## NISTIR 7926

# Review of Research on the Ground Truth Systems for Evaluating Part Identification Systems

Afzal Godil
Roger Eastman
Tsai Hong Hong
*Intelligent Systems Division*
*Engineering Laboratory*

http://dx.doi.org/10.6028/NIST.IR.7926

**National Institute of Standards and Technology**
U.S. Department of Commerce

# NISTIR 7926

# Review of Research on the Ground Truth Systems for Evaluating Part Identification Systems

Afzal Godil
Roger Eastman
Tsai Hong Hong
*Intelligent Systems Division*
*Engineering Laboratory*

http://dx.doi.org/10.6028/NIST.IR.7926

April 2013

U.S. Department of Commerce
*Rebecca Blank, Acting Secretary*

National Institute of Standards and Technology
*Patrick D. Gallagher, Under Secretary of Commerce for Standards and Technology and Director*

## ACKNOWLEDGEMENTS

The authors wish to give special thanks to Stacy Bruss of the NIST Research Library who provided the initial set of the reference papers.

## 1. Abstract:

In this report we discuss different types of potential ground-truth systems that could be used to test part identification systems for manufacturing assembly applications. We discuss four main ways of acquiring ground truth for evaluating part recognition/identification systems: 1) synthetic and physics based simulation, 2) manual annotation or labeling of ground-truth systems, 3) platform-based ground-truth systems, and 4) physically sensed ground-truth systems. We discuss previous efforts and discuss the different physical quantities used for ground-truth systems. In this report, we focus on static parts. The sizes of parts are in the range of 30 cm x 30 cm x 30 cm volume or smaller.

## 2. Introduction

The goal of parts recognition/identification is to correctly identify parts that are present in a 3D scene, usually in Red Green Blue (RGB) color and/or depth/range images. This is a challenging task especially since the scene may be cluttered, the parts in the scene may occlude each other, there could be illumination variation, and the material, reflectivity, and surface properties of parts can make it hard for perception systems to recognize or identify parts in the scene.

For our purposes, we define “identification” as identifying one of N types of parts, with each type defined by an exemplar object. An example would be identifying which specific type of bolt, or flange, or similar part is viewed. We do not attempt to classify a part as a bolt, or a cup, or as a member of a general category; nor do we attempt to identify the object as a specific instance, or individual, of a class. Some exemplar parts used by an identification system are shown in Figure 1.

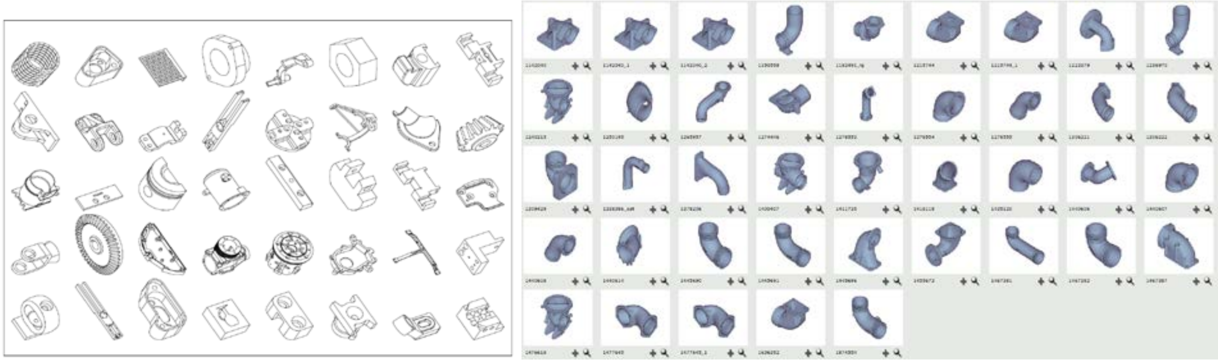


Figure 1. Examples of possible parts from the National Design Repository [1] (Left image) and the Purdue Engineering Shape Benchmark [2] (Right image).

