), resampling (Turk, 1992), wavelet analysis (Eck et al., 1995), vertex deletion (Schroeder et al., 1992) and edge collapse (Silva and Gomes, 2004). In these methods, edge collapse is one of the most popular 3D mesh simplifications. Edge collapse scans all the meshes, computes the collapse cost, and finally folds a side with a cost less than the resolution limit. It repeats the process until the meshes reach the expected simplification degree.

Without considering human visual resolution, most existing edge collapse costs (Hoppe et al., 1993) are measured by the loss of surface roughness. However, since human eyes have limited visual resolution, an algorithm should discard details that humans cannot distinguish regardless of the degree of surface roughness. This requirement illustrates the importance and practical value of applying VRL to model simplification. In the next subsection, we discuss an improved edge collapse algorithm based on VRL.

3.2. Visual resolution limit (VRL)

According to resolution limit theory (Bennett and Rabbetts, 1989), the resolution of the human eye is limited. Angular resolution describes the minimum angle of two object points that visual cells could distinguish. The smaller the angle, the higher the resolution becomes. Currently, there are two basic theories of visual fields: receptor theory, which is based on the physiological structure of the human eye and the wave theory of light, which is based on diffraction when light passes through an optical system. The VRL we present in this paper will combine these two theories.

3.2.1. Receptor theory

Receptor theory says that human eyes can only distinguish two object points when there are non-stimulated cone cells between two stimulated cone cells. Fig. 2 shows the imaging plane of a receptor model. Receptor theory predicts that the human eye cannot distinguish point A from point B but cannot [ME4] distinguish point C from point D. The least distance between two distinguishable imaging points is \(2l\), where \(l\) is the scale of a sensitive cell.

Define \(d\) as the distance between the lens and the imaging plane; then the angular resolution could be calculated as follows:

\[
\delta = 2l/d.
\]

(1)

VRL could be achieved through the formula below:

\[
\Delta S = L \cdot \delta = 2lL/d.
\]

(2)

where \(L\) is the distance between the object and the observer.

3.2.2. The wave theory of light

The wave theory of light says that the image of a pointolite through an optical system is not a point, but a diffractive disc (Airy’s disc). The human eye can distinguish two pointolites clearly only when the distance between the centers of two Airy’s discs is greater than the Airy radius (referred as Rayleigh criterion). According to this criterion, angular resolution could be calculated as follows:

\[
\delta = 1.22 \lambda/D.
\]

(3)

where \(\lambda\) is the wavelength, and \(D\) is the diameter of the pupil. VRL could be achieved through the formula below:

\[
\Delta S = L \cdot \delta = 1.22 \lambda L/D.
\]

(4)

where \(L\) is the distance between the object and the observer.

![Fig. 1. Structure of the 3D recognition system.](image1)

![Fig. 2. The imaging plane of a receptor model.](image2)
3.2.3. A novel VRL

Traditional receptor theory and the wave theory of light only consider the effect of distance, neglecting the effects of object orientation, gesture and velocity. VRL cannot serve as the scale basis of a recognition system without considering these important factors.

Fig. 3 illustrates the VRLs of an object in different directions. In Fig. 3, \( \psi \) is the angle between the optic axis and the line from any point of the object to the lens center, which is called the azimuth angle. We assume that the distance between the object and the lens is much greater than the size of the object, which is true in most situations. Therefore, the distances between the lens and all the points of the object can be considered equal, and the same approximation is valid for azimuth angle \( \psi \) and resolution angle \( \delta \). \( \| AB \|, \| AC \|, \| AD \| \) are VRLs in different directions. The effective VRL of the object is the smallest value of \( \| AB \| \), which is approximately vertical to the sight line.

Fig. 4 illustrates the imaging system model of the human eye to get a more accurate VRL. Assuming that \( L \) is the distance between the object and the lens, \( d \) is the distance between the lens and the imaging plane, and \( \psi \) is the departure angle between the sight line and the optic axis. Denote by \( V \) the velocity of object and by \( T \) the exposure time. Let \( N_1 \) denote the unit vector parallel to the connection line between the object and the lens, and \( N_2 \) denote the unit vector parallel to \( BA \) and perpendicular to \( N_1 \). Assume that \( D \) is the lens diameter, and \( \kappa \) is the wavelength of the light source. \( l(\psi) \) could be defined as the scale of cone cells at angle \( \psi \).

