3D object recognition using the CAD model and stereo vision

Sungil. Kim and Sungjun. Choi and Sangchul. Won

Graduate School of Iron& Steel, POSTECH, Phohang 790-784, Korea
(Tel: +82-54-279-5576; Email:mocoen@postech.ac.kr)

Abstract: 3D object recognition is difficult but important in computer vision. The important thing is to understand about the relationship between a geometric structure in three dimensions and its image projection. Most 3D recognition systems construct models either manually or by training the pose and orientation of the objects. But both approaches are not satisfactory. In this paper, we focus on a commercial CAD model as a third type of model building for vision. The models are expressed in Initial Graphics Exchanges Specification(IGES) output and reconstructed in a pinhole camera coordinate.

Keywords: CAD model, IGES, Stereo Vision,

1. Introduction

The object recognition begins with designing adequate models. It depends largely on the object to be represented and the algorithms at hand how to choose the method of model representations. Generally, most model representations in 3D object recognition come from one of two direction: feature-based and appearance-based.
Feature-based models represent 3D objects through features, their type, and their spatial relations. The identification means finding a set of features which is uniquely distinctive for an object. And, location is to match a number of image and object features and solve for the position and orientation of the 3D object. The advantage of feature-based models is that they generate compact object descriptors, offer some robustness against occlusion(if the features are local), and some invariance against illumination and pose variations. A disadvantage is that they cannot be compared directly with images and require feature extraction and object descriptions is obviously time-consuming and requires detailed knowledge of the internal structure of the object recognition system.[1] It is impractical in applications where the set of objects to be recognized is large or changes frequently.[2]
Appearance-based models represent an object through one or more images, as in eigenspaces method. Models are constructed using prototypical features and extracted from images of the object to be model. Recognition means to find the image in a model set which is most similar to the one to recognize. The advantage is that images and models can be compared directly, and objects with no obvious features can still be modelled. Disadvantages include the fact that illumination, pose and location variations alter the images.[1]
In recent, the integration of mechanical CAD systems and computer vision system introduce a third type of model-building for vision. CAD model is kind of a feature-based solid modeler that aids efficient building of 3D object models. It provides a quick and compact object representation by incorporating view-dependant features and the object-centered representation.
In this paper we focus on a scheme to produce a certain model using the commercial CAD system. IGES, Initial Graphics Exchange Specification, has been used as an interface to the CAD system to infer the geometric information necessary for object recognition. IGES is an industry standard developed by the National Institute of Standards and Technology to allow for transfer of CAD models between various commercial CAD systems. We made several type of CAD models and reconstructed it. To demonstrate the feasibility of our reconstructed CAD model, We recognize 3D coil shaped objects and identify its poses. The object recognition system works in two modes, off-line mode and run-time mode. In the off-line mode, vision oriented data derived from a 3D CAD system are used to generate the model database for object recognition. In the on-line mode, depth map as a result of a stereo vision is generated and edge feature is extracted to recognize the pose and location of object.

2. A camera model and stereo vision

2.1. A camera model

The central projection describes the projection of an object surface point of an object surface point of a three-dimensional scene into the two-dimensional image plane. The projective relationship between the concerned coordinate systems(world coordinates, camera coordinates, ideal image coordinates etc.) is of importance for techniques for object reconstructions.
For the assumed ideal pinhole camera the projection of scene points $\mathbf{P} = (\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ of the XYZ-space on to image points $\mathbf{p} = (\mathbf{x}, \mathbf{y})$ of the *xy*-image plane is a central or perspective projection. If the coordinate system of the XYZ-space is defined to be the *camera-centered coordinate system* (focal point = projection center, optical axis = Z-axis) then the projection equations are given by

$$x = \frac{f_k \cdot X}{Z} \qquad \text{and} \qquad y = \frac{f_k \cdot Y}{Z} \tag{1}$$

XYZ-coordinate system that is adjusted with the camera is called a *camera coordinate system* and is labeled by an index *k*. And, a $X_w Y_w Z_w$ *world coordinate system* is assumed which is adjusted with a selected object that is positioned in the three-dimensional space of the scenes. The camera constant(f_k,focal length) describes the orthogonal distance between the center of the camera lens and the image plane. The image plane is assumed to be behind the $X_k Y_k$ plane in a distance of f_k. World coordinates and camera coordinates can be transformed into each other through a simple Eu-

