A 3D image processing method for manufacturing process automation

Dongming Zhao*, Songtao Li

Electrical and Computer Engineering, University of Michigan-Dearborn, 4901 Evergreen Road, Dearborn, MI 48128, USA

Received 1 December 2004; received in revised form 31 March 2005; accepted 31 May 2005
Available online 26 September 2005

Abstract

Three-dimensional (3D) image processing provides a useful tool for machine vision applications. Typically a 3D vision system is divided into data acquisition, low-level processing, object representation and matching. In this paper, a 3D object pose estimation method is developed for an automated manufacturing assembly process. The experimental results show that the 3D pose estimation method produces accurate geometrical information for an automated material handling application.

Keywords: Range imaging; 3D image processing; Pose estimation; 3D object recognition; Pattern recognition; Robot vision; Manufacturing technology

1. Introduction

Machine vision is an important subject within the machine intelligence systems area. The goal of a machine vision system is to understand given image data and use the extracted information for a high-level task. For a comprehensive machine vision system, the objects in the following cases should be identified: (1) objects may have arbitrary and complicated shapes; (2) objects may be viewed from various directions; (3) objects may be partially occluded by other objects in a scene. In a manufacturing environment, such a machine vision system can be used to determine the locations of grasp sites for a robot to manipulate for an object in the assembly or inspection of parts.

To realize such vision systems, system designers must resolve the following issues: (1) the type of sensors for data collection, (2) the methods of describing the collected data and the object, and (3) the methods of matching the object descriptions obtained from the input data to the desired object model. The sensor determines the resolution and precision. More importantly, it determines whether the data provides two-dimensional (2D) or 3D information of the scene. Object representations are used to describe the collected data and the desired object
model. The descriptions are also used to calculate the various properties of objects in the scene during the matching stage. Matching strategies are performed for real-time on-line processes and must resolve many ambiguities between the input data and the model descriptions. Once the match has been established, the orientation and location information of the target object can be computed for the further robot tasks.

For an automated assembly application, a 3D object pose estimation method is developed in our study using range images as the input data. The local surface geometrical feature is described by a 3D surface descriptor, Angle Distance Map (ADM). A triangular mesh model of the 3D object surface is used for decreasing the cost of computing and storage of applying the ADM descriptions. The principal component analysis (PCA) method is applied on the original ADM description to extract the principal features for matching strategies. The pose information is computed according to the final 3D feature points. The proposed approach is tested in an application for flexible robot assembly. The experimental results show that accurate pose estimation can be obtained for 3D machine vision applications.

2. 3D computer vision system

The basic issues involved in the design of a computer vision system are outlined in Section 1. Fig. 1 shows a general paradigm of a computer vision system. The system includes on-line and off-line processes. In an off-line process, a description of a standard object is established as the model object. In an on-line process, the input data are represented by the same description method that is used for the model description in the off-line process. The input object description and the model description are then compared in the matching stage. The located object and its pose information are used for the further applications.

2.1. Range image sensing

Active imaging acquires range images with dense and accurate range information. Laser radar imaging is one approach that produces high quality range data. The advantages and disadvantages in passive and active sensing can be found in a number of literatures [2,3]. In this study, laser radar imaging is used to acquire images, where a continuous-beam laser is used to send laser light to the object and collect the returned signals [4]. The phase shift in the return signal is used to measure the distance. A range image is generated by scanning the laser beam horizontally and vertically across the scene, as shown in Fig. 2. A laser beam is
sent out onto faceted polygon, and is then reflected onto a nodding mirror. The mirror reflects the beam to the targeted object, as shown in direction Z in Fig. 2. The imaging object is within the viewing scope of the laser beam scanned area on the mirror. When the laser beam returned from the object is received at the mirror, it is transmitted to the sensor and then to the demodulation processor. The laser beam is modulated on a continued amplitude sinusoidal wave with a consistent frequency. The phase shift is compared and interpreted as the geometrical differentiation pixel by pixel on the received image grids, and such phase shift parameters are then translated into the range information.