We intend for this problem to represent one faced by an automated robotic system that needs to pick up a specific part and place it in an assembly. The specific part may one of many mixed parts, so the system must identify each as it appears. The parts may be presented to the robot on a conveyer, in a bin, or in some other way. We assume the part may be in an arbitrary 3D position and not rigidly fixed. We allow the part to be in any of the possible stable poses on a flat surface or loosely held in a kit. This means that there could be some *a priori* constraints on position or pose. Our goal is to find ways to evaluate identification systems for effectiveness and performance so manufacturers that use them can have confidence that they will meet operational needs.

To evaluate identification systems, we could conduct real-time tests in which the system is presented with a series or set of objects under preset conditions. The system would report the identity of each and we would compare the results to ground truth. For simple test scenarios, or for offline datasets, the experimenter places known objects in the scene and has ground truth by definition. However, we want to consider more complex scenarios in which the identification system under test is close to operational deployment. In these scenarios, the system undergoes strenuous, continuous tests where it has to identify long sequences of parts presented in a possibly automated method. Under these conditions it may be tedious or difficult for the experimenter to track part identity, and an automated ground truth system would assist.

A ground truth system for this purpose needs to be able to identify all of the parts placed in the test area, possibly labeling each with a location. The ground truth system should be much more accurate compared to the system under test. Additionally, the ground truth system should work under realistic test and environmental conditions. Therefore, we need to take potential variables into account including the amount and combination of occlusion and clutter in the scene, lighting, background radio and magnetic interference, nature and appearance of the parts, and the scenario by which parts are presented. If there are multiple parts in a scene, the ground truth system may need to report suitably accurate locations, which may be in the world or sensor coordinate system depending on how the system under test reports location.

In this report, we focus on static parts of sizes in the range of 30 cm x 30 cm x 30 cm volume or smaller, and consider tests where the parts are presented on a horizontal flat surface (shown in Figure 3) of size 150 cm x 150 cm. We may place a mount, bin, or compartmented box on the platform to hold parts. A few of the possible parts used for possible experiments are shown in Figure 2, and possible experiments are shown in Figures 3 to 8. These experiments include presenting:

1. One part at a time in the center of the test surface (Figure 3)
2. One part at a time at different locations on the test surface (Figure 4)
3. Few parts at different locations on the test surface with no occlusion and no change in elevation (Figure 5)
4. Few parts at different locations on the test surface with varying levels of occlusion and clutter and parts may have minor changes in elevation (Figure 6)
5. Lots of parts together on the test surface with high level of occlusion and clutter and change in elevation (Figure 7.)
6. Parts in a Kit box for kitting applications (Figure 8.). In a kitting type application the localization information could help identify parts.

The ground truth requirements for each experiment are different. For most, the location information is useful, except experiment 1 where only identity is important. In the case of experiment 5, the ground truth systems and the system under test may need to provide each part's location with relatively fine resolution.

Figure 2. Shows Possible parts for experiments.

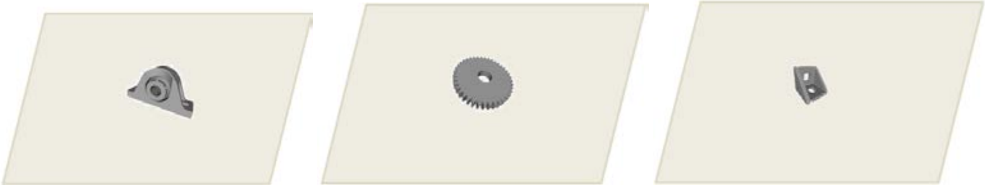


Figure 3. One part at a time in the center of the test surface.

Figure 4. One part at a time at different locations on the test surface.

Figure 5. Few parts at different locations on the test surface with no occlusion or very little occlusion.

Figure 6. Few parts at different locations on the test surface with varying levels of occlusion and clutter

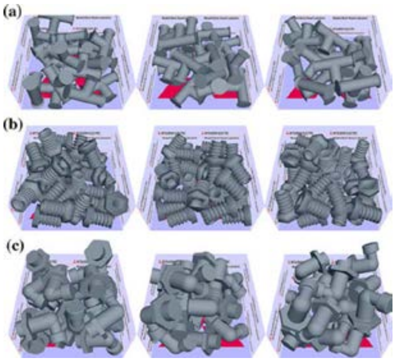


Figure 7. Lots of parts together on the test surface with high level of occlusion and clutter.