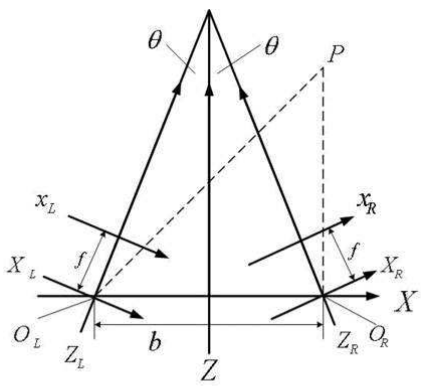


Fig. 1. Geometry of a static stereo system.

clidean transformation. Let x_b and y_b be the coordinates of pixels in the digital image which is considered to be the final result of an image acquisition process. The mapping, object points $\mathbf{P} = (\mathbf{X_w}, \mathbf{Y_w}, \mathbf{Z_w})$ in world coordinates are projected into digital image points $\mathbf{p} = (\mathbf{x_b}, \mathbf{y_b})$, can be split into several coordinate transforms and is described by the following linear transformation.

$$\begin{pmatrix} x_b - c_x \\ y_b - c_y \\ -f_k \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{pmatrix} \cdot \begin{pmatrix} X_w - X_0 \\ Y_w - Y_0 \\ Z_w - Z_0 \end{pmatrix} \quad (2)$$

(c_x, c_y) are the coordinates of the principal point,the intersection point of the optical axis with the image plane, according to the image coordinate system. r_i are the coefficients of the rotation matrix $\mathbf{R} = \mathbf{R_x} \cdot \mathbf{R_y} \cdot \mathbf{R_z}$ of the Euclidean transformation(world into camera coordinates). And, (X_0, Y_0, Z_0) are the coordinates of the projection center in world coordinates.[3]

2.2. A stereo vision

Stereo vision refers to the ability to infer information on the 3D structure and distance of scene from two or more images taken from different viewpoints. For the static stereo analysis it is assumed that no object movements occur during the time interval of image acquisition of the two stereo images. We assume two cameras with identical effective focal length f such that the distance between their focal points $\mathbf{O_L}$ and $\mathbf{O_R}$ is equal to b. Both cameras may be tilted towards each other. This allows that more object faces will be visible in both images. We assume parallel Y-axes for both cameras.

The line between the two optical centers $\mathbf{O_L}$ and $\mathbf{O_R}$ is called *base line* and value b is called *base distance*. Let the angle subtended by the two optical axes be 2θ. The coordinate systems $X_L Y_L Z_L$ and $X_R Y_R Z_R$ are the left and right camera coordinates. The coordinate system XYZ is defined such that Z-axis exactly divides the angle subtended by the Z_L axis and Z_R axis into two identical angles θ.(Fig.1.)

The coordinate system XYZ can be transformed into the coordinate system $X_L Y_L Z_L$ through a rotation by the angle $-\theta$ about Y-axis.

$$\begin{pmatrix} X_L \\ Y_L \\ Z_L \end{pmatrix} = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} X - \frac{b}{2} \\ Y \\ Z \end{pmatrix} \quad (3)$$

Analogously the coordinate system XYZ can be transformed into the coordinate system $X_R Y_R Z_R$ through a rotation bye the angle θ about the Y-axis and a translation by b/2 to the right.

$$\begin{pmatrix} X_R \\ Y_R \\ Z_R \end{pmatrix} = \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} X + \frac{b}{2} \\ Y \\ Z \end{pmatrix} \quad (4)$$

A point $\mathbf{P} = (\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ in 3D scene space is projected onto the points (x_L, Y_L) and (x_R, y_R) in the image planes of the two cameras. For an assumed central projection it follows that

$$x_L = \frac{f \cdot X_L}{Z_L}, \qquad y_L = \frac{f \cdot Y_L}{Z_L}, \quad (5)$$

$$x_R = \frac{f \cdot X_R}{Z_R}, \qquad y_R = \frac{f \cdot Y_R}{Z_R}, \quad (6)$$