If \( \psi = 0 \) and the object is static, angular resolution could be calculated by the traditional receptor theory directly as follows:

\[
\delta_1 = 2l(0)/d. \tag{5}
\]

If the object departs from the optic axis and does not move in the direction of \( N_\psi \), an equation of the geometrical relation in Fig. 4 is given by

\[
\| AC \| = (d/\cos \psi) \cdot \delta_2 = 2l(\psi) \cdot \cos \psi. \tag{6}
\]

Therefore, angular resolution \( \delta_2 \) could be gained by

\[
\delta_2 = 2l(\psi) \cos^2 \psi/d. \tag{7}
\]

In addition, considering exposure time, points of the high-speed object will leave tracks in the imaging system during exposure. To take these tracks into account, the angular resolution should be modified as follows:

\[
\delta_3 = \delta_2 + \delta_T = 2l(\psi) \cos^2 \psi/d + \| \hat{V} \cdot N_\psi \| T/L. \tag{8}
\]

Finally, the VRL based on the receptor theory could be achieved through the formula below:

\[
\Delta S_1(L, \psi, \hat{V}) = \| AB \| = L \cdot \delta_3 = 2l(\psi)L \cos^2 \psi/d + \| \hat{V} \cdot N_\psi \| T. \tag{9}
\]

Half-wave zone method[ME5] Wu et al. (2005) point out that if the azimuth angle \( \psi \) is not very large, angle resolution is independent of \( \psi \). Therefore the VRL based on the wave theory of light could be achieved through the following formula:

\[
\Delta S_2(L, \psi, \hat{V}) = L \cdot \delta_3 = 1.22lD + \| \hat{V} \cdot N_\psi \| T. \tag{10}
\]

Comparing the calculation results of Eqs. (9) and (10) in the same condition, the larger one should be considered the more effective VRL. Thus, the VRL could be calculated as follows:

\[
\Delta S(L, \psi, \hat{V}) = \max(\Delta S_1(L, \psi, \hat{V}), \Delta S_2(L, \psi, \hat{V})). \tag{11}
\]

3.3. Model simplification based on VRL

The main differences between various edge collapse methods are the definitions of collapse cost and target vertex selection methods. Traditional collapse algorithms (Hoppe et al., 1993) use the loss of surface roughness as the collapse cost. For example, if a vertex \( u \) id collapsed to another vertex \( v \), then the collapse cost can be calculated as follows:

\[
\text{cost}(u, v) = \max \min_{f \in F_u \in F_v} \left( 1 - \frac{f_{\text{normal}} \cdot n_{\text{normal}}}{2} \right). \tag{12}
\]
where \( T_u \) is a set of triangles containing vertex \( u \), \( T_v \) is a set of triangles containing vertices \( u \) and \( v \), and \( \mathbf{n} \) is the normal vector of the triangle \( \mathbf{n} \).

Without considering triangular face size, the “roughness-based criterion” might merge large triangular faces, which may contain valuable 3D object information. Therefore, we introduce the in-circle of a triangle (Ira and Thomas, 2001) into edge collapse simplification. Usually, the larger the in-circle of a triangle, the more geometrical features the triangle has. Under this circumstance, the vertices of the triangles with larger in-circle might have higher collapse costs and should be simplified later. So, the maximum in-circle radius of all the triangles around a certain vertex could be defined as the collapse cost of this vertex. In addition, if a vertex has a collapse cost less than the VRL, it cannot be distinguished from the adjacent points and should be collapsed. Thus, an algorithm could use VRL to control 3D model simplification. The entire simplification algorithm could be described as follows:

1. Search all the vertices. For each vertex, find out the maximum radius of inscribed circles of all the triangles around it and set this radius as its collapse cost. Construct the collapse cost list.  
2. Select the vertex with the lowest collapse cost from the collapse cost list, and set it as the collapsing target.  
3. Collapse the target vertex and update the collapse cost list.  
4. If all the radii of inscribed circles are greater than the VRL, the collapsing process is complete. Otherwise, repeat Step 2 and Step 3.