Figure 8. Parts in a Kit box for a kitting application.

The identification system under test may be based on one of many different technologies, or a combination, which may overlap with the technologies used for creating ground truth. We assume remote sensing technologies are non-contact sensors, such as 3D vision, Radio-frequency identification (RFID), or barcodes that work by sensing parts from a small to moderate distance. Any technology should be flexible and should work under unstructured conditions where parts do not have to be in fixed orientations or positions. A default part identification system for our study has a fixed sensor looking at the test surface. For reporting the data, we can assume one of the following approaches:

1. The system uses a proprietary sensor and algorithms, whereby we have no access to the sensor data. The lack of access to internal data means extra care must be taken in synchronizing time stamps and calibrating external coordinate systems. If we have no access to the raw data stream, we may have difficulty using some ground truth approaches, such as hand annotation or physics-based simulation.
2. The system uses a standard sensor and open algorithm, for which we have access to the sensor data. These systems should be easier to use with all forms of ground truth.

Since these are controlled test conditions, we will consider ground truth systems that use optical or radio tags attached to each part. On the other hand, for the systems under test, the use of tags may or may not be allowed, depending on the nature of the system and its expected application. A complication arises if the tags used by a ground truth system can be "hijacked" by a system under test – highly visible optical ground truth tags could be visible and used by the system under test, and similarly for radio frequency tags. However, the tags may not lead to an unwanted interaction – a barcode tag ground truth system may not interact with a shape-based system. It may be useful to have multiple options for ground truth, so one can be chosen that does not interact with the system under test.

In the next section, we will discuss different ways of acquiring ground truth. We also present a summary of the advantages, disadvantages, and limitations for each approach.

## 3. Ground-truth systems

We organize the approaches for ground-truth systems for evaluating part recognition/identification systems into four categories:

1) Synthetic and physics-based simulation ground-truth systems
2) Annotation/label based ground-truth systems
3) Platform-based ground-truth systems
4) Physically-sensed ground-truth systems

The following subsections discuss each in turn, with emphasis on the ground-truth systems that we conclude are most appropriate for our test procedures.

### 3.1. Synthetic and Physics Based Simulation Ground Truth Systems

We can simulate the virtual environment by placing one or more parts at different locations in the scene and/or by physically simulating their behavior. Realistic synthetic images can then be rendered with real-time or offline rendering [3]. We then know the identity, the location, and illumination variation and material and surface properties of every part and can also calculate the amount of occlusion and clutter in the scene. On the other hand, sensor data generated from a simulation does not exactly match real world sensor data despite advances in algorithms for realistic graphics rendering. The performance of parts recognition on synthetic images will differ from that on real images because sensor noise types and levels are not yet modeled with sufficient fidelity. In their paper, English et al. [4] showed that for verification and validation processes, the best solution to the evaluation problem may be through synthetic data generated by a physics-based simulation. To evaluate a pose estimation algorithm applied to range data in a bin picking algorithm, Park et al. [5] generated synthetic images by dropping a large number of parts with a physics-based simulation. The authors used the simulated data both to generate *a priori* statistics for likely part orientation and to evaluate their pose algorithm. They also used data from real world experiments in their evaluation, but with limited metrics since they had no ground truth.

Physics-based simulations can be applied to any scenarios where suitably accurate physical models exist. There are two ways in which it can have an advantage over real sensors and equipment in real environments:

1) Usually there is a cost saving associated with physics-based simulation, since there is less need for expensive sensors and equipment for experiments; initially many problems can be modeled and simulated. However, effective simulations can be expensive to create – the payoff may come with standardized simulation systems that support many scenarios.
2) Complex scenes with many parts including occlusion and clutter, as in the case of bin picking, can be simulated by dropping multiple parts randomly, whereas it might be more difficult and expensive to capture the ground truth locations and poses of all these parts in a real environment, as shown in Figure 9.

As a disadvantage, physics-based modeling can be limited in producing accurate sensor data, since those models are not yet mature.