2.2. Edge extraction and region segmentation

Range scanners sample points on the object surface. The surface properties, such as surface normal and surface curvature, are then computed for segmentation and edge detection. Many algorithms are well developed to deal with improving image quality, modeling and reducing noise effects \cite{5}, edge extraction and segmentation \cite{6,7}, region growing \cite{5,8}, and region segmentation \cite{9,10,27}.

In range image processing, unlike traditional 2D image processing, there are three primary types of edges in 3D range images. These types of edges are step edges which are discontinuous in depth value, roof edges which are discontinuous in surface normals, and curvature edges which are discontinuous in variety of surface normals. Since the roof and curvature edges do not correspond to depth discontinuity, the edge detection in range images is more complicated than in traditional intensity images \cite{11–14}.

2.3. 3D object description

Design of an appropriate surface representation scheme is a crucial factor for 3D object recognition (see the surveys \cite{1,2,15,28}). Both local and global based representations, such as points or groups of points, straight line segmentations and polylines, planar and quadric surface patches, constructive surface geometry and superquadrics, have been proposed to describe a 3D object surface. The methods can also be found to deal with surface normals \cite{16–20}, and non-restricted smooth surfaces \cite{15,21}.

2.4. Matching and pose estimation

Once the appropriate descriptions are derived from the range images, the 3D computer vision system is able to match the two descriptions for completing the task of object recognition. This is performed in two steps. First, a correspondence is established between the two sets of descriptions. For some partial occlusion cases, only partial descriptions of objects can be obtained. The matching strategy establishes the correspondence between the partial description of the object and its full model description. The calculation of correlative coefficients is used to match the simple descriptions, such as feature vectors. In the second step, using the established correspondence, a geometrical transformation is derived. The 3D orientation information of the object in the scene is then estimated according to the geometrical transformation.

Object pose estimation is an important application in 3D object recognition. There are two primary approaches to obtain pose estimation: feature based approach and model parameter based approach. In feature based approaches, pose information is directly computed from the derived features corresponding to model features \cite{22,23}. In model parameter based approaches, a parametric object model is obtained from raw data at first. Pose information is computed from the model parameters \cite{24,25}.

3. A pose estimation method for an automated assembly application

In this study, a 3D object pose estimation method is developed for guiding flexible robots on an automated assembly line. Accurate pose information, including the position and orientation of torque converters, is needed to guide a robot to grab a target torque converter. A laser scanning camera is mounted on top of the working cell. Fig. 3 shows the work field and a grasping example of the automated assembly application.

Fig. 4(a) and (b) show respectively the intensity image and its corresponding range image. There are six torque converters laid in a bin as shown in Fig. 4. The objects, torque converters, are stacked in a bin of four layers. Using the range images, the developed 3D
object pose estimation method provides the orientation and position information of each torque converter. The framework includes two stages: (1) segment each torque converter and locate the center positions of torque converters; (2) determine the orientation of the located torque converters. A diagram of the integrated 3D pose estimation method is shown in Fig. 5.

3.1. Edge extraction and processing

The first step in our system is to process the input workcell range image and find localizations of the stack area and each torque converter inside the stack area. Edge detection and segmentation approaches are applied on the original range images.

As shown in Fig. 3(b), there are only step edges between the stack area and the background. A normal intensity-based edge detection method is used to extract the edges. Fig. 6(a) shows the edge detection result of Fig. 4(b). The stack area is a large rectangle area in the range images. Using this knowledge, a Hough transform is applied to the detected edge results. The Hough transform is a traditional image processing algorithm to extract some geometrical shapes, such as line and circle. After the edges in the stack area are extracted, nine sub-areas are equally segmented. These sub-areas are applied as inputs for the orientation and position estimation algorithm. Fig. 6 shows the edge detection and segmentation results.
3.2. 3D surface representation using angle distance map

In range images, surface shape is described by a dense collection of 3D points. The 3D points with their associated 3D positions and surface normals are used as the fundamental components of our surface representation. An Angle Distance Map is a 3D surface descriptor proposed to describe the local geometric features using global surface information. Using point P and neighborhood point Q, as shown in Fig. 7, angle $\alpha$ between surface normals $n_1$ and $n_2$ of these two points and their 3D distance $d$ are the two parameters in the ADM descriptions.