Taking the aircraft F4 as an example, Fig. 5 illustrates the results of comparing the traditional 3D model simplification method with our method. The simplification results show that some important features, such as wings and empenagles, have disappeared gradually during the traditional simplification process, which uses the loss of surface roughness as the collapse cost. In contrast, most basic features have been retained during the simplification process proposed in this paper, which uses the maximum radius of inscribed circles as the collapse cost.

4. Multi-resolution recognition

4.1. Multi-resolution aspect graph library foundation and recognition

In changing conditions, expressions of the model at different resolution levels could be obtained after the model simplification, and the object could be described by a set of multi-resolution aspect graphs. The foundation of multi-resolution aspect graph library is shown in Fig. 6. In the vertical direction, the VRL could be calculated according to the current observing condition. In the horizontal direction, VRL will be used to simplify a 3D model first. Then the algorithm will apply viewpoint space partition to the simplified model and cluster it. After these steps, the multi-resolution aspect graph library has been founded successfully.

Aspect graphs obtained after viewpoint space partitioning could not express the shapes of the model in all directions accurately. Therefore, recognition methods based on traditional Euclidean distance can be affected by these inaccurate expressions. On the other hand, the Hausdorff distance transform (Huttenlocher et al., 1993) can preferably bear the deformation, displacement and rotation transformations. Consequently, in this paper, we introduce Hausdorff distance based on projection contours to compute the similarity of two images.

After the target object is segmented from the observed image, its contour could be extracted by the Canny operator (Ali and Claudi, 2001), and be adjusted by translation, rotation, and mirror transformations. Then the Hausdorff distance could be used to compare the observed object and the multi-resolution aspect graphs in the library.

4.2. Gradual recognition

If the observed object is moving, then the VRL will change from time to time. To recognize the moving object, we introduce a multi-level matching strategy. For inter-class recognition, matching is as follows:

- Compare each of the progressive 2D images with a set of aspect graphs of one particular 3D model, which has the most approximate resolution with the current visual resolution limit. \( R_{k0} \) describes the matching result at level \( k \). If the current object is proximal to model \( i \) at level \( k \), then \( R_{k0} = 1 \), otherwise \( R_{k0} = 0 \).  
- A weighting coefficient \( W = 1/\Delta S \) describes the weight of the matching result at the current resolution level. The lower the value of \( \Delta S \), the more accurate the model is, and the more reliable the matching result is.  
- According to the matching results at different resolution levels, the matching value of model \( i \) is computed as follows: 
\[
S_i = \sum W_i R_{k0}.
\]
- If \( S_j = \max(S_i) \), the current object is recognized as model \( j \).

For intra-class recognition, the 3D model of the current object is already known. We could find the gesture of the observed object.
after comparing it with the aspect graphs with the most approximate resolution based on the current VRL.

However, practical incremental recognition methods or systems should be real-time. Thus, the computational complexity of Hausdorff distance transform will limit its use in real-time recognition systems. On the other hand, Fourier Descriptor (Folkers and Samet, 2002) replaces the Hausdorff distance transform in the incremental recognition process. The Fourier Descriptor is a detail descriptor that can also bear deformation, displacement and rotation transformations. Besides, the Fourier Descriptor can lower the processing time significantly, and it is about ten times faster than Hausdorff distance transform. With the Fourier Descriptor, the entire recognition process, specifically, segmentation, contour extraction and comparing Fourier coefficients, can complete within 1 s, which makes our 3D gradual recognition method practical and promising.

5. Experimental results

All the 3D models used in this paper are obtained from the standard 3D model library published by Princeton University (Princeton, 2007). We chose seven types of models from this library as test objects. Most of these models are airplanes, including F117, F4, F16 and Bomber. In addition, we also selected car models, including Bmobile and Racecar, to validate inter-class recognition, as well as including ship models to verify the recognition results for objects at different distances. The aspect graphs of the seven types of models (one for each) are shown in Fig. 7.

The experiments have three parts: VRL computation, multi-resolution aspect graph library foundation and multi-resolution recognition. Finally we will discuss some comparisons with other multi-resolution recognition methods, demonstrating the effectiveness of our method.

5.1. VRL computation

Table 1 lists the parameters of the imaging system in this experiment.

