THREE-DIMENSIONAL LASER-ASSISTED IMAGE ANALYSIS FOR ROBOTIC SURFACE OPERATION WITH CAMERA-SPACE MANIPULATION

A Dissertation

Submitted to the Graduate School of the University of Notre Dame in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Biao Zhang, B.S., M.S.

Steven B. Skaar, Director

Graduate Program in Aerospace and Mechanical Engineering

Notre Dame, Indiana

December 2007
THREE-DIMENSIONAL LASER-ASSISTED IMAGE ANALYSIS FOR ROBOTIC
SURFACE OPERATION WITH CAMERA-SPACE MANIPULATION

Abstract

by

Biao Zhang

In light of the apparent inability of alternative strategies to guide robots satisfactorily, in real-world settings of daily operation, relative to work pieces without precise, permanent fixture as required by “teach/repeat”, the present dissertation proves real-world robustness and workability of the presented paradigm with actual tasks found in industry. Two fundamental experiments -- thickness-reduction-gauging and five-component positioning -- prove experimentally that the high precision, robustness, and workspace-extent versatility of CSM can be extended to general robot operations.

Three types of general applications are implemented based on the present method. The first, peg-in-hole assembly, is a challenging robotics application in industry today, one that requires high precision and robustness of five-component positioning. The second, surface-finishing, task requires a combination of gauging the surface change and positioning the tool with high precision. The third, palletizing and de-palletizing, applications demonstrate the robustness of three-dimensional image analysis in real-world tasks.
To my wife Zhengjie who supports me unconditionally in all things

To my children Theodore and Angela

To my parents
CONTENTS

Figures........................................................................................................................................ v

Tables.......................................................................................................................................... ix

Chapter 1: Introduction............................................................................................................... 1
  1.1 Motivation......................................................................................................................... 1
  1.2 Research Objectives........................................................................................................ 5
  1.3 Outline of Dissertation................................................................................................. 9

Chapter 2: Background .............................................................................................................. 11
  2.1 Mainstream Methods of Vision-Guided Robotics ...................................................... 11
  2.2 The survey of optical 3D shape measurement ............................................................ 13
    2.2.1 Time of Flight ..................................................................................................... 14
    2.2.2 Interferometry Measures Depth. ...................................................................... 14
    2.2.3 Triangulation..................................................................................................... 15
    2.2.4 Comparison of the Three Categories of Approach in Visually Guided
        Robotic Applications .................................................................................. 18

Chapter 3: The Basics of Camera-Space Manipulation, Laser-spot Identification and
           User Point-and-click ............................................................................................. 22
  3.1 Introduction to Camera-Space Manipulation (CSM).............................................. 22
  3.2 User Point-and-Click and Laser Spot Identification............................................. 28

Chapter 4: Three-dimensional Laser-assisted Image Analysis for Surface Operation
           with CSM.................................................................................................................. 31
  4.1 Difficulties and Limitations in 2D Image Analysis ............................................... 31
  4.2 Motivation for Three-Dimensional Image Analysis.............................................. 33
  4.3 Three Dimensional Image Analysis....................................................................... 34
  4.4 Edge Detection Based on 3D Image Analysis....................................................... 36

Chapter 5: Fundamental Experiments of Laser-spot-assisted, 3D Image Analysis........ 38
  5.1 Thickness-Reduction-Gauging Experiment Based on Laser-Spot-assisted,
      3D Image Analysis ................................................................................................. 38
    5.1.1 Objective of Experiment.................................................................................... 38
    5.1.2 Experimental Setup........................................................................................... 39
FIGURES

Figure 1.1 Valve body assembly........................................................................................................ 6

Figure 1.2 A plane may be fit that locates the uppermost boundary of the hole of interest..................................................................................................................... 7

Figure 1.3 Turbo engine blade ........................................................................................................ 8

Figure 2.1 Principles of non-contact 3-D shape measurements...................................................... 13

Figure 2.2 Principle of a time-of-flight sensor.............................................................................. 14

Figure 2.3 Interferometry 3-D reconstruction............................................................................. 15

Figure 2.4 Hierarchy of the most important principles of triangulation techniques................. 16

Figure 2.5 Triangulation Principle for 3D imaging....................................................................... 16

Figure 2.6 Simple stereo camera geometry with parallel optical axes ..................................... 17

Figure 3.1 Coordinate frames of a typical vision system............................................................. 23

Figure 3.2 Coordinate frames of Camera-Space Manipulation vision system........................ 23

Figure 3.3 Projection according to the pinhole camera model ................................................... 26

Figure 3.4 Graphic User Interface for User’s Point-Click............................................................ 28

Figure 3.5 Point-Click CSM System .......................................................................................... 29