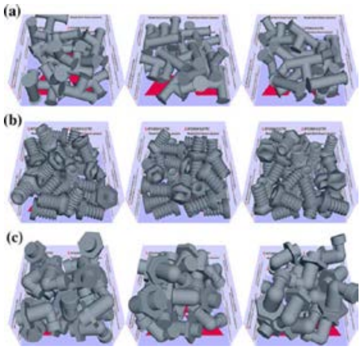
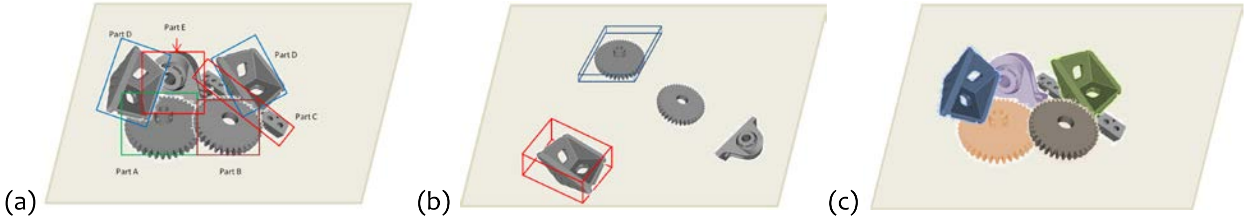


Figure 9. Synthetically generated scene images with physics-based simulation and modeling. Three bins with: a) 20 T-pipe models, b) 30 bolt models, c) 20 elbow pipe models in each bin
[Permission granted by the author [5] to use the Figure]

## 3.2. Annotation/Label Based Ground-Truth Systems

One popular way to create ground truth for evaluating part recognition/identification and tracking systems is by human annotation of images (including video, depth, or combined images). One of the commonly used desktop tools for image/video annotation is VIPER-GT [6]. It is possible to annotate images by drawing 2D or 3D bounding boxes depending on the sensor data, as shown in Figure 10a-b, which the users have annotated by drawing a 2D or 3D bounding box around a part indicating its identity. In the case of 3D point cloud data, as shown Figure 10c, 3D points associated with each part can be manually colored differently based on their identity.



(a) (b) (c)

Figure 10a. Shows 2D annotation of parts by drawing the outlined by a 2D bounding box. Figure 14 b. shows a 3D annotation of parts by drawing 3D bounding boxes. Figure 3. In case of 3D point cloud data, 3d points associated with each part are different color based on the identity.

LabelMe [7] is a very popular web based tool for image annotation and has been used for creating very large imaging benchmarks [28], including for the Pascal visual object classes challenge [26] and TinyImages [27]. LabelMe allows the annotation boundaries to be drawn along the actual boundaries of the object instead of drawing bounding boxes as in other annotation tools. Crowd-sourced annotation of images and video can be accomplished at low cost through platforms such as Amazon Mechanical Turk (MTurk) [29]. This has revolutionized the annotation of static and dynamic image datasets and has led to massive image and video datasets. Finally, in the KITTI Vision Benchmark Suite for Autonomous Driving [8], the 3D laser range data were annotated by drawing 3D bounding boxes around cars in the data.

In case we have 3D range data of a scene, more accurate localization can be obtained by manually registering the parts in the scene with the corresponding complete parts. This can be done by selecting corresponding points, or by initializing Iterative Closest Points (ICP) for finer registration.

Annotation-based approaches are typically applied to test areas/scenes where an area is sensed by a sensor whose data is suitable for human interpretation. Annotation-based approaches have the following advantages:

a) Complex scenes can be annotated by hand in cases where effective ground truth systems do not exist.
b) The software is often free, allowing a low-cost entry into the project.
c) The resulting annotated data supports analysis by multiple groups using multiple algorithms, so repeatability is good and cross-comparisons can be easily done.

Disadvantages include the labor cost of performing annotation, the variable and often unknown accuracy and reliability of the labels (identifications), and the fact that the annotations are mainly based on the images and recorded in sensor-based coordinates instead of 3D world coordinates.

## 3.3. Platform-based Ground Truth Systems

Platform-based systems give ground truth for part poses by placing the part on a platform that fixes the pose in advance of a test. There is no need to sense part position unless greater precision is needed than the fixture can provide. Part identity is also known in advance. Strength of platform-based systems is the ability to set fairly precise pose and location in addition to identity.

