3D vision assisted flexible robotic assembly of machine components

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: OGUN, P. ... et al., 2015. 3D vision assisted flexible robotic assembly of machine components. IN: Proceedings of 2015 8th International Conference on Machine Vision (ICMV 2015), Barcelona, Spain, 19-21 November 2015 (Proceedings of SPIE, 9875, DOI: 10.1117/12.2229053).

Additional Information:

- © 2015 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

Metadata Record: [https://dspace.lboro.ac.uk/2134/25454](https://dspace.lboro.ac.uk/2134/25454)

Version: Published

Publisher: © SPIE

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: [https://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Please cite the published version.
3D Vision Assisted Flexible Robotic Assembly of Machine Components

Philips S. Ogun, Zahid Usman, Karthick Dharmaraj, Michael R. Jackson
EPSRC Centre for Innovative Manufacturing in Intelligent Automation, Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, UK

ABSTRACT

Robotic assembly systems either make use of expensive fixtures to hold components in predefined locations, or the poses of the components are determined using various machine vision techniques. Vision-guided assembly robots can handle subtle variations in geometries and poses of parts. Therefore, they provide greater flexibility than the use of fixtures. However, the currently established vision-guided assembly systems use 2D vision, which is limited to three degrees of freedom. The work reported in this paper is focused on flexible automated assembly of clearance fit machine components using 3D vision. The recognition and the estimation of the poses of the components are achieved by matching their CAD models with the acquired point cloud data of the scene. Experimental results obtained from a robot demonstrating the assembly of a set of rings on a shaft show that the developed system is not only reliable and accurate, but also fast enough for industrial deployment.

Keywords: 3D vision, robotics, automated assembly, object recognition, point cloud matching

1. INTRODUCTION

Industrial robots are widely used for automated assembly due to their repeatability and support for mass production [1, 2]. Currently, fixtures are used to place parts in predefined poses and the robots are operated through pre-programmed paths. So far, automation has been less attractive to small and medium enterprises (SMEs) because the costs of designing and fabricating fixtures are very high. Automation is currently economically viable for high volume and low mix production. Moreover, fixture-based systems operate blindly and they cannot handle subtle variations in part geometry and locations. Flexible and adaptive automation approaches are needed to eliminate the use of dedicated fixtures. In this case, a single robotic setup could be used to assemble different parts through software reconfiguration only.

Vision assisted fixture-less robotic assembly systems have already been developed and reported in several academic and technical white papers. However, most of the existing solutions are limited to three degrees of freedom because they can only be calibrated for a single plane and rotation [3, 4, 5, 6 and 7]. The vision challenge is simplified from three dimensional (3D) to two dimensional (2D) by placing components on flat worktables without any overlap.

The primary challenge in 3D vision assisted robotic assembly is recognising and estimating the full poses of objects within the robot workspace. A substantial number of 3D object recognition and pose estimation pipelines using point cloud data (PCD) have already been developed [8-13]. 3D feature matching techniques find correspondences between the point cloud of a rigid object and that of a scene containing the object. The goal is to recognise the object from a single viewpoint and then estimate its pose in the scene. Geometric shape descriptors are computed for every point or selected key points in the scene and then matched with those of the object’s in order to establish points’ correspondences. The shape descriptors are features that encode information about a point and its surrounding geometry. The performances of some of these algorithms are reported in [14, 15].

Despite the considerable number of algorithms already developed in the field of computer vision, application to real industrial assembly tasks is still in the early days. This is because the requirements in terms of stability, reliability and speed are much higher than in other applications such as mobile robotics. Also, there is a low margin for error as the assembly components would usually have sub millimeter tolerances. The working environment and the surface finish of the components could also introduce additional challenges. For example, machine vision systems work best under controlled lighting environments and imaging of objects that exhibit specular highlights (shiny) is still challenging. Therefore, the goal of this paper is not to create any new 3D object recognition and pose estimation algorithms, but to identify suitable existing technologies and use them to develop a flexible robotic assembly system that will meet general industrial needs and process requirements.

CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2229053

Proc. of SPIE Vol. 9875 98751O-1

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 03/02/2017 Terms of Use: http://spiedigitallibrary.org/ss/termsofuse.aspx
Assembly of rings on a shaft has been chosen in this paper because the assembling of components, such as gears and disks, on a shaft is one of the most common assembly tasks in many industrial sectors. One of the tasks to which 3D vision has been applied significantly is bin-picking [16, 17]. However, object pose estimation accuracy requirements in bin-picking are not as critical as in assembly. Also, most bin-picking applications do not require object classification since the bins contain objects of the same type.

There are two main aspects in the work reported in this paper. The first one is the 3D vision perception and the other is the integration of the vision system with the robot to perform an assembly task. The 3D vision perception involves sensor selection and the development of an object recognition and pose estimation pipeline that is accurate and fast enough for a potential industrial deployment.

2. ASSEMBLY COMPONENTS DESIGN

The shaft and the rings (Figure 1) used for the assembly task were designed to represent an industrial assembly challenge that requires 3D visual perception. The shaft has a tapered section on which three rings of gradually increasing diameters and with tapered bores are to be placed. A rectangular key feature was designed for the shaft and the rings in order to eliminate the rotational symmetry. The taper on the rings makes the vision challenge 3D because an upside down ring will not fit on the shaft. Therefore, the vision system must be able to recognise the correct face of the rings for them to fit properly on the shaft. The clearance between the rings and the shaft is 0.15 mm. Therefore, the overall grasping accuracy of the 3D vision system and the robot must fall within this range.

![Figure 1. CAD models of rings, shaft and their assembly](Image)

The vision-guided assembly robot setup consists of a six axis industrial robot (ABB IRB120), a 3D laser scanner and the assembly components. The robot system is shown in Figure 2.

![Figure 2. The vision-guided robotic assembly setup](Image)
3. 3D OBJECT RECOGNITION AND POSE ESTIMATION

3.1 Scene Acquisition and Object Matching

The field of computer vision has experienced great advancements over the years especially in 3D sensor technologies. There are many state-of-the-art 3D sensor technologies on the market. Classification of these sensors based on their working principles can be found in [18]. Sensor selection in machine vision is generally based on the application [19] but some of the factors that are usually considered are accuracy, resolution, depth of field and field of view.

Considering the tolerance between the shaft and the rings, the minimum sensor accuracy is 0.15mm. Also, the sensor must be robust to specular highlights because the components are made of anodised aluminium, which could exhibit specular highlights under uncontrolled lighting conditions. In order to select the most suitable sensor, some commercial sensors were evaluated. The details of the experimental results are beyond the scope this paper but it was found that imaging of relatively shiny objects still presents a technological challenge especially for area scan sensors operating within the visible light spectrum. So far, laser triangulation sensors are the only ones found to be suitable for imaging machine components in an unstructured environment.

The main requirement is that the vision system should be able to handle situations where the rings may be overlapping or placed one on top of the other in any order. Figure 3 shows examples of the various possible rings and shaft placements that the vision system should cope with. It should be able to recognise the topmost ring in the scene and also determine its pose. The scene is arranged so that the shaft can be picked first without being obstructed by the rings.

The object recognition and pose estimation are achieved by matching the CAD models of the objects with the point cloud data of the scene in which the objects are located. The scene is reconstructed by combining point cloud datasets captured from two sensor views. This is because a single view is not sufficient to capture enough details of the objects. Although a number of algorithms for estimating the relative transformation between two point cloud datasets exist, they are required only if such information cannot be obtained from external sources. Also, pairwise matching requires that there are significant overlapping regions between the datasets. Since the 3D laser sensor is mounted on the robot, the two point cloud sets were registered by using the known robot poses during the scanning stage. Figure 4 shows the union of two views of a scene containing the assembly components.
The object matching and pose estimation pipeline is based on the work published in [9, 10 and 11]. The processing pipeline for the object recognition and pose estimation is shown in Figure 5.

Feature descriptors are used to represent the local geometry around a point for 3D point cloud datasets. In this work, the feature descriptors are computed similarly to [9] but the correspondence group is performed using the Hough voting scheme described in [10 and 11] instead of the sample consensus based initial alignment algorithm used in [9]. The feature descriptors vector computed in [10] has four scalar values whereas the fast point feature histogram computed in [9] has three values. Tests have shown that the fourth feature characterising the Euclidean distance between points can be removed without any significant decrease in robustness and discriminately power [9]. The Hough voting scheme for correspondence grouping is found to be more robust than the sample consensus approach because multiple instances of the rings could be present in the scene. In the case of multiple instances, the sample consensus approach may fail to find any of the instances. The result of this combined feature descriptors computation and correspondence grouping algorithm is a faster and more robust object matching and pose estimation pipeline.

