



Autonomous metrology for robot mounted 3D vision systems

Peter Kinnell, Tom Rymer, John Hodgson, Laura Justham, Mike Jackson

EPSRC Centre for Innovative Manufacturing in Intelligent Automation, Loughborough University, Loughborough, UK

Submitted by Lin Li

Using a metrology system simulation approach, an algorithm is presented to determine the best position for a robot mounted 3D vision system. Point cloud data is simulated, taking into account sensor performance, to create a ranked list of the best camera positions. These can be used by a robot to autonomously determine the most advantageous camera position for locating a target object. The algorithm is applied to an Ensenso active stereo 3D camera. Results show that when used in combination with a RANSAC object recognition algorithm, it increased positional precision by two orders of magnitude, from worst to best case.

Metrology, Cognitive Robotics, 3D Image Processing

1. Introduction

3D vision systems are used in a wide range of manufacturing tasks, such as locating tools and parts, inspecting part geometry and checking alignment in assemblies. They are also widely used to give robots the ability to locate objects in their environment; this is useful for pick and place operations, assembly jobs and weld path following [1][2][3].

When setting up a 3D vision system, selecting the camera pose that gives the most useful data is a significant challenge. This is because the quality and quantity of captured data is dependent on the pose of the vision system relative to the target object. This is particularly true for shiny surfaces, which can adversely scatter the illumination provided by the vision system such that it cannot be detected by cameras in the system [2][4]. To mitigate this problem, an expert user might be able to specify viewing position that maximises the data captured from an object. However, considering only the quantity of measured data is not sufficient to select the optimum viewing position; for true optimisation the intended use of the data must also be considered. To determine how good a pose is, the user is required to setup the vision system, collect data, process it and evaluate the performance of each pose; continuing to try new configurations until a suitable setup is found. Optimising the setup of vision systems can be extremely time consuming and heavily reliant on the expertise of the human operator [5][6].

For applications in which the pose of target object relative to the vision system are well defined, such as car body parts on an assembly line, the optimisation only needs to be competed once; so the time taken to optimise the vision system can be easily justified. However, for applications that are subject to large inherent variation, it might be necessary to undertake the optimisation for a wide range of possible situations; this is true in challenging pick and place operations that require components, presented in a random orientation, to be located and manipulated by a robot.

In this paper the challenge of accurately locating randomly presented parts, using a 3D vision system mounted to the end of a robot arm, is considered. This setup might be deployed in applications such as robot bin picking or other pick and place

tasks [7]. The aim of the vision system is to allow the robot to locate an object, of known nominal geometry, and determine its position and orientation relative to the robot. Mounting the vision system on the robot gives the advantage that, in response to variation of the target objects, the vision system can be moved by the robot to an optimal viewing position. To make this possible the robot needs the capability to autonomously select good camera positions, removing the need for input from a human expert.

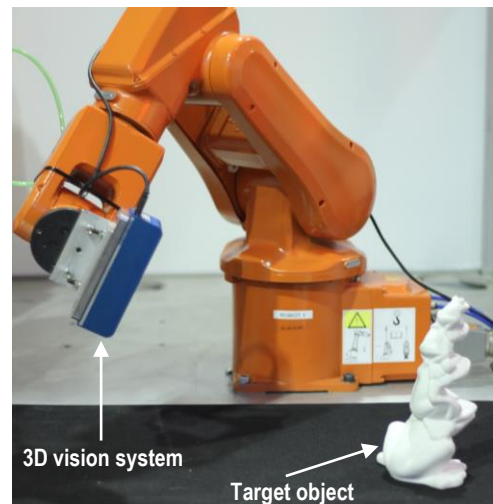


Figure 1. Photograph of the 3D vision system mounted to an industrial robot and positioned to view a target object.

To give the robot the necessary intelligence to select good camera positions, a metrology simulation approach is proposed. To demonstrate this approach, a simulation of the physical metrology system was created to support the positioning of an Ensenso active stereo camera, which is mounted to a robot arm, as shown in Figure 1. The approach taken in this work is similar in essence to the concepts of virtual metrology, or model based metrology, that allow the performance of a metrology system to be predicted and optimised prior to undertaking the real measurement [8]. It is also compatible with the concept of Cyber-physical systems which are being increasingly applied to a wide

range or manufacturing activities [9]. The aim of the system created in this work is to provide a ranked list of viewing angles for a given object, from which the 3D vision system is able to collect the most information-rich data to give the best estimation of the object's position and orientation.