The area of one sensitive cell could be calculated as follows:

\[
S(\psi) = (5.76 \times 10^{-3} \times 4.29 \times 10^{-3})/(7 \times 10^6)
= 3.53 \times 10^{-12} \text{ (m}^2). \tag{13}
\]

The calculation results of Eqs. (15) and (16) show that \(d_2\) is greater than \(d_w\). Combining with Eq. (11), the VRL calculation could be simplified as follows:

\[
\Delta S = 2l(x, y) \cos^2 \psi / d + ||V \cdot N_\perp|| \hat{T}
= (2 \times 1.88 \times 10^{-6}) / (5.8 \times 10^{-3}) L \cos^2 \psi + ||V \cdot N_\perp|| \hat{T}
= 6.483 \times 10^{-4} L \cos^2 \psi + ||V \cdot N_\perp|| \hat{T}. \tag{17}
\]

According to Eq. (17), the computation results of the VRL value in different conditions are listed in Table 2.

When \(\psi = 0\), and \(||V \cdot N_\perp|| = 0\), the calculation results obtained from our method are the same as the results obtained from traditional receptor theory. However, if \(\psi \neq 0\), or \(||V \cdot N_\perp|| \neq 0\), the difference between the calculation results obtained through these two methods will be large. As illustrated in Table 2, when the distance \(L\) is small, for example, \(L = 500\) m, the influence of the velocity is stronger than that from the azimuth angle \(\psi\). And when the distance \(L\) is great, for example, \(L = 5000\) m, the influence from the azimuth angle \(\psi\) is greater. Therefore, considering both the

\[
H(\psi) = \sqrt{S(\psi)} = \sqrt{3.53 \times 10^{-12}} = 1.88 \times 10^{-6} \text{ (m)}. \tag{14}
\]

Supposing the visual angle limit is 63.5°, the part \(d_2\) in Eq. (8) of receptor theory could be calculated as follows:

\[
\delta_2 = 2l(\psi) \cos^2 \psi / d \geq 2l(\psi) \cos^2 (63.5/2) / d
= (2 \times 1.88 \times 10^{-6} \times \cos^2 31.75) / (5.8 \times 10^{-3})
= 4.688 \times 10^{-4}. \tag{15}
\]

For the wave theory of light, red light is used as the light source, with wavelength 770 nm. The part \(d_w\) in Eq. (10) is computed as follows:

\[
\delta_w = 1.22/\lambda \times (1 \times 770 \times 10^{-9}) / (2.07 \times 10^{-2})
= 4.538 \times 10^{-5}. \tag{16}
\]

The experiments have three parts: VRL computation, multi-resolution aspect graph library foundation and multi-resolution recognition. Finally we will discuss some comparisons with other multi-resolution recognition methods, demonstrating the effectiveness of our method.

Table 1

<table>
<thead>
<tr>
<th>Parameters of the imaging system.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCD (pixels)</td>
</tr>
<tr>
<td>7,000,000</td>
</tr>
</tbody>
</table>

\[
l(\psi) = \sqrt{S(\psi)} = \sqrt{3.53 \times 10^{-12}} = 1.88 \times 10^{-6} \text{ (m)}. \tag{14}
\]

Supposing the visual angle limit is 63.5°, the part \(d_2\) in Eq. (8) of receptor theory could be calculated as follows:

\[
\delta_2 = 2l(\psi) \cos^2 \psi / d \geq 2l(\psi) \cos^2 (63.5/2) / d
= (2 \times 1.88 \times 10^{-6} \times \cos^2 31.75) / (5.8 \times 10^{-3})
= 4.688 \times 10^{-4}. \tag{15}
\]

For the wave theory of light, red light is used as the light source, with wavelength 770 nm. The part \(d_w\) in Eq. (10) is computed as follows:

\[
\delta_w = 1.22/\lambda \times (1 \times 770 \times 10^{-9}) / (2.07 \times 10^{-2})
= 4.538 \times 10^{-5}. \tag{16}
\]

The calculation results of Eqs. (15) and (16) show that \(d_2\) is greater than \(d_w\). Combining with Eq. (11), the VRL calculation could be simplified as follows:

\[
\Delta S = 2l(x, y) \cos^2 \psi / d + ||V \cdot N_\perp|| \hat{T}
= (2 \times 1.88 \times 10^{-6}) / (5.8 \times 10^{-3}) L \cos^2 \psi + ||V \cdot N_\perp|| \hat{T}
= 6.483 \times 10^{-4} L \cos^2 \psi + ||V \cdot N_\perp|| \hat{T}. \tag{17}
\]