A simple way to create a platform-based ground truth system is to use a Lego baseboard, pegboard, or optical bench as shown in Figure 11. Parts of known identity are placed in known locations in this external coordinate system, so identity can be derived from where the system under test places an identified part. Relative calibration of the ground truth system and system under test are needed.

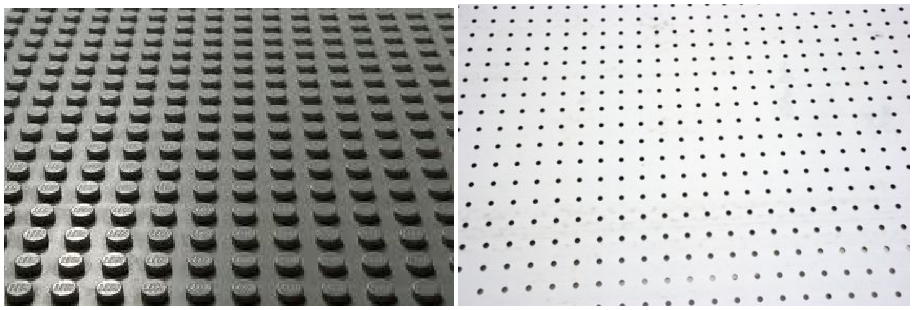


Figure 11. Lego baseboard or a pegboard can be used for a ground truth platform.

In Marvel et al. [9], the authors discuss multiple platform-based systems developed to estimate poses of parts in 6 degree of freedom (6DOF) Cartesian space (X, Y, and Z coordinates plus roll, pitch, and yaw) to determine performance and measurement accuracy of different systems. They describe three 6DOF ground truth systems utilized at the National Institute of Standards and Technology (NIST) (as shown in Figure 12): a laser-tracker-based system for pose measurement while the part is moving, an aluminum fixture-based system that can be used to fix the pose of an part while it is static, and a modular, medium-density fiberboard (MDF) fixturing system for location and pose for static parts. The authors provide descriptions, characterizations, and measured accuracies of these systems. The NIST systems were inspired in part by previous work (Radu et al. [10]), which used a pan-tilt mechanism to control a platform.

The systems share the characteristic that a single part is held on a platform so that its pose is precisely known. The first system uses a robot arm, the second a rotatable metal stage, and the third a fiberboard fixture. The three systems have varying accuracies – the robot arm with laser tracker is accurate to about 0.01 mm in position and 0.03° in rotation; the metal platform is accurate to about 0.5 mm in position and 0.17° in rotation; and, the fiberboard is accurate to about 0.67 mm in position and 0.17° in rotation. Similarly, the three decline in precision – the robot arm can be oriented with repeatability to several decimal places (in degrees), the rotating platform has repeatable fixed stops at limited orientations, and the fiberboard is constructed with stops at 15° intervals. The first two can be made to move dynamically, while the fiberboard must be configured for each test. If the robot arm is precise enough in its movements, it alone can be used to define a ground truth pose. In the NIST system, the laser tracker further refines the known robot position.

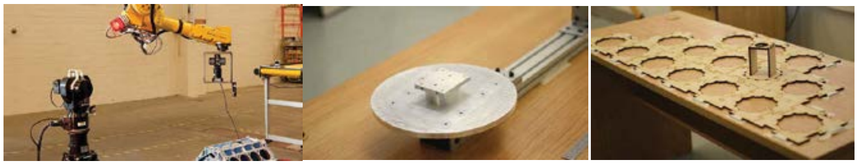


Figure 12. The three systems developed and tested at NIST. [Permission granted by the author [9] to use the Figure]

Marvel et al. [11] presented the results and performance metrics of the 2011 Perception Challenge. The artifacts used in the challenge and the ground truth rigs with the sensor are shown in Figure 13.

Figure 13. The artifacts used in the challenge and the ground truth rig with the sensor mounted at top left. [Permission granted by the author [11] to use the Figure]

Platform-based systems have the following advantages over other options:

a) They are potentially of low cost, although increasing precision and accuracy can be expensive.
b) The desired pose and location can be set in advance, although possibly in limited increments.