We consider the case that the two corresponding image points(x_L, y_L) and (x_R, y_R) are correctly determined in the left and in the right image, respectively. Then the three-dimensional position of the point $\mathbf{P} = (\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ can be calculated from the equations (3),(4),(5) and (6). It holds that

$$[-x_L \cdot \sin\theta - f \cdot \cos\theta]X + [x_L \cdot \cos\theta - f \cdot \sin\theta]Z = -[\frac{b}{c} x_L \sin\theta + \frac{b}{2} f], \quad (7)$$

$$[x_R \cdot \sin\theta - f \cdot \cos\theta]X + [x_R \cdot \cos\theta + f \cdot \sin\theta]Z = -[\frac{b}{c} x_R \sin\theta - \frac{b}{2} f], \quad (8)$$

$$[-y_L \cdot \sin\theta]X + [-f]Y + [y_L \cdot \cos\theta]Z = [\frac{b}{2} y_L \cdot \sin\theta] \quad (9)$$

$$[y_R \cdot \sin\theta]X + [-f]Y + [y_R \cdot \cos\theta]Z = -[\frac{b}{2} y_R \cdot \sin\theta] \quad (10)$$

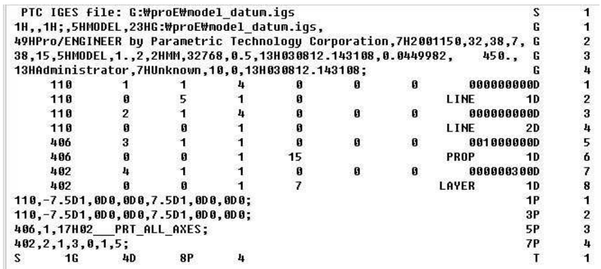
The coordinates X,Y and Z can be determined by solving this linear equations. This process is called triangulation.[3]

3. CAD model

3.1. Neutral data exchange format

Currently, an enormous number of CAD systems are available on the market. The CAD date exchange between other systems is often impossible due to different encoding methods applied by different vendors or even the same vendor at different versions. So, the neutral file mechanism is introduced for translation and reduce the complexity the difficult of direct data transfer between any two systems. Many neutral data exchange systems have been proposed.

IGES(The Initial Graphics Exchange Specification) is standard which has been widely accepted. PDES(Product Data Exchange Specification),derived from IGES, expands and updates its functions with greater applications. The Germans and French have also developed their own version of standardization:SET(Standard d'Echange et de Transgert) and VDA-FS(Verband des Automobilindustrie Flächen-Schnittstelle). And, the Internationsl Standard Organization(ISO) combine all proposals and conclude STEP(The



```
 PTC IGES file: G:₩proE₩model_datum.igs                                   S      1
1H,,1H;,5HMODEL,23HG:₩proE₩model_datum.igs,                               G      1
49HPro/ENGINEER by Parametric Technology Corporation,7H2001150,32,38,7,   G      2
38,15,5HMODEL,1.,2,2HMM,32768,0.5,13H030812.143108,0.0449982,    450.,    G      3
13HAdministrator,7HUnknown,10,0,13H030812.143108;                         G      4
     110       1       1       4       0       0       0        000000000D      1
     110       0       5       1       0                    LINE        1D      2
     110       2       1       4       0       0       0        000000000D      3
     110       0       0       1       0                    LINE        2D      4
     406       3       1       1       0       0       0        001000000D      5
     406       0       0       1      15                    PROP        1D      6
     402       4       1       1       0       0       0        000000300D      7
     402       0       0       1       7                   LAYER        1D      8
110,-7.5D1,0D0,0D0,7.5D1,0D0,0D0;                                        1P      1
110,-7.5D1,0D0,0D0,7.5D1,0D0,0D0;                                        3P      2
406,1,17H02___PRT_ALL_AXES;                                              5P      3
402,2,1,3,0,1,5;                                                         7P      4
S      1G      4D      8P      4                                          T      1
```

Fig. 2. IGES file example.

Standard for Exchange for Product Model Data) as the future international standard.

Although there are several neutral data exchange format currently implemented, we will limit our interest to IGES. It is still widely accepted and enough to represent the model.[4]

3.2. IGES

The Initial Graphic Exchange Specification(IGES) was formed in 1979 with technical team members from the National Bureau of Standards, Boeing, and General Electric. The main purpose of IGES is to exchange product definition data between different CAD/CAM systems. In 1981 it was accepted by ANSI as standard for the exchange of product definition data.