Figure 3.6 Applying a mask to each pixel provides data regarding its value as well as surrounding pixel values......................................................................................... 30
Figure 4.1 Box Picking-Up Task ................................................................. 32
Figure 4.2 multiple laser spots are cast on object surface ....................... 35
Figure 4.3 Three surface meeting ............................................................ 36
Figure 5.1 The setup of system ................................................................. 40
Figure 5.2 Metal plates ........................................................................... 40
Figure 5.3 Laser spots fall on the top surface of top plate ......................... 41
Figure 5.4 Laser spots fall on top surface of lower plate ......................... 42
Figure 5.5 Cue-bearing plates approach workpiece .................................. 42
Figure 5.6 Multiple paper cues on surface ............................................. 43
Figure 5.7 Histogram of surface points vs. thickness .............................. 46
Figure 5.8 Normal distribution of individual thickness assessment ........... 47
Figure 5.9 System configuration ............................................................. 53
Figure 5.10 Position the end member (L-shape) onto surfaces 1 and 2 ...... 54
Figure 5.11 Intersection line ................................................................. 54
Figure 5.12 Four points in each of two surfaces were selected ................ 55
Figure 5.13 Multiple laser spots on two surfaces .................................... 56
Figure 5.14 Spots in two groups ............................................................. 56
Figure 5.15 The normal vectors of two planes and intersection line .......... 57
Figure 5.16 The end effector approaches target ...................................... 57
Figure 5.17 Sample the cues on end effector .......................................... 58
Figure 5.18 Robot positions the end member on two surfaces ........................................ 58
Figure 5.19 Measurement .............................................................................................. 59
Figure 5.20 Experimental outcome .............................................................................. 60
Figure 5.21 Definition of out of plane and in plane component positions and orientations ................................................................. 60
Figure 5.22 Cues on end member ............................................................................... 62
Figure 5.23 Variation of the values of Z component of the nominal coordinates of the points on a flat surface ........................................................................... 63
Figure 5.24 Cues on horizontal plate ........................................................................... 64
Figure 6.1 Camera on the robot end-effector ............................................................... 68
Figure 6.2 Perspective camera model, as example of depth inference ......................... 69
Figure 6.3 Engine head assembly ............................................................................... 70
Figure 6.4 3D laser vision system developed by Fanuc .............................................. 71
Figure 6.5 Force control torque convert assembly ...................................................... 72
Figure 6.6 Force control insertion process .................................................................. 72
Figure 6.7 Two insertion parts .................................................................................... 73
Figure 6.8 Peg on robot end-effector .......................................................................... 74
Figure 6.9 Multiple lasers on cylinder receptacle ....................................................... 75
Figure 6.10 Illustration of spot segmentation ............................................................... 75
Figure 6.11 Laser spots on outer and inner surface of the cylinder receptacle .......... 76
Figure 6.12 Tests on various location and orientation of the cylinder receptacle ....... 78
Figure 6.13 At various scales, from a small mold to a ship’s hull, human action is often needed in order to effect the required finishing process. ........................................ 79

Figure 6.14 Setup for sanding project.............................................................................................................. 81

Figure 6.15 Select the points to specify the region of interest, using computer monitor. 82

Figure 6.16 Multiple lasers are cast onto the surface and identified .................................................. 82

Figure 6.17 Illustration of interference between a sanding block and workpiece surface. .......................................................... 83

Figure 6.18 White crosses on wood board indicate where measurement occurred .......... 85

Figure 6.19 Pattern of bag stacking ........................................................................................................ 87

Figure 6.20 Bank of laser pointers........................................................................................................... 88

Figure 6.21 Gaps on paper bags stack .................................................................................................. 89

Figure 6.22 Gap insertion ..................................................................................................................... 89

Figure 6.23 Vision guided de-palletizing system overview............................................................... 90

Figure 6.24 Multiple laser spots on three surfaces of bag stack .................................................. 91

Figure 6.25 Edges and corner of the bags stack ....................................................................................... 91

Figure 6.26 Rough estimation of the location of the gap ............................................................... 92

Figure 6.27 Laser spots on front surface ............................................................................................... 92

Figure 6.28 The robot inserted the tool into the gap and pick up the bags ........................................ 93

Figure A.1 Coordinate frames of Camera Space Manipulation vision system .................. 98
TABLES

Table 5.1 Experiment result................................................................. 44
Table 5.2 Standard deviation of the individual assessments of plate thickness........ 47
Table 5.3 Experimental results .............................................................. 61
ACKNOWLEDGMENTS

I would like to thank the University of Notre Dame for offering me this opportunity of studying in this great campus. To my advisor and mentor, Dr. Steven B. Skaar, I would like to express my grateful appreciation for his advice, support and motivation through these past five years, which will be beneficial to my career for the rest of my life.

I want to thank my committee members: Dr. Stanisic, Dr. Goodwine, Dr. Bowling and Dr. Madey for their valuable comments and suggestions.

I also want to thank William Aman, who has been working closely with me on the vision-guided robotic palletizing system. His experience and dedication are the key to the success of this system. I am thankful to Jack Forbes and Wayne Schumm’s great support and direction to my work in Smurfit-Stone on this successful project.

Finally, I would like to thank my wife Zhengjie for her unconditional support and encouragement over the years in all the things, and my parents for their love.
1.1 Motivation

In 1920, Czechoslovakian playwright Karel Capek introduces the word robot in the play R.U.R. - Rossum's Universal Robots. The word comes from the Czech robota, which means tedious labor. About thirty years later, George Devol designs the first programmable robot and coins the term Universal Automation, planting the seed for the name of his future company - Unimation. This manipulator is the first of many Unimates to be deployed, and also the beginning of the modern robots. After that, many different types of robots were developed and built, and many robotics companies were founded. A new robotics industry emerged. Meanwhile, academic robotics research became popular. Researchers studied robot kinematics, robot dynamics, artificial intelligence and robot vision. Now after half a century, the public believes robots can do what humans can do since researchers in the robotics field claim and show their success in many applications. Robots are working on dangerous jobs, doing factory work, and providing household help. Robots are playing in robot wars, robot soccer or as toy robotic dogs. Robots explore Mars. Robots are sensing light, sound and smells. Robots are talking to you and thinking about answering your questions.

Do all of these evidences really mean robots can do what humans can do? People might say, “No. Because robots still do not have enough intelligence to do complex tasks
associated with thinking.” But if we believe that artificial intelligence can solve the last obstacle of using the robots to do all or a portion of jobs now done by humans, how can we explain the fact that millions of human workers are still doing simple, repetitive and tedious tasks of factory production? Remember the word robot comes from the Czech word robota, meaning drudgery or slave-like labor. Why is it that the robots still do not replace humans on even these simple jobs? Especially the current industrial robots have enough power, speed and dexterity for almost any task. This has been proved as thousands of robots are employed in certain kinds of production lines. However, as a chief researcher at one of the largest robot manufactures said “We don’t actually make robots; what we make are just programmable machines”. It means that, generally, for 3D operation, robots work in the teach-and-repeat mode. The robot can only move to particular poses or along particular paths previously taught by a human. This mode takes advantage of the excellent repeatability of modern industrial robots. But it also limits robots only to handle inflexible work environments, which means everything in the work cell should be the same as when the robots were taught; for instance the position, orientation and dimensions of the parts must not change.