According to Eq. (17), the computation results of the VRL value in different conditions are listed in Table 2. When \(\psi = 0\), and \(||V \cdot N_\perp|| = 0\), the calculation results obtained from our method are the same as the results obtained from traditional receptor theory. However, if \(\psi \neq 0\), or \(||V \cdot N_\perp|| \neq 0\), the difference between the calculation results obtained through these two methods will be large. As illustrated in Table 2, when the distance \(L\) is small, for example, \(L = 500\) m, the influence of the velocity is stronger than that from the azimuth angle \(\psi\). And when the distance \(L\) is great, for example, \(L = 5000\) m, the influence from the azimuth angle \(\psi\) is greater. Therefore, considering both the

\[
H(\psi) = \sqrt{S(\psi)} = \sqrt{3.53 \times 10^{-12}} = 1.88 \times 10^{-6} \text{ (m)}. \tag{14}
\]

Supposing the visual angle limit is 63.5°, the part \(d_2\) in Eq. (8) of receptor theory could be calculated as follows:

\[
\delta_2 = 2l(\psi) \cos^2 \psi / d \geq 2l(\psi) \cos^2 (63.5/2) / d
= (2 \times 1.88 \times 10^{-6} \times \cos^2 31.75) / (5.8 \times 10^{-3})
= 4.688 \times 10^{-4}. \tag{15}
\]

For the wave theory of light, red light is used as the light source, with wavelength 770 nm. The part \(d_w\) in Eq. (10) is computed as follows:

\[
\delta_w = 1.22/\lambda \times (1 \times 770 \times 10^{-9}) / (2.07 \times 10^{-2})
= 4.538 \times 10^{-5}. \tag{16}
\]

The calculation results of Eqs. (15) and (16) show that \(d_2\) is greater than \(d_w\). Combining with Eq. (11), the VRL calculation could be simplified as follows:

\[
\Delta S = 2l(x, y) \cos^2 \psi / d + ||V \cdot N_\perp|| \hat{T}
= (2 \times 1.88 \times 10^{-6}) / (5.8 \times 10^{-3}) L \cos^2 \psi + ||V \cdot N_\perp|| \hat{T}
= 6.483 \times 10^{-4} L \cos^2 \psi + ||V \cdot N_\perp|| \hat{T}. \tag{17}
\]

According to Eq. (17), the computation results of the VRL value in different conditions are listed in Table 2. When \(\psi = 0\), and \(||V \cdot N_\perp|| = 0\), the calculation results obtained from our method are the same as the results obtained from traditional receptor theory. However, if \(\psi \neq 0\), or \(||V \cdot N_\perp|| \neq 0\), the difference between the calculation results obtained through these two methods will be large. As illustrated in Table 2, when the distance \(L\) is small, for example, \(L = 500\) m, the influence of the velocity is stronger than that from the azimuth angle \(\psi\). And when the distance \(L\) is great, for example, \(L = 5000\) m, the influence from the azimuth angle \(\psi\) is greater. Therefore, considering both the
velocity and orientation of the moving object, our method is preferable.

5.2. Multi-resolution aspect graph library foundation

The multi-resolution aspect graph library of six types of 3D models could be founded as follows:

1. Compute the VRL ($D_S$) by Eq. (17).
2. Change the value of $D_S$, which is used to control the model simplification.
3. Get the aspect graph representations of the simplified model at the current resolution level by viewpoint partitioning.

Fig. 8 shows multi-resolution aspect graph representations of the aircraft F4, Fig. 9 shows graphs for the model ship.

5.3. Multi-resolution recognition

Our recognition test consists of an intra-class gesture recognition test and an inter-class recognition test. Intra-class gesture recognition compares the test image with all aspect graphs of this model type, to determine whether the testing image can be classified into the right viewpoint region. Inter-class recognition compares the testing image with the aspect graphs of all the models to determine whether the testing image can be classified as the right model type. Fig. 8 shows the results of multi-resolution gesture recognition of F4 are shown in, and Fig. 10 illustrates one typical intra-class gesture recognition result for the model ship. Table 3 lists the viewpoints of input images and matched images. The holistic results of intra-class gesture recognition and inter-class recognition are shown in Tables 4–6. From the results of Tables 3–6, we can see that the intra-class gesture recognition results are satisfactory, but the inter-class recognition rates of some models are not very high. In order to improve matching performance, we can search for a new matching algorithm.