This paper is structured as follows; first, a top level description of the metrology system simulation approach is given. This is followed by a detailed description covering how the system is able to autonomously select optimum camera poses. Finally, to demonstrate the performance of the autonomous metrology system, the predicted best and worst camera positions for a test object are verified by comparison to experimental data.

2. The metrology system simulation approach

The metrology system simulation approach combines information regarding the configuration of the physical measurement task with a computer simulation of the measurement process; in this case the measurement process is a 3D vision sensor that is used to determine the location and orientation of an object. The computer simulation gives the opportunity to quickly test the expected quality of the data collected from multiple camera positions. This could be done either in real time, or in advance by considering a large range of possible scenarios. The results of the simulation then provide the knowledge needed to choose the best camera positions for the actual physical measurement process.

The metrology system simulation created in this work has the following key elements: a meshed 3D CAD (computer aided design) model of the target object; an empirical model of performance for the 3D vision system; and a method to simulate, and assess the quality of data generated by the vision system. First, the 3D CAD model is analysed to determine the most information-rich positions on its surface. Then a simulation of the measurement process is undertaken for a set of camera viewing positions. The simulation requires the 3D CAD model of the target object and the performance model of the 3D vision system. Each of the simulated camera viewing positions is then assigned a score. The score is based on the probability that each of the information-rich surface points, previously identified, can be measured by the 3D vision system.

3. Autonomous camera pose selection

3.1. Sensor performance simulation

To simulate the data generated by the modelled sensor the following procedure is followed. First, a 3D model of the target object is created in the form of a triangulated surface mesh. This is represented as a set of triangular faces with associated points at their vertices. If the target object is highly sculpted, it is preferable for the mesh triangles to be of similar size, and to have an individual surface area at least smaller than 1% of the total surface area. The camera's field of view is then modelled in orientations relative to the target object. To model the field of view, a pinhole camera model is used with the focal point of the camera, and the associated imaging plane, positioned in a set of poses relative to the object. For each camera pose back-face culling removes the mesh faces of the 3D model that are obscured from the camera's field of view. Ray tracing is then performed to determine the surface points imaged by each pixel of the camera. The result of this step is a set of surface points that could be measured under perfect conditions, with no corruption of the data due to noise or loss of data due to imperfect light scattering at the target surface. Then to account for measurement imperfections, the expected missing data on each triangular surface facet is determined using a performance model of the 3D

vision system. The performance model specifies the probability that data is recovered as a function of the orientation and position of each facet.

3.2. Empirical model of sensor performance

The ability to model the expected performance of a 3D vision system is critical to the metrology system simulation approach. A suitable model must provide information about the likelihood of obtaining data from all parts on the target object, and it must also provide an estimation of the quality of measured data. Currently, manufacturers of 3D vision systems generally provide only very basic performance data. As there are no standards to cover expected performance descriptions of 3D vision systems, on non-ideal surfaces, the data provided are limited to basic system properties, such as the spatial resolution of points or working distance. Crucially, there is typically no objective description of how an object's orientation or surface finish will affect performance [10].

Recent work by Hodgson et al. highlighted the inadequacy of standard evaluation methods for 3D vision systems, and defined a new set of metrics to describe their performance [10]. In addition to the new metrics, the work also presents a clear methodology for the empirical characterisation of 3D vision systems, with an associated performance model. The proposed model is based on the empirical evaluation of a vision system when measuring a flat target surface over a range of orientations, which includes angular position as well as distance from the sensor.

To set up the metrology system simulation in this work, the empirical model of the vision system's performance was established by following the methodology defined by Hodgson et al. [10]. To represent the surface finish at the intended target object, which was made of a non-glazed white ceramic material, a matt white 60 mm x 60 mm square test sample was created. The test sample was then positioned in front of an Ensenso active stereo imaging system on a pan tilt stage, at a distance of 500 mm. Using the pan tilt stage the sample was moved through a set of 1008 orientations, covering a range of 360° in pan and 55° tilt. At each orientation the test sample was measured using the Ensenso vision system. The measured data was then analysed to determine the expected quality and quantity of data for each orientation. From the collected data a model that predicts the surface point density and noise expected for a target surface, as a function of the position and orientation of the surface relative to the imaging system, was created, again using the method described by Hodgson et al. [10].

3.3. Ranking algorithm for camera viewing positions

The problem of camera positioning to gain the best coverage of a subject was first described by Chvatal [11] in the classic art-gallery problem. For 3D point cloud data, approaches to the view planning problem have typically focused on determining the visibility and accessibility of a set of surface points, or region of interest on an object [5][12][13]. In this work, the aim of the position ranking algorithm is to score each selected viewpoint based on the quantity and quality of the data points captured; with point quality being dependent on how information-rich the point is, and the probability that it can be measured. The algorithm [13] first considers the ability to visualise information-rich surface points, however in contrast to previous work it also considers the expected probability that these points will be captured. This is achieved due to the use of the 3D vision system performance model, and it is a step necessary to ensure optimum viewing positions are as robust as possible, especially on challenging surfaces.