Someone may argue that the robot is required by the objective of the task to be used in this teach-repeat way. Is this true? If the robot can work automatically out of the teach-and-repeat mode, who would want to spend time and money to build the expensive specialized jigs and fixtures? Moreover, if the cost of using robot would be lower than human labor and also robot could do what human worker do, why we did not see more and more robots replace the human worker in production line? In fact, the replacement trend went in the opposite direction. In Japan, the world’s most robotized economy, the
number of robots has dropped steadily from a peak of 413,000 in 1997 to 350,000 in 2000, as companies choose not to replace some aging robots [1]. Meanwhile, the cost per robot unit keeps declining and robot quality keeps improving.

If a “simple” task, such as grasping a part cannot be done by a robot, what can be expected that a robot might do under some high level artificial intelligence? Therefore the question, how to make robots do all or portion of routine jobs now done by humans, becomes a critical and fundamental question, which needs to be answered for breaking that present limitation of robots.

This dissertation presents work related to camera-space manipulation (CSM) in order to answer the question. CSM is robustly capable of delivering a level of three-dimensional positioning accuracy, relative to a flat surface, that exceeds the accuracy of the kinematical calibration of the robot mechanism by two orders of magnitude and exceeds the pixel resolution of participant cameras by at least one order of magnitude. This ability is tied to CSM’s use of multiple-pose camera-image samples of the robot’s end-member in the joint-space vicinity of the terminus, samples that ordinarily may be acquired as the robot’s end member approaches its current-maneuver terminus. Because the most current end-member samples acquired en route to the terminus need not be applied until critical high-precision maneuver junctures are neared, and the maneuver can therefore ensue based upon earlier-processed samples even as newer samples are being processed, the method is compatible with nonstop motion in real time.

The individual-camera-space objectives associated with such a maneuver are understood in terms of “common points” – discrete physical surface points identified in terms of two-dimensional camera-space coordinates as detected in two or more
participant “CSM cameras.” These surface junctures lie on the workpiece or “target” surface(s) of interest. Camera-space coordinates of the junctures are identified in advance of the robot’s approach using “matched laser spots” or appearances of the centers of individual dots of laser light reflected off the surface and into the image planes of the same stationary-but-uncalibrated cameras subsequently used to guide the robot. The matching of the laser spots ensures that the system sorts the two-component camera-space positions or pixel addresses of spot-center locations in accordance with the common physical-juncture identity of each. Several such dots or spots of light may be accumulated in advance of the action of the manipulator using multiple images in each camera, where each image is associated with a particular “blast” of single- or multiple-beam laser pointers, each blast directed by a slightly different pan/tilt angle of the laser-pointer-bearing pan/tilt unit to illuminate slightly different surface junctures each time. The result, provided the region over which the spots are accumulated is nearly planar, is a kind of averaging that, on the “target side” (as opposed to the complementary “CSM side” which controls the robot), allows for precision of positioning far in excess of the pixel resolution of the participant cameras. A similar exploitation of information redundancy on the CSM side – redundancy involving multiple exposures to the CSM cameras of the robot end effector and in particular of end-effector-borne “cues” – ensures similar sub-pixel precision in delivery of the positioned point. The manipulated point that is brought into coincidence with the surface may, depending upon the task description, be a point physically located on the manipulated body’s surface, a point embedded within that body, or a point that lies outside the physical end effector. In each case, however, the manipulated point moves as a rigid body with the end effector, and is
presumed herein to previously have been characterized precisely in position relative to that body’s visual cues.

While the use of high redundancy and current-maneuver, local data can ensure both precision and robustness, the above procedure limits robot exploitation to just three degrees of freedom compared to the six degrees of freedom often available. That is because a flat surface alone presents information that permits only one component of maneuver-position specificity between manipulated point and target, together with just two components of rigid-body orientation with respect to the target body. Hence, three additional constraints must be imposed on a six-axis robot in order to complete the maneuver. While this extent of control is sufficient for a range of tasks such as surface finishing, many more tasks could be included with a generalization of the same mode of operation to tasks and target bodies that present additional surfaces as the basis for movement. Additionally, non-flat surfaces, particularly if their geometry is known a priori, can be addressed with high precision by means of further extension of [2] herein.

1.2 Research Objectives

The objective of this thesis is to motivate, develop and prove experimentally that the same precision, robustness, and workspace-extent versatility shown in [2] can be extended to these more general cases. Moreover, in light of the apparent inability of alternative strategies to guide satisfactorily, in real-world settings of daily operation, robots relative to workpieces without the known and permanent fixturing as required by “teach/repeat”, the present dissertation strives to prove real-world robustness and
workability of the presented paradigm with actual tasks found in industry such as assembly, surface-finishing and palletizing/depalletizing tasks.

Figure 1.1 presents an assembly-task example. The left image shows a piece already in one of the holes. The arrow of the next image indicates desired insert-piece direction into nearest hole. Laser spots that are identified as falling on the upper lip of this nearest hole can be applied to control robot’s terminal depth of insert action. And laser spots falling on the inner and/or outer walls of the hole of interest can be used for in-plane positioning or alignment purposes. We assume here that laser spots have been cast and matched, and that, as indicated on the third image from the left, certain spots have been identified as lying on the lip of the hole into which the next insert piece is to be placed. As discussed in [2], multiple cast and matched spots known to lie on a common physical plane can be used, as indicated in Figure 1.2, to refine the position of that infinite plane relative to a “nominal world coordinate system”. This coordinate system has an important attribute attained through the use of CSM: It is the coordinate system relative to which the nominal kinematics model – the same model given by the manufacturer – is locally virtually perfect. The word “locally” here means the local region of joint space, physical space and time near which the maneuver will culminate.