The ADM $M_P$ of point P is defined as the function that projects its neighborhood 3D points to the 2D coordinates corresponding to point P

$$M_P : \mathbb{R}^3 \rightarrow \mathbb{R}^2, \quad M_P(Q) \rightarrow (\alpha, d).$$

If the coordinates of points P and Q are $(x_1, y_1, z_1)$ and $(x_2, y_2, z_2)$, and their surface normals are $(f_{x1}, f_{y1}, -1)$ and $(f_{x2}, f_{y2}, -1)$, angle parameter $\alpha$ and distance parameter $d$ are given by

$$\alpha = \arccos\left(\frac{f_{x1} f_{x2} + f_{y1} f_{y2} + 1}{\sqrt{f_{x1}^2 + f_{y1}^2} + 1 \sqrt{f_{x2}^2 + f_{y2}^2} + 1}\right),$$

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$  

In Angle Distance Map descriptions, parameter $\alpha$ is an angle in degrees between 0 and 180. Parameter $d$ is a 3D distance in Cartesian coordinates. Normally the proposed ADM descriptions are square maps. The size of an ADM description is determined by the degree scale of parameter $\alpha$. The distance area changes with different distance scales of parameter $d$.

The ADM based surface representation describes the geometric features of a 3D point. For an object...
a triangular mesh model of the 3D object is used to contain the computing complexity and the size of data representation of the ADM description. A triangular mesh is a piece-wise linear surface, which consists of vertices, edges and faces. In this study, the vertices of the triangular meshes on an object surface are used to generate the ADM descriptions for the object surface representation.

Fig. 8 shows an example of the ADM description of a point on a cone surface. As shown in Fig. 8(a), for vertex P on the cone surface, ADM description $M_P$ is obtained. The size of $M_P$ is selected to be $180 \times 180$, and the scale of distance is selected to be 1. The ADM description $M_P$ is shown in Fig. 8(b).

In the ADM based surface description, each vertex has an ADM description. The surface geometric features are represented according to the distributions of ADM descriptions.

The ADM descriptions are compared using their correlation coefficients. Given two ADM descriptions $M_P$ and $M_Q$ with a size of $N \times N$, the linear correlation coefficient, $R(M_P, M_Q)$, is given by

$$R(M_P, M_Q) = \frac{N \times N \sum_{i} \sum_{j} p_{ij}q_{ij} - \left( \sum_{i} \sum_{j} p_{ij} \right) \left( \sum_{i} \sum_{j} q_{ij} \right)}{\left( N \times N \sum_{i} \sum_{j} p_{ij}^2 - \left( \sum_{i} \sum_{j} p_{ij} \right)^2 \right)^{1/2} \left( N \times N \sum_{i} \sum_{j} q_{ij}^2 - \left( \sum_{i} \sum_{j} q_{ij} \right)^2 \right)^{1/2}}.$$

(4)

Coefficient $R$ is between $-1$ (anti-correlated) and 1 (completely correlated), and it provides a measure for comparing two ADM descriptions. The surface representation of an ADM description is estimated through the correlation coefficient.

3.3. 3D object pose estimation for automated assembly application

In this automated assembly application, the 3D orientation and position information of each torque converter is required for the robot system. In this 3D object pose estimation method, the position and orientation information is computed using separate methods.

3.3.1. Center position estimation

The contour of torque converters is circle from the camera’s view direction. The circle center is used as the center position of each torque converter. In the low-level processing, each torque converter is

![Fig. 8. An ADM description example of a cone: (a) the range image and surface mesh of part of a cone; (b) the ADM description result of vertex P.](image)
segmented within each sub-area. The Hough transform is applied to extract the circle shape within each sub-area. The center position of the extracted circle is taken as the position of the torque converter. Fig. 9 shows the center position results of Fig. 6(b).