5.4. Comparison with other 3D object recognition methods

Most existing multi-resolution recognition methods are based on 2D projective images. The method of (Zhang et al., 2006) obtains a set of aspect graphs at the original scale with viewpoint space partitioning and clustering to the original 3D model. The multi-resolution aspect graph library will be founded after applying sub-
The multi-resolution aspect graph library obtained through the method of (Zhang et al., 2006). (a) is a set of aspect graphs at level 3. The image size of each graph is 96 × 96 (pixels × pixels). (b) is a set of aspect graphs at level 6. The image size of each graph is 32 × 32 (pixels × pixels).

Table 3
Viewpoint matching result.

<table>
<thead>
<tr>
<th>Input viewpoint</th>
<th>$\Delta s = 3.5$</th>
<th>$\Delta s = 1.75$</th>
<th>$\Delta s = 0.175$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 30$</td>
<td>$x = 20$</td>
<td>$x = 30$</td>
<td></td>
</tr>
<tr>
<td>$y = -20$</td>
<td>$y = -10$</td>
<td>$y = -20$</td>
<td></td>
</tr>
<tr>
<td>$z = -15$</td>
<td>$z = -30$</td>
<td>$z = 25$</td>
<td></td>
</tr>
<tr>
<td>Viewpoint space number</td>
<td>4</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Matched image viewpoint</td>
<td>$x = 29.28$</td>
<td>$x = -23.00$</td>
<td>$x = 28.80$</td>
</tr>
<tr>
<td>$y = -29.28$</td>
<td>$y = -9.53$</td>
<td>$y = -19.25$</td>
<td></td>
</tr>
<tr>
<td>$z = -17.15$</td>
<td>$z = -37.26$</td>
<td>$z = 28.43$</td>
<td></td>
</tr>
<tr>
<td>Hausdorff distance</td>
<td>40.50</td>
<td>46.17</td>
<td>17.89</td>
</tr>
</tbody>
</table>

Table 4
Intra-class recognition result: different models, same distance.

<table>
<thead>
<tr>
<th>Test model</th>
<th>F117</th>
<th>F4</th>
<th>Racecar</th>
<th>F16</th>
<th>Bomber</th>
<th>Btmobile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition rate</td>
<td>28/30</td>
<td>28/30</td>
<td>30/30</td>
<td>27/30</td>
<td>27/30</td>
<td>30/30</td>
</tr>
</tbody>
</table>

Table 5
Intra-class recognition result: model ship, different distances.

<table>
<thead>
<tr>
<th>Test distance (m)</th>
<th>0–2350</th>
<th>2350–3900</th>
<th>3900–4700</th>
<th>4700–6000</th>
</tr>
</thead>
<tbody>
<tr>
<td>VRL ($\Delta s$)</td>
<td>0.1</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Recognition rate</td>
<td>51/51</td>
<td>49/52</td>
<td>30/32</td>
<td>24/32</td>
</tr>
</tbody>
</table>

Table 6
Inter-class recognition result.

<table>
<thead>
<tr>
<th>Test model</th>
<th>F117</th>
<th>F4</th>
<th>Racecar</th>
<th>F16</th>
<th>Bomber</th>
<th>Btmobile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition rate</td>
<td>24/30</td>
<td>16/30</td>
<td>14/30</td>
<td>20/30</td>
<td>13/30</td>
<td>22/30</td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper, we have summarized and analyzed existing 3D object recognition algorithms based on aspect graphs, as well as
presenting some new ideas. Our 3D model simplifier uses an edge collapse algorithm based on the maximum radius of inscribed circles, which could effectively reduce redundancy while preserving the key features of the model. In addition, we have presented a novel visual resolution limit for the image system to control the simplification degree. Then we create multi-resolution aspect graph representations after applying viewpoint partitioning to the simplified models at different resolution levels, which lay a solid foundation for a real-time gradual recognition imitating human vision. Our results demonstrate that the multi-resolution recognition method introduced in this paper could be applied effectively both in intra-class gesture recognition and inter-class recognition.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. 60502013), and by 863 project (No. 2006AA01Z115).

References