Figure 1.1 Valve body assembly
As indicated in Figure 1.2, using 3D coordinates of the indicated spots relative to the pertinent “nominal physical coordinate system”, a plane may be fit that locates the uppermost boundary of the hole of interest.

![Figure 1.2 A plane may be fit that locates the uppermost boundary of the hole of interest](image)

If the number of matched spots is large, and if the requirements on the “CSM side” for en-route sampling of end-of-arm “cues” are adequate as discussed in [2] then, once the plane shown in Figure 1.2 is found, the robot may be controlled to guide the insertion piece to a precise depth with precise orientation relative to that plane. That would provide for one component of position and two components of orientation, or control of three degrees of robot freedom.

Two additional components of position, in-plane position, may be identified using matched laser spots indicated in the right-most part of Figure 1.1 – those that fall on the inner and outer surfaces of the cylindrical receptacle. Provided the physical shape of the cylinder is known, the 3D coordinates of these spots relative to the aforementioned nominal physical frame can be fit precisely to a geometric model of the feature, thus providing the needed pair of coordinates for rigid-body positioning. (For this 5DOF task, in-plane orientation need not be controlled.)
It is worth noting that the fit and location of the hole itself can, if for example the outer shape is conic but not necessarily cylindrical, benefit from knowledge gained about the location of the top plane. Orientation of the top plane, once established, reduces the number of degrees of arbitrary orientation of the cylinder. Furthermore, the modeled and best-fit hole can, along with the plane, serve as a basis for out-of-plane orientation control of the insertion.

Matched laser spots and local camera-space kinematics, however, combine to allow for a three-dimensional understanding of individual-spot coordinates. This helps with the segmentation problem in two ways: It removes from consideration surface coloration and ambient lighting. And it provides a three-dimensional basis for sorting the spots according to surface region.

Another category of problem considered herein can be seen in Figure 1.3. Here, prior knowledge of the geometry of the surface of interest may not be needed for purposes of surface treatment such as deburring, cleaning, or polishing. Yet the surface may not be adequately close to planar to allow for the use of a plane assumption in aggregating the information of multiple spots, for example within the indicated region, to create the desired robot motion.

Figure 1.3 Turbo engine blade
1.3 Outline of Dissertation

The dissertation consists of seven chapters. An overview of each chapter is given in this section. Following this introductory chapter is a chapter discussing some background issues. First the mainstream methods of vision-guided robotics are reviewed. Then the survey of optical 3D shape measurement is given. Chapter 3 is to explain the basics of CSM, laser-spot identification and user point-and-click, which are the bases of the three-dimensional laser-assisted image analysis for surface operation with CSM.

Chapter 4 constitutes a large portion of the original work presented in the dissertation. It begins with a discussion of the difficulties and limitations associated with the two-dimensional traditional image analysis. It provides the motivation for what we here term three-dimensional image analysis. The target vision information is specifically for a CSM system. For that particular goal, the feature extraction and location procedure discussed herein should be reliable and robust.

In the Chapter 5, two fundamental experiments, thickness reduction gauging and five-component positioning experiment are presented. The high precision and robust results prove experimentally the high-precision robustness of surface operation based on the three-dimensional laser-assisted image analysis with CSM. They also lay the foundation of the extension of these to more general cases.

Chapter 6 deals with three types of general applications performed with the three-dimensional laser-assisted image analysis for surface operation. The first peg-in-hole assembly task requires the high precision and robustness five-component positioning. This type of assembly task is a challenging robotics application in industry now. The current methods are compared with the presented approach. And the assembly procedure
is described. The result shows how high-precision five-component positioning extended into the general case. The second surface-finishing task requires the combination of gauging the surface change and positioning the tool with high precision. It is applied both to the thickness gauging method and high precision five-component positioning. The third palletizing and de-palletizing applications show the robustness of three-dimensional image analysis in the real world.

Chapter 7 concludes the dissertation, which ties together the material presented in the dissertation. It reviews the original work presented, the experimental results achieved and the real world applications developed. Finally recommendations for future research are presented.
2.1 Mainstream Methods of Vision-Guided Robotics

Calibration [3] [4] [5] [6] [7] [9] [28] and visual servoing [8] [9] are two mainstream methods of vision-guided robotics. They are based on opposite philosophies. Calibration is an open loop control strategy. Visual servoing is a close-loop control strategy.

The calibration schemes establish robot-kinematics parameters; they also identify intrinsic camera parameters, which include focal length, pixel scale factors and lens distortion, and extrinsic camera parameters, which decide the position and orientation of the camera pose with respect to world coordinate systems. Calibration builds a global geometric characterization of the mapping between each camera’s image space and 3D space in a pre-selected world coordinate system as well as the mapping between the 3D space and the robot coordinate systems. Calibration relies entirely on an accurate camera model and robot kinematics model to deliver accurate positioning results. Any error at any stage of such a system will contribute to a final positioning error. Also, in the real world, the noise in an image or a slight shift, for example temperature-induced, of the parameters in camera or robot will corrupt the whole elaborate global model, which was carefully calibrated.
Visual servoing takes a close-loop control approach to drive the positioning error in the image toward zero. It makes use of a Jacobian matrix describing the relative changes between the robot’s internal joint coordinates and coordinates of manipulated junctures of interest in camera space. Visual servoing does not require identification of an accurate camera model, robot kinematics model, or Jacobian matrix to achieve precise positioning. Visual servoing is a very adaptive method. However, it does not take advantage of the excellent repeatability of most commercial robots. One of the biggest drawbacks in visual servoing is that they need to access the terminal error between the current pose and target pose in order to adjust the end-effector to close in toward the target. The target must be visually accessible throughout the entire approach. In some applications this would be impossible such as where visual access becomes obscured, or where the target gets occluded from a camera as the system nears the target. Moreover, visual servoing is overly demanding in terms of high-speed image processing to robustly detect the image error. This is usually the weakness of visual servoing.