There is a potential to use robot arms with known repeatability. The disadvantages of the platform-based approach include the limited position options for the low cost systems, the artificiality of the scenario, and the difficulty of integrating multiple parts, clutter, and complex backgrounds. This approach is best for measuring pose and location of smaller individual parts.

## 3.4. Physically-Sensed Ground Truth Systems

To evaluate the performance of part recognition/identification systems reliably, the ground truth system needs to know the identities of all the parts with a very high probability of correctness compared to the system under test. Ideally, ground truth identity should be known perfectly. For some of the stationary part recognition tests, the ground truth identity is exact, since the experimenter places known parts in the scene. For some tests with moving parts or different sensors, as mentioned earlier, if we know the identities and locations of all the parts at the start, then if we can track them accurately we will also know their identities throughout the test. There are some more complex scenarios where we might not be able to use the previous approaches to get the true identities of parts, such as when parts appear randomly on a conveyer-belt or are randomly dropped into a bin for testing. In these cases, we need to rely on physically-sensed identification systems for the ground truth. An advantage of physically-sensed identification systems is the ability to use them to conduct long randomized tests, where parts are continuously fed in for identification and where manually keeping track of identity is tedious. The identification ground truth system can be classified based on the physical method used for identification (optical, RF, etc.).

### 3.4.1. Optical and Vision based identification

Optical identification systems can use markers or tags attached to the object, or can be markerless so they use inherent characteristics of the object such as shape or pre-existing texture. Since we are considering controlled experiments with the time to attach markers, and markerless systems have not demonstrated the reliability to serve as ground truth systems for identification, we consider marker-based systems here.

Bar codes [12,13] are widely used for part identification in logistic environments, but have many limitations in manufacturing environments. Bar code scanners usually require a line-of-sight to the bar codes. They typically have to be close to the parts that are being identified and can scan only one at a time. Additionally, the bar code scanners or part must physically move between different reads. This process limits the read rate to at best only a few per second. This has led to the development of many 2D symbols where a human is still typically involved in aligning the camera to read the code, such as Quick Response (QR) codes [13, 14, 15, 16, 23]. Augmented reality (AR) tags are a related technology that allows identification. There are commercial systems based on visual pattern recognition technology that provide reliable and robust vision solutions for particular categories of parts

[17][18]. Bar codes or 2D symbols must be applied so as to be visible to the ground truth system, and in some cases preferably not visible to the system under test.

Active optical systems use one or multiple Light Emitting Diodes (LEDs) that are successively powered in phase with the capture system. This provides a unique identification of each LED. The ability to identify each marker in this manner is useful in real-time applications and for collecting data from large numbers of objects/parts [19][20]. Time modulated active markers based on high-speed cameras provide unique identification.

For both passive bar codes and active optical systems, as noted, the presence of tags may interact with the system under test. Since they provide an extra clue to the object's identity, they make the test less valid. One option is to use covert barcodes that are only visible and read in Ultraviolet (UV) light (http://www.uvreaders.com/). Current readers only work at close ranges.

The main drawbacks of optical systems are that line-of-sight is required and the targets have a small information payload.

## 3.5. Radio Frequency Identification Based Systems

Radio Frequency Identification (RFID) based systems [21][22] are widely used for identification and can provide a large payload of information about the detected parts/objects. RFID is a wireless non-contact system that uses radio frequencies to transfer data from a RFID tag for the purpose of identification. A typical RFID setup consists of readers and tags and a back-end computer that receives, stores, processes, and transmits the information that the readers have collected from the different tags. There are three basic types of RFID tags: active, passive, and semi-passive. RFID readers can read hundreds if not thousands of tags per second without requiring line-of-sight as in optical systems (bar code, 2D symbols, etc). Thus, they allow the simple automation of the reading process and making RFID-based identification very attractive for parts identification studies. As the identification process is automated, however, we must ensure the successful reading of all the tags within the readers' field. RFID tags need to be selected for the application, considering part composition (metal parts will reflect some RFID frequencies), distance between reader and part, and whether the tag is passive or active (self-powered).