The first operation performed on the acquired point cloud is the removal of noisy data points caused by internal reflection from the shiny robot cell table. This is achieved through statistical analysis of points and their neighborhoods. The mean distances from a point and its k-nearest neighbours are computed. The points whose mean distances are outside a set interval are regarded as noise. Points in the robot workcell table plane are also removed based on a priori knowledge of the distance from the table plane to the sensor. The data is then resampled at a specified resolution before passing it to the feature matching block. The object models are transformed using the estimated poses and then overlaid on the scene point cloud. Figure 6 shows a sample of the registration results for overlapping rings.

Figure 5. Processing pipeline for object recognition and pose estimation

Figure 6. Registration results for overlapping rings
The number of data points for the objects at 1.0 mm, 1.5 mm and 1.7 mm voxel grid sampling distances are shown in Table 1. The minimum time it takes to find each object at various sampling distances when there are no other objects in the scene are shown in Table 2. The registration time could be significantly higher if there are other objects in the scene. The processing is done in C++ on an Intel Core i7 – 3770 3.4GHz, 16GB RAM machine.

<table>
<thead>
<tr>
<th>Object</th>
<th>1.0 mm</th>
<th>1.5 mm</th>
<th>1.7 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shaft</td>
<td>14762</td>
<td>6336</td>
<td>4776</td>
</tr>
<tr>
<td>Ring 1</td>
<td>5697</td>
<td>2371</td>
<td>1839</td>
</tr>
<tr>
<td>Ring 2</td>
<td>4432</td>
<td>1848</td>
<td>1465</td>
</tr>
<tr>
<td>Ring 3</td>
<td>3388</td>
<td>1463</td>
<td>1148</td>
</tr>
</tbody>
</table>

Table 1 No of data points at various sampling distances

<table>
<thead>
<tr>
<th>Object</th>
<th>1.0 mm</th>
<th>1.5 mm</th>
<th>1.7 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shaft</td>
<td>9.24 s</td>
<td>4.34 s</td>
<td>3.33 s</td>
</tr>
<tr>
<td>Ring 1</td>
<td>3.81 s</td>
<td>1.76 s</td>
<td>1.48 s</td>
</tr>
<tr>
<td>Ring 2</td>
<td>2.88 s</td>
<td>1.28 s</td>
<td>1.13 s</td>
</tr>
<tr>
<td>Ring 3</td>
<td>2.21 s</td>
<td>0.99 s</td>
<td>0.86 s</td>
</tr>
</tbody>
</table>

Table 2 Object pose estimation computation time at various sampling distances

As shown in Table 3, the number of data points affects the processing speed significantly. After several experimental runs, it was observed that at 1.7mm sampling distance the rings and shaft could be invariably identified from the point cloud within a minimum time. Therefore, an optimum sampling distance of 1.7 mm was used.

### 3.2 Matching Error Estimation

The matching accuracy is quantified by using a synthetic point cloud generated from the CAD model of the shaft. The point cloud is corrupted by removing some actual data points and also adding some randomly generated points. A plane is also added in order to simulate the points on the workcell table. A known transformation is then applied to the original CAD model before matching it with the synthetic scene data. The average poses estimation errors for 1.0 mm, 1.5mm and 1.7 mm sampling distances are shown in Table 3. The results in Table 3 show that the object matching error is very small. The inconsistently high errors in the $\alpha_z$ rotation angle are due to the shaft symmetry. The keyway features in the shaft and the rings are lost at the 1.7mm optimum sampling distance. Consequently, the registration could not establish a unique rotation angle around the objects’ axes of symmetry.