Camera Space Manipulation (CSM) has a different philosophy, which separates it from calibration and visual servoing. Using stationary cameras, CSM pursues local accuracy by making local to the maneuver terminus a mapping relationship directly between robot joint rotations and appearances of the manipulated body in image space. Unlike calibration which relies upon highly accurate global kinematics and optical models, CSM applies the preplan data to roughly initialize overall camera space kinematics, and then acquires data local to the current-maneuver terminus that lies within a local, model-asymptotic-limit sub-domain of wherever these local samples are applied. The method takes advantage of the excellent repeatability available with most
commercial robots. CSM absorbs the errors of camera and robot models into locally refined camera-space-kinematics parameters. Small time-varying errors, such as the changes of relative positions of cameras and robot due to temperature changes or vibrations also can be absorbed into the localized camera space kinematics parameters by a judicious weighting of the latest samples used to produce the estimates.

2.2 The survey of optical 3D shape measurement

Three-dimensional optical shape measurement delivers absolute 3-D geometry of objects that should be independent of the object’s surface reflectivity, its distance from the sensor, and illumination [10]. From knowledge of the underlying physical principles, optical 3D shape measurement methods can technically be put into three categories: triangulation, time-of-flight measurement (TOF) including broad-band interferometry and classical interferometry [10] [15][17], as shown in Figure 2.1.

Figure 2.1 Principles of non-contact 3-D shape measurements.
2.2.1 Time of Flight

The time-of-flight method for measuring shape is based on the direct measurement of the time of flight of a laser or other light-source pulse. During measurement, an object pulse is reflected back to the receiving sensor and a reference pulse is passing through an optical fiber and received by the sensor. The time difference between the two pulses is converted to distance. A typical resolution for the time of flight method is around a millimeter. The principle behind a time-of-flight sensor is depicted in Figure 2.2.

![Figure 2.2 Principle of a time-of-flight sensor](image)

2.2.2 Interferometry Measures Depth.

Classical interferometry is a technique to measure smooth (polished) surfaces. A coherent wavefront is split into a measuring (or object) and a reference wavefront. These are superimposed (correlated) again in a detector as illustrated in Figure 2.3. If a 2-D detector is used, an interferogram or correlogram is generated, indicating the phase shift over the detector area. With at least three measurements with different phase positions of the reference, the phase shift between the reference and the signal wavefronts can be determined according to the phase-shift principle. Unfortunately, this technique cannot
determine absolute depth. Because of the ambiguity of the signal in multiples of $\lambda/2$, ($\lambda$ is wavelength) a unique depth determination is only possible in this narrow depth range.

![Interferometry 3-D reconstruction](image)

**Figure 2.3 Interferometry 3-D reconstruction**

### 2.2.3 Triangulation

Triangulation normally determines the coordinates and distance of an unknown visual point within a triangle by calculating, based on optics, the length of one side of a triangle with the related side angles pointing to the unknown point.

Figure 2.4 shows the hierarchy of the most important variants, which have the same basic principle of triangulation. At the highest level, they are divided into: (1) focus/defocus techniques; (2) active triangulation with structured illumination; (3) passive triangulation based on the digital photogrammetry and stereoscopy (4) theodolite-measuring systems; and (5) shape-from-shading techniques [10][20][21][22][23][24][25].
The most widely used optical triangulations are active triangulation with structured light, passive triangulation with digital photogrammetry and stereoscopy.

Active triangulation needs structured light, either a laser spot or laser line projected onto the object and reflected to the detector, either a CCD line sensor or a position-sensitive photodetector (PSD). Figure 2.5 shows schematically the triangulation principle for 3D imaging.
Passive triangulation techniques basically include the digital photogrammetry and stereoscopy (stereovision), which gain the 3-D information from the difference of images taken of the same two-dimensional scene. Stereo vision refers to the ability to infer information on the 3-D structure and distance of a scene from two or more images taken from different viewpoints [23]. The stereo camera geometry is depicted in Figure 2.6.

![Figure 2.6 Simple stereo camera geometry with parallel optical axes](image)

Figure 2.6 Simple stereo camera geometry with parallel optical axes

\[ Z = \frac{Tf}{X_r - X_l} \]  

(2.1)

where T is the baseline length - the distance between the lenses centers of the cameras). \(X_l\) and \(X_r\) are the image coordinates of object point P projected onto image planes with respect to left camera coordinate system and right camera coordinate system.
2.2.4 Comparison of the Three Categories of Approach in Visually Guided Robotic Applications

Time of flight is particularly appropriate in applications involving long distances with centimeter precision as well as applications not requiring fast measurement such as scanning [15]. Therefore it does not satisfy the millimeter or sub-millimeter precision and fast measurement for visually guided robot positioning.

Interferometry requires a smooth or polished surface for measuring. This requirement greatly limits its application. Because of the ambiguity of the signal in multiples of $\lambda/2$ ($\lambda$ is wavelength), a unique depth determination is only possible in this narrow depth range. For getting the absolute depth, interferometry needs help from other measurement approaches such as time of flight. Moreover, achieving high-depth accuracy of interferometric measurements requires a mechanically very stable instrument. This is hard to achieve in some real world applications.

Triangulation is the most widely used approach for a visually guided robot system. Most calibration and visual-servoing approaches rely on triangulation to provide the 3D target information. It can be fast and can precisely measure most objects with varying surface structures [10]. Though triangulation is very popular in academic robotics research, there are few successful applications in industry. The reason is that several drawbacks limit real-world application of triangulation for visually guiding robots.

Before estimating the 3D coordinates of targets in a World coordinate system, in order to characterize the stereo-camera rig, the parameters of the camera model in the stereo rig must be calibrated first by acquiring samples on a “calibration fixture”. Either eighteen or twenty-two parameters (or even more in an elaborate camera model [26])
need to be calibrated. The more parameters that need to be calibrated, the easier the error of samples propagated into the 3D data measurement. Real-world error in samples acquired in the calibration process is inevitable. These errors would reduce the 3D data measurement accuracy.