There are still some issues with recognizing many tags simultaneously, and using them with metal parts. In a cluttered scene, the RFID tags may not give adequate localization, so we may know that a particular part is present, but we may not be able to tell where it is relative to other parts. In a scene with metal parts and considerable electromagnetic interference, the RFID tags may not work. One alternative, in a controlled experiment, is to create a set of plastic parts, possibly with 3D printing, and create an experimental set-up that minimizes Electromagnetic (EM) interference. RFID tags could also be embedded inside the parts.

The other RF technologies used for identification are WiFi, Ultra-wideband, near field communication (NFC) and Bluetooth. Given that RFID technology is well-established and offers a wide range of tag types and readers, we have limited our discussion to this technology.

## 3.6. Physically-sensed localization systems

A physically sensed system can be used to determine the position of parts in the scene, which can then be used as localization ground truth. Some possible systems are a laser tracker, 3D coordinate measuring machines, optical trackers, mechanical arms, etc. Any possible 3D point or 3D scanning system can be used. An example is shown below in Figure 14. We will not discuss these systems, because there are a number of good review papers [24][25].

Figure 14. 3D coordinate measuring system for parts localization.

## 4. Conclusion

In this report we have reviewed several technologies for part identification that could be used as a ground truth to test proposed operational manufacturing systems. Given the wide number of scenarios in which identification could be applied, and the variety of technologies that could be used in operational systems, test procedures may need to have multiple options for ground truth.

Based on this review, we find the following ground truth options worth further investigation for real-time, physical experiments; simulation and annotation can be useful for offline analysis when appropriate:

1) Two-dimensional optical markers or tags that can be read from a distance. Questions to be answered are whether the tags can be applied to smaller parts and remain readable from a distance, at what distance can tags be reliably read, and how they can be applied so as to be visible to the ground truth sensor under most conditions. There is an option for human supervision to use a handheld device to read the barcode, making the test procedure semi-automated and using the barcode primarily for ease of bookkeeping.

2) Covert, or ultraviolet, markers not visible to the system under test. Similar questions as above apply here.

3) RFID tags on regular parts. There are multiple types of RFID tags, both active and passive, of different sizes and operating on difference frequencies. Questions here are what type(s) are most useful, under what conditions can they be reliably read, from what distances can they be read, and what localization information can be derived to place an object in a scene.

4) RFID tags embedded in specially made plastic parts. In this case the RFID tag can be larger, and possibly self-powered, since it is hidden. A question is whether the parts can be made with appearance suitably close to real parts.

## 5. References

[1] Bespalov, Dmitriy, Cheuk Yiu Ip, William C. Regli, and Joshua Shaffer. "Benchmarking CAD search techniques." In Proceedings of the 2005 ACM symposium on Solid and physical modeling, pp. 275-286. ACM, 2005.
[2] Jayanti, Subramaniam, Yagnanarayanan Kalyanaraman, Natraj Iyer, and Karthik Ramani. "Developing an engineering shape benchmark for CAD models." Computer-Aided Design 38, no. 9 : pp. 939-953, 2006.
[3]Software rendering, http://en.wikipedia.org/wiki/Software_rendering, December 15, 2012.
[4] English, Chad, Galina Okouneva, Pierre Saint-Cyr, Aradhana Choudhuri, and Tim Luu. "Real-time dynamic pose estimation systems in space: Lessons learned for system design and performance evaluation." Special issue on “Quantifying the performance of intelligent systems,” International Journal of Intelligent Control and Systems 16, no. 2 : 79-96, 29011.
[5] Park, In Kyu, Marcel Germann, Michael D. Breitenstein, and Hanspeter Pfister. "Fast and automatic object pose estimation for range images on the GPU." Machine Vision and Applications 21, no. 5 (2010): 749-766.
[6] Mihalcik, David, and David Doermann. The design and implementation of ViPER. Technical report, 2003, http://www.cs.umd.edu/Grad/scholarlypapers/papers/davidm-viper.pdf, 2003.
[7] Russell, Bryan C., Antonio Torralba, Kevin P. Murphy, and William T. Freeman. "LabelMe: a database and web-based tool for image annotation." International journal of computer vision 77, no. 1: 157-173, 2008.

[8] Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 3354-3361. IEEE, 2012.

[9] Marvel, Jeremy A., Joe Falco, and Tsai Hong. "Ground truth for evaluating 6 degrees of freedom pose estimation systems." In Proceedings of the Workshop on Performance Metrics for Intelligent Systems, pp. 69-74. ACM, 2012.