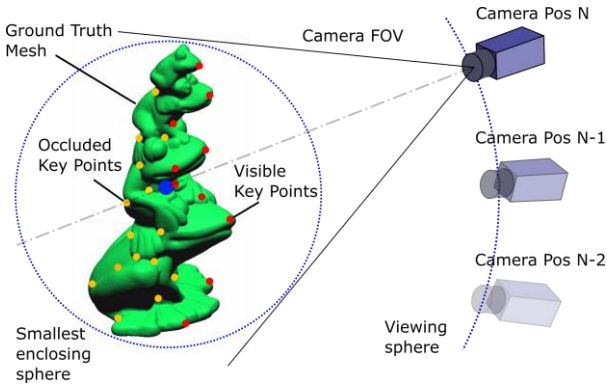


Figure 2. Identification of visible model key points per camera view point.

When locating objects in 3D space, using point cloud data, the most common algorithms are based on a key point matching approach. Typically, a random sample consensus (RANSAC) type algorithm is used to search for, and match, key point pairs that are present in both the 3D model of the target part, as well as the measured point cloud data. When using a RANSAC algorithm it is not sensible to consider all measured points, as this can lead to unacceptably long computation times. Therefore there have been a number of approaches presented to automatically select a subset of the most useful, information-rich, key points. In this work, Intrinsic Shape Signatures (ISS) presented by Zhong et al. [14] are used to determine these information-rich points.

To rank a set of viewpoints the following process is followed. First, a sphere is tightly fitted to the object scene using Gartner's smallest enclosing ball approach [15]. A working distance from the object surface to the camera is added to this radius to produce another sphere. This larger sphere is uniformly sampled with N points which will act as camera origin locations for viewpoint optimisation. The smallest enclosing sphere and multiple camera positions surrounding the target object are illustrated in Figure 2. These camera origin locations position the camera line of sight directly at the centre of the tightly fitting sphere.

For each camera origin location a simulated point cloud is generated, following the approach detailed in section 3.1. This determines key points that are visible, shown in red on Figure 2, and those that are occluded from the vision system shown in yellow on Figure 2.

Upon completion of all iterations, the key points in each view are each then analysed. Figure 3 illustrates a key point that is surrounded by a set of four triangular surface facets. To determine the probability that the key point will be measured, the vision system simulation is used to assess the likelihood that points will be recovered from each facet. This process is represented in Figure 3 by the link from a facet n_4 to the graph giving the probability that a point is recovered as a function of facet orientation. Each facet surrounding the key point is considered to see if data points will be recovered from it. Then from each facet where data points will be recorded, the maximum probability value, λ is found and assigned to the key point. The sum of all the key point λ values for each view is then calculated and stored against the camera position for that view. This score is therefore a function of the number of key points theoretically visible, but also as λ itself is the probability of the key point being captured by the vision system, a high score is also indicative of a set of key points that can be easily measured with the vision system. Therefore the combination of these two factors provides a scoring system that considers quantity as well as quality of the information-rich key points which are visible.

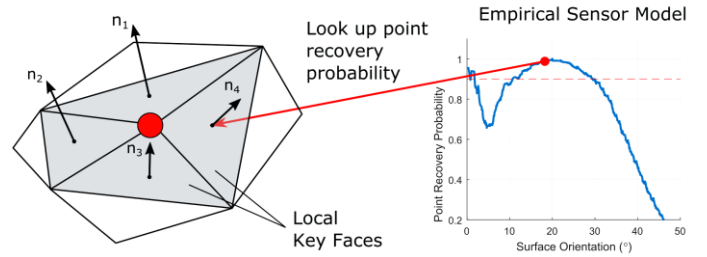


Figure 3. Scoring key points based on the probability of capturing data on local faces using an empirically determined sensor model.

4. Experimental evaluation procedure

To verify the output of the position optimisation algorithm an experiment was designed to capture point clouds from a range of sensor orientations relative to the object. The viewpoint position ranking algorithm was run sampling a set of 200 evenly distributed viewpoints. Using the output of the algorithm, a set of eight viewpoints was selected; these included the four highest, and four lowest scoring viewpoints. The Ensenso active stereo camera was then positioned at a stand-off distance of 0.5 m, which is equal to the stand-off distance of the considered viewpoints, from the target object. The pose of the Ensenso vision system with the target object were manipulated to allow each of the eight selected viewing positions to be achieved. During measurements, ambient lighting conditions were controlled, with no direct lighting above or surrounding the experimental setup.