Separation between stereo cameras affects the resolution and precision of the stereo-camera method with higher depth precision associated with wider separation. This restriction limits the visibility of the cameras for the whole surface of the object as well as the scale of the workspace.

The approach to measuring a 360-deg shape of an object is either to transport the stereo camera rig or to apply multiple stereo-camera rigs around the object [14][17]. But 3D data estimated from the stereo rig is with respected to a coordinate system corresponding with the stereo rig. All the local (individual) 3D coordinates are transformed into the global (common) coordinate system and patched together using a least squares fit method [17]. The inevitable errors in these transformations will reduce the 3D data-measurement accuracy.

For applying the acquired 3D data into visual guidance of a robot system, the 3D coordinates with respect to the stereo rig need to be transformed into a robot coordinate system. The error in the transformation will also reduce the 3D data-measurement accuracy.

Once the stereo camera rig is calibrated, the model for estimating 3D data is fixed. Slight changes or vibration of stereo camera rig will reduce the measurement accuracy. It lacks robustness to real-world environmental changes.
The 3D object measurement based on the CSM technique is different from each of the above 3D object measurement approaches. It takes advantage of CSM to overcome all of the drawbacks.

There are only six parameters in the localized model, which are estimated with a combination of deweighted global data and current-maneuver local data. The error in samples for estimating the model, the error of target in camera space and the error from slight changes of camera can be absorbed into model and averaged out. With adequately dense sampling near the terminus, accurate and robust 3D object measurement relative to the frame with respect to which the nominal robot kinematics are near-perfect becomes possible.

The physical reference frame relative to which surface-point position is inferred is one with respect to which the forward kinematics of the robot is locally near-perfect. Multiple cameras can be seamlessly combined together, with the actual position region of this nominal World frame migrating with surface and joint-space. It becomes easy and accurate for the CSM 3D-object measurement system to measure from 360-degrees of vantage-point variation and do the positioning. Because various cameras and various regions of the robot’s joint space shift during execution, there is some distortion of the surface geometric characterization, but locally, in any given subspace, the robot can position relative to the surface with near perfection.

CSM’s uncalibrated cameras may be widely separated, because each camera has its own camera-space-kinematics estimates, rather than applying a constraint between cameras, as with a stereo-camera rig. This makes the CSM approach able to address the
object precisely in a large-scale workspace with individual cameras distant from the work piece. It is adaptable to a wide range of applications.

With multiple cameras in a seamless combination, the system has the capability of self-diagnosing failure among the multiple cameras by comparing the results of 3D estimates from among different combinations of cameras at any moment. It’s an important function of the system to assure reliable operation under real-world conditions. Also it’s useful and convenient for an operator to be notified about any exceptions that occur.

These unique advantages of the CSM 3D object measurement provide the capability of applying 3D image analysis to robust detection of the feature on the object in a visually guided robot system. Chapter 3 presents the details.
CHAPTER 3:
THE BASICS OF CAMERA-SPACE MANIPULATION, LASER-SPOT IDENTIFICATION AND USER POINT-AND-CLICK

3.1 Introduction to Camera-Space Manipulation (CSM)

The method of camera-space manipulation (CSM) emerged in the mid-1980s as a way to achieve both robustness and precision in visually guided manipulation without the need to acquire and sustain precise calibration of cameras and manipulator kinematics, as required by calibration-based methods. Additionally, CSM avoids the visual-servoing requirements for very fast, real-time image processing and for visual access to image-plane errors through to maneuver closure.

Figure 3.1 shows the Coordinate Frames of a typical system for visual guidance of a robot. With calibration, the relationships among all of these frames must be established and the parameters in each transformation model must be calibrated to within whatever degree or extent of precision the maneuvers demand.

In contrast with that, CSM uses six parameters to identify locally the mapping relationship from the internal - and directly controllable – robot-joint rotations within the relative workspace to local 2D camera-space. As indicated in Figure 3.2, the physical 3D points, which scatter around a local origin (flattening point), are projected into the 2D image-plane, with Xc-Yc, as “camera-space coordinates”. These physical 3D points are designated with respect to a local frame, Δx-Δy-Δz, axes
Figure 3.1 Coordinate frames of a typical vision system

of which are nominally parallel to the robot’s world frame and the origin of which is close to the 3D points within a model-asymptotic-limit region. The definition of asymptotic-limit region is in Appendix A. The frame denoted by x-y-z is the robot frame, the coordinate frame attached to the robot base. The frame X-Y-Z is the camera-fixed frame, and the Z axis is aligned with the optical axis of the camera. The X and Y axes are parallel to the axes of the 2D image frame Xc-Yc, and the origin is on the system’s equivalent focal point.

Figure 3.2 Coordinate frames of Camera-Space Manipulation vision system
This local mapping relationship is described in equations (3.1) and (3.2), which correspond to the assumption of an orthographic camera model.

\[ X_c = A_{11} \cdot \Delta x + A_{12} \cdot \Delta y + A_{13} \cdot \Delta z + A_{14} \quad (3.1) \]

\[ Y_c = A_{21} \cdot \Delta x + A_{22} \cdot \Delta y + A_{23} \cdot \Delta z + A_{24} \quad (3.2) \]

Where \( X_c, Y_c \) are with respect to the 2D image frame and \( \Delta x, \Delta y, \Delta z \) are with respect to local frame \( \Delta x-\Delta y-\Delta z \), with origin on the focal axis.

and where each of \( A_{11}, A_{12}, \ldots, A_{24} \) groups a nonlinear expression dependent upon six view parameters \([C_1, C_2, \ldots, C_6]\) as follows:

\[ A_{11} = C_1^2 + C_2^2 - C_3^2 - C_4^2 \quad (3.3) \]

\[ A_{12} = 2(C_2C_3 + C_1C_4) \quad (3.4) \]

\[ A_{13} = 2(C_2C_4 - C_1C_3) \quad (3.5) \]

\[ A_{14} = C_5 \quad (3.6) \]

\[ A_{21} = 2(C_2C_3 - C_1C_4) \quad (3.7) \]

\[ A_{22} = C_{12} - C_{22} + C_{32} - C_{42} \quad (3.8) \]

\[ A_{23} = 2(C_3C_4 + C_1C_2) \quad (3.9) \]

\[ A_{24} = C_6 \quad (3.10) \]

The first four parameters \( C_1-C_4 \) are proportional to four Euler parameters used to characterize a relative orientation between the camera frame, where the camera-space
target coordinates are based, and the nominal World-frame. The last two parameters $C_5$, $C_6$ define the nominal location, in camera-space, of the origin of the local frame.