[10] Rusu, Radu Bogdan, Gary Bradski, Romain Thibaux, and John Hsu. "Fast 3d recognition and pose using the viewpoint feature histogram." In Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, pp. 2155-2162. IEEE, 2010.

[11] Marvel, Jeremy A., Tsai-Hong Hong, and Elena Messina. "2011 solutions in perception challenge performance metrics and results." In Proceedings of the Workshop on Performance Metrics for Intelligent Systems, pp. 59-63. ACM, 2012.

[12] Schmidt, Malte, Lars Thoroe, and Matthias Schumann. "RFID and Barcode in Manufacturing Logistics: Interface Concept for Concurrent Operation." Information Systems Management, just-accepted , 2013.

[13] Xu, Anqi, and Gregory Dudek. "Fourier Tag: A Smoothly Degradable Fiducial Marker System with Configurable Payload Capacity." In Computer and Robot Vision (CRV), 2011 Canadian Conference on, pp. 40-47. IEEE, 2011.

[14] Chu, Chung-Hua, De-Nian Yang, and Ming-Syan Chen. "Image stablization for 2D barcode in handheld devices." In Proceedings of the 15th international conference on Multimedia, pp. 697-706. ACM, 2007.

[15] Wagner, Daniel, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. "Pose tracking from natural features on mobile phones." In Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 125-134. IEEE Computer Society, 2008.

[16] Fiala, Mark. "Automatic projector calibration using self-identifying patterns." In Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on, pp. 113-113. IEEE, 2005.

[17] Tanaka, Hideyuki, Yasushi Sumi, and Yoshio Matsumoto. "A high-accuracy visual marker based on a microlens array." In Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, pp. 4192-4197. IEEE, 2012.

[17] Evoretail LaneHawk, http://www.evoretail.com/lanehawk/ ,December 15, 2012.

[18] StopLift checkout vision systems, www.stoplift.com , December 15, 2012.

[19] Phase Space motion capture system, http://www.phasespace.com/impulse_motion_capture.html, December 15 2012.

[20] Raskar, Ramesh, Hideaki Nii, Bert Dedecker, Yuki Hashimoto, Jay Summet, Dylan Moore, Yong Zhao et al. "Prakash: lighting aware motion capture using photosensing markers and multiplexed illuminators." In ACM Transactions on Graphics (TOG), vol. 26, no. 3, p. 36. ACM, 2007.

[21] Klaus Finkenzeller. RFID–Handbuch. Hanser Fachbuch, 1999. Also available in English as RFID Handbook: Radio-Frequency Identification Fundamentals and Applications, JohnWiley & Sons, 2000.

[22] Vogt, Harald. "Efficient object identification with passive RFID tags." Pervasive Computing -2002: 98-113. 2002.

[23] The On Target system, http://www.eliteeng.com/on%20target%203d%20camera.htm, December 15 2012.

[24] Cuypers, Wim, Nick Van Gestel, André Voet, J-P. Kruth, J. Mingneau, and P. Bleys. "Optical measurement techniques for mobile and large-scale dimensional metrology." *Optics and Lasers in Engineering* 47, no. 3: 292-300, 2009.

[25] Mautz, Rainer, and Tilch, Sebastian. Survey of optical indoor positioning systems. In Indoor Positioning and Indoor Navigation (IPIN), IEEE 2011 International Conference on, pp. 1-7, 2011.

[26] Everingham, Mark, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. International journal of computer vision, 88, no. 2: pp. 303-338, 2010.

[27] Torralba, Antonio, Fergus, Rob and Freeman, William T.. 80 million tiny images: A large data set for nonparametric object and scene recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30, no. 11: pp. 1958-1970, 2008.

[28] Yao, Benjamin, Yang, Xiong and Zhu, Song-Chun. Introduction to a large-scale general purpose ground truth database: methodology, annotation tool and benchmarks. Energy Minimization Methods in Computer Vision and Pattern Recognition. Springer Berlin/Heidelberg, 2007.

[29] Amazon Mechanical Turk (MTurk), https://www.mturk.com/mturk/welcome, March 05, 2012.