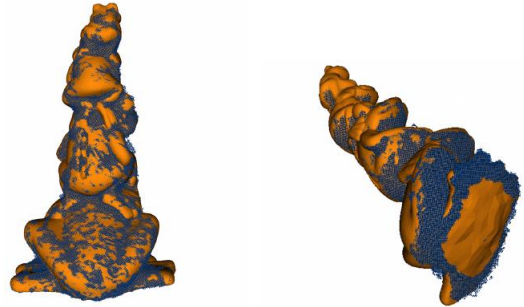


Figure 4. The best and worst viewpoints are shown on the left and right respectively.

The target object, a matt white sculpture of a stack of frogs, was chosen so the performance of the algorithm can be demonstrated on an object with significant geometric complexity, which makes it hard for a human expert to judge which viewpoints are best. A 3D mesh model representative of this object is required to run the position optimisation simulation. As a pre-existing CAD model was not available for the chosen target object, an approximated 3D model was reverse engineered using a Roland LPX-250 3D laser scanner to produce the 3D mesh file required. The 3D mesh model can be seen in Figure 2. Figure 1 shows the Ensenso camera mounted on an industrial robot to, and Figure 4 shows the best and worst ranked views of the target object.

To assess the usefulness of the point clouds captured from each view point, the data was processed using a RANSAC algorithm designed to determine object location in the sensors coordinate system. The RANSAC variant object recognition algorithm used was created by Papazov and Burschka [16]. This algorithm takes a mesh model of the target part, selects key points and tries to match these to the sensor data. The output from the algorithm is a homogeneous rigid transformation matrix giving the position of the measured object in the Ensenso camera's coordinate system.

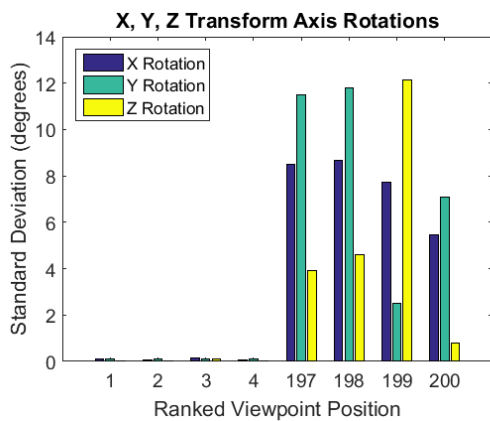


Figure 5. Standard deviation of the transform axis rotations

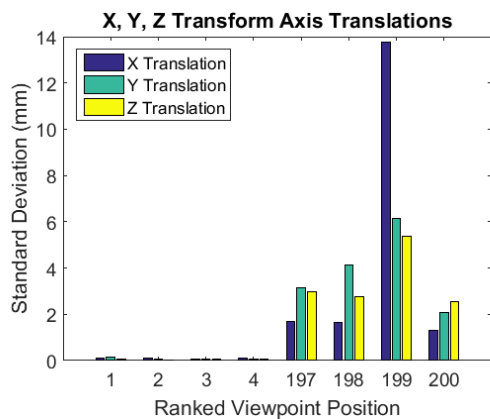


Figure 6. Standard deviation of the transform axis translations

5. Results and discussion

To verify the performance of the viewpoint position ranking algorithm, errors in the absolute position of the object were not considered. This is because these errors are a function of the uncertainty generated by both the 3D vision system and the RANSAC algorithm; instead errors generated by the RANSAC algorithm as a function of the quality of the measured data were the main concern. The random sampling component of the RANSAC algorithm means that it provides a different result each time it is run. Therefore to assess how good the data collected for each viewpoint was, the repeatability of the algorithm for a single set of data from each viewpoint was tested; based on the assumption that viewing positions with a large quantity of information-rich points lead to a more repeatable result from the RANSAC algorithm. For each data set collected from the eight viewing positions, the RANSAC algorithm was run 50 times. In this way the variation of the target object's position estimation, due to the quality of the measured data, was considered. Variation was assessed in terms of both spatial translations and rotations. Spatial translations were considered in three orthogonal axes (x, y and z), and rotations were considered about these three axes. The standard deviation of rotations and translations are shown on the charts in Figure 5 and Figure 6 respectively; these charts illustrate the large positional variations that can be expected with data that is collected from the least optimal viewing positions, ranked 197 to 200. For the worst ranked positions, translation standard deviations up to as much as 14 mm were found, and rotational standard deviations up to 12 degrees were found. In contrast, for the best ranked viewpoints all the translation standard deviations were less than 0.15 mm and all the rotational standard deviations

were less than 0.14 degrees; therefore, by selecting the best ranked viewing positions, it is possible to increase the precision of object location by a factor of 85 for rotational errors and 93 for translational errors.