The view parameters establish a local relationship (camera-space kinematics) between the internal robot joint rotations and the camera-space location of any point on the manipulated body. Together with laser-spot-based assessment of maneuver objective in each camera space, the camera-space-kinematics relationships permit precise calculation of the 3D coordinates of target points in the “nominal World frame”. The nominal World frame is a small, gradually shifting translation and rotation of the actual World frame because of the local differences between the nominal forward kinematics and real forward kinematics of robot. The definition of nominal World frame is in Appendix A. Also the system can calculate the joint rotations required for the robot to position given junctures on its end member onto target points in the nominal World frame. It is important that view parameters of the orthographic camera model are only valid within the asymptotic-limit region, which refers to the region both in physical space and joint space. This means two things: One is that an adequate number of end-member samples for estimating the view parameters should be acquired within the asymptotic-limit region. Also, the target point should be within the same asymptotic-limit region for high-precision positioning. In order to enlarge the asymptotic-limit region a flattening procedure has been used. The flattening procedure is based on a presumption of a pinhole projection of physical points into the two-dimensional image plane, as depicted in Figure 3.3. This procedure consists of modifying the raw camera-space samples of junctures on the robot end effector, so that they become more consistent with the orthographic model of equations (3.1), (3.2).
The X coordinate of an $i^{th}$ raw camera-space sample of a particular juncture on the robot end effector is $X_{ci}$. The flattened sample is determined by $X_{ci}Z_i/Z_o$ based on the assumption of a pinhole or perspective lens model, where $Z_i$ represents the location of the sample along the optical axis of the camera, and $Z_o$ is the location of the origin of the local frame $\Delta x-\Delta y-\Delta z$ with respect to the camera frame.

The Y coordinate of the $i^{th}$ raw camera-space sample $Y_{ci}$ is determined by $Y_{ci}Z_i/Z_o$.

With the use of a weighting scheme on sample data, one which gives more emphasis to the sample close to the target point when estimates of the view parameters are updated, enlarging the asymptotic-limit region not only helps include more sample data, but also reduces the error of noise in sample data propagated into the positioning.

After the camera-space kinematics is established for each camera in the CSM vision system, we have separate camera-specific expressions for equations (3.1) and
With at least 2 cameras and corresponding camera-space coordinates of the target, the target 3D coordinates in the nominal World-frame can be estimated. With more than 2 cameras the accuracy of estimation will be improved because of the geometric advantage of any new viewpoint combined with the averaging affect. The estimation procedure is as follows:

Choose an origin of the local frame, the closer to the target, the better.

Compute \([C_1, C_2, \ldots, C_6]\) for each camera using samples flattened about this local frame’s origin.

Estimate the relative position of the target point with respect to local frame by solving the non-linear questions of (3.1) and (3.2).

Shift the origin of the local frame to the newly estimated target position.

Repeat from step 2 until the shift of target location changes very little between corrective iterations.

Given nominal World-frame coordinates of a target, the process of finding the camera-space coordinates is to choose the target as the origin of the local frame, then compute \([C_1, C_2, \ldots, C_6]\) for each camera. \(C_5\) and \(C_6\) become the \(X_c\) and \(Y_c\), the camera-space coordinates of the target point.
3.2 User Point-and-Click and Laser Spot Identification

As illustrated in Figure 3.4, user involvement (the user may be very remote from the physical system) entails a mouse click onto an image of the work piece acquired from a conveniently placed, often user-directed, camera, called herein the “selection camera”.

![Figure 3.4 Graphic User Interface for User’s Point-Click.](image)

This procedure can designate the target location by clicking onto a single point on the object surface of interest. The user can click onto two points to define an in-plane direction of the orientation and/or movement of the end-effector. Three points on the physical surface can designate a plane, which specifies out-of-plane orientation of desired end-effector positioning. Three or more than three points could define the region of interest for robot operation, for example across a surface. Also, more than one group of points on different surfaces of an object could be used to convey more complicated information about desired action relative to the as-formed, as-located surfaces of the object.
As discussed in the previous section, CSM needs camera-space targets in at least two cameras to do positioning; this is accomplished by estimating the nominal-World-frame coordinates of the target. To establish camera targets for a common point clicked on by the user in the selection camera, a laser pointer mounted on a pan/tilt unit is used to direct a single laser spot onto the surface of the object in order to coincide with the clicked-on point. Then, as depicted in Figure 3.5, the other cameras detect the location of the same physical center of the laser spot in each camera space.

![Figure 3.5 Point-Click CSM System](image)

For detecting the location of the center of the laser spot in each camera space, the laser-spot identification procedure is the following:

Turn on the laser pointer to highlight the juncture of interest on the object surface with a laser spot. Acquire the image of the object surface with the selection camera.

Turn off the laser pointer and acquire the image of the object surface with the camera.
Image difference between these two images to make only the laser spot stand out.

Apply a “mask”, as indicated in Figure 3.6, in order to condition the differenced image, replacing all pixel values, except those in the rightmost, leftmost, uppermost, and lowermost 3 columns/rows with a new value calculated based upon the mask formulation. The pixel with the largest value in this result is detected as the center of the laser spot from the differenced image.

Mask

```
<table>
<thead>
<tr>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
```

Remaining elements are zero.