3.3.2. Orientation estimation using ADM descriptions

The torque converters appear in different sizes and directions due to the height of a stack, and torque positions in a stack and the different orientations of torque converters themselves. Fig. 10 shows two examples with their intensity images of size $130 \times 130$ pixels and size $105 \times 105$ pixels. The top circular region of a torque converter is a ring-shape band, as illustrated in Fig. 9. This ring-shape band is an area identified as a region that is most representative to the pose orientation of the object of interest. The normal of the tangent plane of the ring-shape area is taken as the torque converter orientation. Notice that the surface normals are taken from range images, and the ring shape bands are superimposed on intensity images for easy viewing by human.

In this orientation estimation, the proposed ADM descriptor is applied to represent the 3D surface features. The correspondences between pairs of the features are obtained through the ADM descriptions. Since the ADM description is a local surface representation, the points belonging to the ring-shape areas are extracted first. The tangent plane of the ring-shape area is then generated through fitting the extracted feature points. Finally, the pose is estimated according to the surface normal of the extracted tangent plane.

For an efficient comparison between two original ADM descriptions, the principle component analysis is applied to compress ADM descriptions. The PCA or the Karhunen–Loeve expansion provides a method to automatically determine a linear subspace with a minimal number of dimensions. Such a representation is optimal in terms of the least-square error [26]. By using only the most important eigenvectors as defined by the corresponding eigenvalues, the dimension of the eigenspace is significantly reduced with a minimal loss in precision and reliability. This compression process greatly reduces computing complexity.

In this orientation estimation method, the PCA is used to compress the original ADM descriptions consisting of all vertices on the object surface. Since the original ADM vectors are correlated, the Compressed ADM (CADM) descriptions are obtained through eliminating the eigenspace dimensions whose eigenvalues are relatively small. The projected CADM dimension, $s$, is determined according to the needed fidelity in reconstruction and the variance among ADM vectors.

The extracted subeigenspace constructed by $s$ significant eigenvectors and the $s$-dimensional CADM vectors of model features are used to match the feature points in the pose estimation method.

As shown in Fig. 11, a general procedure of the 3D object pose estimation method is divided into two sections, the off-line training and the on-line testing. In the off-line training section, the feature points used to estimate orientation information are specified on the
model object surface. The model object surface is first described in terms of the surface meshes. The ADM descriptions of mesh vertices are then obtained. The eigenspace of all ADM descriptions is constructed based on the PCA approach. According to the reconstruction accuracy and the variance among the ADM vectors, the CADM dimension $s$ and corresponding $s$ significant eigenvectors are generated. The CADM vectors of the specified feature points are then obtained by projecting the original ADM descriptions onto the subeigenspace constructed using the extracted $s$ significant eigenvectors. The CADM vectors of the specified feature points and the $s$ significant eigenvectors of the subeigenspace are the two components for the online testing process.

In the online testing section, the surface mesh representation for object surface is first generated and then the ADM descriptions of the mesh vertices are obtained. All ADM descriptions are projected into the subeigenspace using the $s$ significant eigenvectors which are extracted during the online training process. The projected CADM vectors are matched with the obtained CADM vectors of the specified feature points of the model object. The correspondences of the feature points are established by the maximal correlation coefficient between the training sampler and the testing objects.

Once the feature points are matched, the tangent plane of the ring-shape area is estimated according to the least-square surface fitting. Some points not belonging to the ring-shape area are also extracted due to the sensor error or the difference between the model torque converter and the tested torque converter. Therefore, an iterative outlier elimination method is applied based on the 3D projection distance between the points and the estimated tangent plane. For each elimination step, the point with the largest distance to the fitted tangent plane is found first. If the largest distance is larger than a predefined threshold, the point is eliminated from the feature point set. This iterative process terminates only if no point is eliminated from the current feature point set. The tangent plane fitted by the final feature points is used to compute the orientation of a torque converter.