6. Conclusion

This work has demonstrated the benefits of using a metrology system simulation approach. A viewpoint selection algorithm was created to support the process of using a 3D vision system mounted to a robot to locate target objects using point cloud data. The algorithm creates a ranked list of useful viewpoints for locating a given object, and therefore provides the means for autonomous positioning of the camera by the robot. This results in a worst to best case increase in measurement precision of almost two orders of magnitude.

7. Acknowledgements

This work was funded by UK Engineering and Physical Research Council (grants EP/L01498X/1 and EP/1033467/1). Thanks also to Helen Curtis for proof reading and Professor Lin Li of The University of Manchester, for support of this work.

References

- Ahmad, R., & Plapper, P. (2015). Generation of safe tool-path for 2.5D milling/drilling machine-tool using 3D ToF sensor. *CIRP Journal of Manufacturing Science and Technology*, 10, 84–91.
- Bi, Z. M., & Wang, L. (2010). Advances in 3D data acquisition and processing for industrial applications. *Robotics and Computer-Integrated Manufacturing*, 26(5), 403–413. doi:http://dx.doi.org/10.1016/j.rcim.2010.03.003
- Michalos, G., Makris, S., Papakostas, N., Mourtzi, D., & Chryssolouris, G. (n.d.). Automotive assembly technologies review: challenges and outlook for a flexible and adaptive approach. *CIRP Journal of Manufacturing Science and Technology*, 2, 81–91. doi:10.1016/j.cirpj.2009.12.001
- Hodgson, J. R., Kinnell, P., Justham, L., & Jackson, M. R. (2015). Characterizing the influence of surface roughness and inclination on 3D vision sensor performance. In *Proceedings of SPIE - The International Society for Optical Engineering* (Vol. 9875). doi:10.1117/12.2228826
- Scott, W. R., Roth, G., & Rivest, J.-F. (2003). View planning for automated three-dimensional object reconstruction and inspection. *ACM Computing Surveys*, 35(1), 64–96. doi:10.1145/641865.641868
- Mavrinac, A., Chen, X., & Alarcon-Herrera, J. L. (2015). Semiautomatic Model-Based View Planning for Active Triangulation 3-D Inspection Systems. *IEEE/ASME Transactions on Mechatronics*, 20(2), 799–811. doi:10.1109/TMECH.2014.2318729
- Pérez, L., Rodríguez, Í., Rodríguez, N., Usamentiaga, R., & García, D. (2016). Robot Guidance Using Machine Vision Techniques in Industrial Environments: A Comparative Review. *Sensors*, 16(3), 335.
- Schmitt, R. H., Peterek, M., Morse, E., Knapp, W., Galetto, M., Härtig, F., ... Estler, W. T. (2016). Advances in Large-Scale Metrology – Review and future trends. *CIRP Annals - Manufacturing Technology*, 65(2), 643–665. doi:10.1016/j.cirp.2016.05.002
- Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., ... Ueda, K. (2016). Cyber-physical systems in manufacturing. *CIRP Annals - Manufacturing Technology*, 65(2), 621–641.
- Hodgson, J. R., Kinnell, P., Justham, L., Lohse, N., & Jackson, M. R. (2017). Novel metrics and methodology for the characterisation of 3D imaging systems. *Optics and Lasers in Engineering*, 91, 169–177. doi:http://dx.doi.org/10.1016/j.optlaseng.2016.11.007
- Chvatal, V. (1975). A combinatorial theorem in plane geometry. *Journal of Combinatorial Theory, Series B*, 18(1), 39–41.
- Raffaelli, R., Mengoni, M., Germani, M., & Mandorli, F. (2013). Off-line view planning for the inspection of mechanical parts. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 7(1), 1–12.
- Leifman, G., Shtrom, E., & Tal, A. (2012). Surface regions of interest for viewpoint selection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 414–421). IEEE.
- Zhong, Y. (2009). Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on* (pp. 689–696). IEEE.
- Gärtner, B. (1999). Fast and Robust Smallest Enclosing Balls (pp. 325–338). Springer, Berlin, Heidelberg. doi:10.1007/3-540-48481-7_29
- Papazov, C., & Burschka, D. (2010). An efficient ransac for 3d object recognition in noisy and occluded scenes. In *Asian Conference on Computer Vision* (pp. 135–148). Springer.