IGES retains the 80 column card format where 1-72 columns contain data in ASCII code and 73-80 columns are serial numbers leading by an alphabetic character indicating the section. IGES separates the file into five sections. To distinguish between the five sections, each section has its own identification character letters. "S" ,"G","D","P" and "T" are used as identification letters for start, global, directory, parameter data and terminate section respectively.

The start section is used to provide a human-readable prolog to the file. This prolog is not formated and should be written using the ASCII characters.

The Global section contains information describing the sending system and reference data required by the receiving post-processor. The parameters in the global section are in a free format.

The directory section has the same fixed format for all entities. The directory entry for each entity is fixed in size and contains 20 fields of 8 characters each. It provides an index for the file and contain attribute information for entity.

The parameter section contains related parameters needed for an entity. It has a different format for each type of entity, with at least one line for each entity. The first parameter in a parameter data entry is always the entity type number. In the parameter data entry of each entity, the parameter delimiter character is used to separate two consecutive parameters.

The terminate section contains the total number of lines in each of the previous sections and is represented at the end of the file.[5]

3.3. Model reconstruction from IGES file

We used the pro-engineer2001 as commercial 3D CAD modeler. The pro-engineer2001 is well known CAD system, so it supports many neutral CAD data exchange format. The IGES format is also provided by the pro-engineer2001. In IGES file a product is described as a collection of geometric and non-geometric entities. The geometric entities(type 100, 110 and 116 etc.) represent the definition of the physical shape,i.e. points, lines, circles, parametric splines, etc. while the non-geometric entities(type 212, 214, 222, 404, 406 and 410 etc.) are used to enhance the model by providing annotations and dimensions. All entities are not used. We chose only geometric entities.

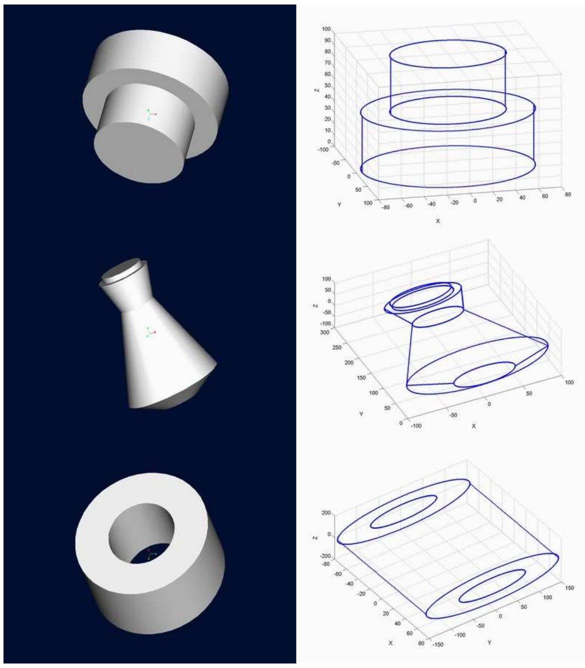


Fig. 3. CAD models and reconstructed CAD models.

Using the circular arc entity(type 100) and line entity(type 110), we represent the wire frame model from the IGES file. Circular arc entity(type 100) composite of more then one point. Its parameters are list of central point, starting point and terminal point. The starting point leaves in the directions of count-clockwise and arrive at the terminal point. The arc lies on the XT, YT plan.

$$C(t) = (X_1 + R\cos t, Y_1 + R\sin t, ZT), \qquad t_2 \le t \le t_3 \quad (11)$$

where i=2 and 3,
(1) $R = \sqrt{(X_i - X_1)^2 + (Y_i - Y_1)^2}$
(2) $(R\cos t_i, R\sin t_i) = (X_i - X_1, Y_i - Y_1)$
(3) $0 \le t_2 \le 2\pi$,$0 \le t_3 - t_2 \le 2\pi$
It is transformed to 3D XYZ-coordinate through a transformation matrix entity(type 124).