4. Experiments

In our experiments, 45 torque converters placed in eight stacks with different positions and levels in a workcell are tested by the proposed orientation estimation approach. Ten torque converters with different positions and levels are used as model parts to obtain the pose information of all of the 45 torque converters. The orientation information of each torque converter is computed according to the Cartesian coordinates of range data. The errors between the computed accurate orientation and the estimated orientation of each torque converter are calculated.
to evaluate the accuracy of this approach. Fig. 12(a) shows a one-level stack with five torque converters. A model torque converter with specified feature points is shown in Fig. 12(b).

Fig. 13 show the final matched feature points of the torque converters in Fig. 12(a) according to model part in Fig. 12(b). The mean angle errors of these torque converters are shown in Table 1. The maximum mean angle error of these torque converters is 0.339°, and the minimum mean angle error is 0.020°.

For evaluating the proposed feature based pose determination approach, a general pose error $E$ is defined to estimate the experimental results of all of the 45 torque converters obtained by 10 model torque converters. Let $\bar{\epsilon}_{ij}$ represent the mean angle error of torque converter $j$ obtained by model torque converter $i$. The estimation error $E$ is computed by

$$E = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \bar{\epsilon}_{ij},$$

(5)

where $M$ and $N$ are the number of model torque converters and test torque converters. In our experiments, the estimation error $E$ is 0.263° for the 45 test torque converters and 10 model torque converters.

The experimental results show that 3D surface features of a local point can be well described using the proposed ADM descriptions in range images. In this automated assembling application, the feature points used to determine the pose are extracted according to the projected ADM descriptions. Accurate 3D orientation data of automobile parts are obtained by fitting the feature plane based on the refined feature points.

Table 1
The mean angle errors (°) of the torque converters in Fig. 12(a)

<table>
<thead>
<tr>
<th>Torque converter</th>
<th>T-1</th>
<th>T-2</th>
<th>T-3</th>
<th>T-4</th>
<th>T-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean angle error</td>
<td>0.02</td>
<td>0.339</td>
<td>0.245</td>
<td>0.297</td>
<td>0.078</td>
</tr>
</tbody>
</table>

Fig. 12. An experimental example displayed in intensity image: (a) a one-level stack with five torque converters; (b) a model torque converter with specified feature points.

Fig. 13. Final matched feature points of the torque converters shown in Fig. 11(a) obtained by the model torque converter; (a) T-1; (b) T-2; (c) T-3; (d) T-4; (e) T-5.
5. Conclusion

In 3D image processing, the range data of the environment are collected and analyzed for completing a specified vision task. In this paper, an introduction to a 3D computer vision system is presented. A 3D object pose estimation method is developed for an automated manufacturing assembly application. Using the traditional image processing methods, the target parts are segmented from the original range images. The center position is then computed through the circle Hough transform algorithm. For the 3D orientation estimation, a 3D free-form surface descriptor, Angle Distance Map, is computed through the circle Hough transform algorithm. For accurate orientation information is estimated according to the edges and critical points.

The experimental results show that accurate object pose is obtained according to the proposed 3D object pose estimation method.

References


Dongming Zhao is currently an associate professor in Electrical and Computer Engineering at the University of Michigan-Dearborn, MI. He received his BSE, MSE in electronic information engineering in Huazhong University of Science and Technology, Wuhan, China, and an MSEE from the University of Michigan, Ann Arbor, MI, and PhD from Rutgers, the State University of New Jersey, New Brunswick, NJ. His research includes 3D imaging, image processing, machine vision and pattern recognition, statistical modeling and pattern forecasting.

Songtao Li received his BSE and MSE in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, and PhD in pattern recognition and intelligence system from Tsinghua University, Beijing, China. He had been a visiting scholar in Electrical and Computer Engineering at the University of Michigan-Dearborn from 1999 to 2001. He is currently a research scientist at Cogent Systems, Inc., South Pasadena, CA. His research interests include image processing, machine vision, pattern recognition and computer graphics.