<table>
<thead>
<tr>
<th>Error</th>
<th>1.0 mm</th>
<th>1.5 mm</th>
<th>1.7 mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>0.01 mm</td>
<td>0.06 mm</td>
<td>0.07 mm</td>
</tr>
<tr>
<td>$y$</td>
<td>0.03 mm</td>
<td>0.05 mm</td>
<td>0.08 mm</td>
</tr>
<tr>
<td>$z$</td>
<td>0.02 mm</td>
<td>0.04 mm</td>
<td>0.06 mm</td>
</tr>
<tr>
<td>$\alpha_x$</td>
<td>0.02°</td>
<td>0.03°</td>
<td>0.05°</td>
</tr>
<tr>
<td>$\alpha_y$</td>
<td>0.03°</td>
<td>0.06°</td>
<td>0.07°</td>
</tr>
<tr>
<td>$\alpha_z$</td>
<td>48.15°</td>
<td>23.59°</td>
<td>67.34°</td>
</tr>
</tbody>
</table>

Table 3 Average pose estimation error

where

$x, y, z$ represent the translation vector

$\alpha_x, \alpha_y, \alpha_z$ are the Euler rotation angles
The pose estimated from the point cloud matching represents the transformation of the object in the sensor coordinate system. Hand-eye calibration was performed in order to determine the transformation between the sensor and the robot tool coordinate system so that pose of the object is known in robot coordinates. This enables the robot to grasp the object and move it accurately within its workspace as required.

3.3 Results Discussion

The overall grasping pose accuracy was measured by using the robot gripper to place the objects in a particular location and orientation. The poses of the robot at those points were then compared with the poses calculated from the overall vision system. The maximum overall grasping pose errors for each object are shown in Table 4. As seen in the table, the rotation angle around the \( z \) axis is not unique due to the symmetry of the objects as mentioned earlier in section 3.2.

<table>
<thead>
<tr>
<th></th>
<th>Ring 1</th>
<th>Ring 2</th>
<th>Ring 3</th>
<th>Shaft</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x )</td>
<td>1.13 mm</td>
<td>1.23 mm</td>
<td>1.02 mm</td>
<td>1.26 mm</td>
</tr>
<tr>
<td>( y )</td>
<td>1.15 mm</td>
<td>1.26 mm</td>
<td>1.05 mm</td>
<td>1.06 mm</td>
</tr>
<tr>
<td>( z )</td>
<td>0.62 mm</td>
<td>0.52 mm</td>
<td>0.58 mm</td>
<td>0.43 mm</td>
</tr>
<tr>
<td>( \alpha_x )</td>
<td>0.14°</td>
<td>0.11°</td>
<td>0.13°</td>
<td>0.15°</td>
</tr>
<tr>
<td>( \alpha_y )</td>
<td>0.16°</td>
<td>0.15°</td>
<td>0.12°</td>
<td>0.16°</td>
</tr>
<tr>
<td>( \alpha_z )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Point cloud object matching algorithms are computationally intensive because of their complexities and the immense amount of data that is being processed. Therefore, the size of the data was reduced by down sampling, which resulted in the loss of the keyway features. A denser point cloud (0.5 mm sampling distance) is required to retain the keyway features but at the expense of processing speed, which is unacceptable for real industrial applications. It can also be seen that the translation errors are greater than the 0.15 mm tolerance between the shaft and the rings. This means that the overall vision system on its own is not able to meet the accuracy requirements. The errors are compensated for by designing the gripper fingers such that the objects are re-centred in the jaws as the grippers are closed.

4. CONCLUSION

This paper presents the development of a 3D vision-guided robot for assembling randomly placed and overlapping mechanical components. Since most machine elements are metallic, it was necessary to select a 3D sensor that is robust to specular highlights. The main challenges in this research were to recognise the components and also determine their poses in the scene accurately and fast enough for real industrial applications. These were achieved by making effective use of an existing point cloud matching algorithm to match the CAD models of the components with the acquired point cloud data of the scene. The maximum achieved grasping errors in translation and rotation were 1.26 mm and 0.16° respectively.

Although the achieved object pose estimation accuracy is not high enough for the chosen assembly task, the misalignment is compensated for by using grippers to re-align the parts as they are grasped. Further improvements are required for tighter tolerance fits such as transition and interference fits. In order to achieve this, a combination of vision sensing and force-torque control will be required. This will provide robots with both the visual information feedback and the manual dexterity used by a skilled worker to perform such complex assembly tasks.
REFERENCES