Line entity(type 110) is defined as the end of two points.It is defined in the 3D defined coordinate space.

$$C(t) = P_1 + t(P_2 - P_1), \qquad 0 \le t \le 1 \quad (12)$$

We reconstructed several wire frame models from the IGES file. Fig3. shows them.[6]

4. Experiment

To demonstrate the feasibility of our reconstructed CAD model, we matched reconstructed CAD model from the IGES

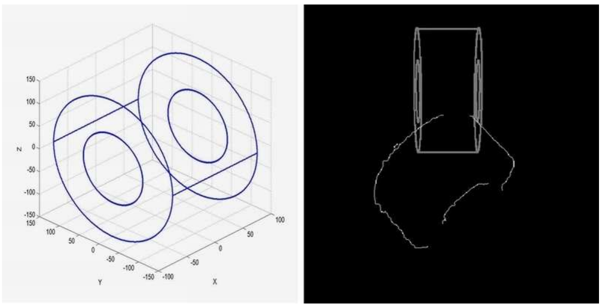


Fig. 4. Model and initial matched camera projected image.

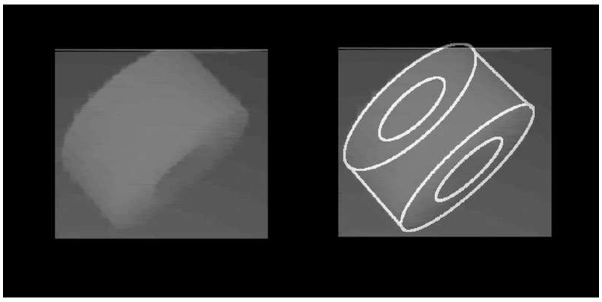


Fig. 5. Object range image and result image.

file to a range image. The range image is obtained from a stereo vision. We used the object-centered representations which attach a reference frame to an object and express the object geometry in that frame. The center of the reconstructed CAD model is placed at the origin of the world coordinate at the first time. The camera extrinsic parameters are $[T_x, T_y, T_z] = [0, -692, 1200]$(mm) and $[R_x, R_y, R_z] = [-157, 0, 180](^\circ)$ respectively. The object,coil shaped object, is arbitrarily placed. The edge of the object is extracted using the Canny's edge detecting algorithm. Our model,i.e reconstructed wire CAD model, is placed at the origin of the world coordinate. Fig4. shows wire model in the world coordinate(left) and the camera projected image, initially placed the wire model and the edge image of the range image(rignt). We minimized the mean squared error between the wire model and edge image and found an wire model corresponding to the object was correctly detected. Fig5. shows the range image and result image. We can identify the object is placed at $[T_x, T_y, T_z] = [7, -241, 150]$(mm) and $[R_x, R_y, R_z] = [0, 0, 49](^\circ)$ in the world coordinate.

5. Conclusion

In this paper, we have presented a method of building a model from a CAD systems and recognize the 3D position and orientation of coil shaped object. Unlike other object recognition systems, we used a commercial 3D CAD system to design the models. And, IGES has been used as an interface to the CAD system to infer the geometric information necessary for object recognition. We reconstructed several models. One of the models is matched and recognized to check the its feasibility. Through the experiment, we identified that the CAD model derived from its neutral data exchange format is easy to build and flexible to use whenever the object is frequently changed and large.

References

[1] Emanuele Trucco,Alessandro Verri, "Introductory Techniques for 3D Computer Vision", Prentice-Hall,Inc,1998.

[2] Patrick J. Flynn, and Anil K. Jain, " CAD-Based Computer Vision: From CAD Models to Relational Graphs", IEEE Transctions on pattern analysis and machine intelligence,vol 13,no 2 1991.

[3] R. Klette, K. Schluns, A.Koschan "Computer Vision, Three-Dimensional Data from Images", Springer,1998.

[4] C.Y.Ho, Shyh-Dong Tsuei, Feng-Gi Yang "IGES and PDES, The Current Status of Product Data Exchange Standard", IEEE ,Univ. of MO-Rolla,1988.

[5] M. Kalta and B.J. Davies, "Converting 80-Character ASCII IGES Sequential Files into More Conveniently Accessible Direct-Access Files," The International journal, advanced manufacturing technology ,vol 8,no 3 1993.

[6] J.Y.Seong and M.S.Kim, "A Study on the Utlization of CAD Design Data Using IGES Format," POSTECH Graduate School for Information Technology,1994.