Figure 3.6 Applying a mask to each pixel provides data regarding its value as well as surrounding pixel values

This laser-spot-identification procedure reliably and robustly establishes the camera space targets under the various illumination, color and texture conditions of the object surface. Laser spots are a powerful tool to help access the visual information of selected junctures of the object surface. And with CSM, the laser spots can be utilized to characterize the object surface prior to being addressed by the robot.
CHAPTER 4:
THREE-DIMENSIONAL LASER-ASSISTED IMAGE ANALYSIS FOR SURFACE OPERATION WITH CSM

The difficulties and limitations associated with the two-dimensional traditional image analysis provide the motivation for what we here term three-dimensional image analysis. The target vision information is specifically for a CSM system. For that particular goal, the feature extraction and location procedure discussed herein are shown to be reliable and robust.

4.1 Difficulties and Limitations in 2D Image Analysis

In order to achieve effective manipulation, the target object or work piece must be characterized geometrically. In a vision-guided robotic system, this characterization is based on the visual information from an object. For example, Figure 4.1 shows a workspace of box-engagement task viewed from three cameras.

The target box position and orientation must be detected and represented in a way that permits calculation of input for successful engagement by the robot. A common approach is to detect the edges of the box in each camera view. Then, combining the edge information in each camera space, to estimate the edges’ 3D information. This characterizes the box surface and provides the box position and orientation to the robot. After the target has been established, the vision-guided robot system would, with a
traditional calibration approach, calculate the set of joint poses that would create a suitable trajectory to approach and pick up the box.

Two-dimensional image analysis provides the algorithms to identify and locate each edge in each image. The goal of traditional edge detection is to mark the points in a digital image at which the luminous intensity changes sharply. Sharp changes in image intensity are often due to changes in the surface geometry of the object. Light-intensity changes could, however, represent discontinuities of the surface coloration or reflectivity, changes in material properties, or variations in scene illumination. So the fundamental problem in traditional edge detection is that the edge detected by 2D image analysis might not be a real physical edge. It could be changes in material properties or scene illumination. Unless these changes are controlled well, the edge detection based on 2D image analysis is not reliable. “Due to the presence of noise and quantization of the original image, during edge detection it is possible to locate intensity changes where
edges do not exist.”[27] The requirement of material surface or scene illumination limits
the applications of edge detection in the visual guidance of robots. Another difficulty
with 2D edge detection is that the intensity-gradient threshold applied to determine edge
presence and location is problematic. A low threshold is susceptible to noise and also to
picking out irrelevant features from the image. Conversely, a high threshold may miss
edges. Slight changes in illumination or the material properties of an object surface will
yield different edge locations in a given image. Edges detected in different camera views
might therefore not correspond with any one, particular physical edge location. All these
things affect the precision of edge detection and finally, therefore, the robot-positioning
precision.

Another limitation of 2D image analysis is that, if distinct edges do not exist on
the object, no useful visual information of the surface can be extracted directly using 2D
image analysis.

4.2 Motivation for Three-Dimensional Image Analysis

The difficulties and limitations of two-dimensional image analysis are a primary
obstacle for applying vision-guided robot technology in the real world. Though robots
may have the dexterity and steadiness to do any given, repetitive job better than a human
in many respects, if the image analysis cannot deliver reliable, precise and robust target
visual information to the robot, even a simple task, such as picking up a box, will not be
possible.

These issues led to the development of new image analysis in three-dimensions
using an approach that complements CSM technology. The target information from three-
dimensional image analysis is independent of changes in illumination or the material properties of the object surface and only relates to the geometric characteristics of the object surface. Another important advantage of doing image analysis in threedimensional space is that it directly uses prior knowledge of three-dimensional geometric characteristics of the object’s surface, which are partially lost after the 3D object is projected into a 2D image plane. This three-dimensional information, for example from a CAD file, would facilitate the reliability and robustness, and enhance the utility of results gained from three-dimensional image analysis.

4.3 Three Dimensional Image Analysis

The first step is to acquire and estimate the 3D positions, relative to the nominal World frame, of laser-spot centers cast onto an object surface. Because of the advantage of CSM 3D shape measurement approach over the other approaches presented in Chapter 1, and the ambient-illumination independence of using laser-spot identification addressed in Chapter 2, the 3D data on an object surface are acquired by casting the multiple laser spots onto the surface and identifying or matching these spots among images from each camera, as shown in Figure 4.2. Then the laser-spot 3D coordinates in the nominal World frame are estimated, as presented in Chapter 3.
These data provide the geometric information of the surface addressed by the robot. This means the robot can position given junctures on its end member at any required place on this surface in high precision.

The second step is to characterize the geometry of the surface based on 3D-coordinate data of the surface points. Because the laser-spot-array direction can be shifted slightly using the pan/tilt unit to cast down new surface spots, allowing for accumulation of a virtually unlimited density of points on the surface region of interest, the characterization also takes advantage of the effect of averaging to filter out the image-discretization and other noise. This characterization is applied either to a previously known model of the object’s surface geometry or to quadratic or other polynomial in order to approximate segmented portions of an unknown surface.

The third step is to analyze the characterized 3D surface to identify the feature of interest for robot positioning or otherwise determine how to operate the robot.
4.4 Edge Detection Based on 3D Image Analysis

Consider for example the box-engagement task. After the 3D coordinates of points on three indicated surfaces of the box are estimated, a plane is fitted to the top, front and side surfaces, as depicted in Figure 4.3. These three surfaces intersect to form edges and the corner of the box as the data is extrapolated. Preferred weight is given to spots near the corner. This stands in contrast with the traditional means of identifying edges directly in 2D images.

![Figure 4.3 Three surface meeting](image)

Figure 4.3 Three surface meeting

There are three advantages of edge detection based on 3D image analysis.

The edge identification procedure is independent of variation of illumination and various materials’ reflective properties; because the edges are the intersection of surfaces and the surfaces are fitted from the laser-spot data, which are independent of lighting conditions. This makes the vision-guided robot run reliably and robustly under real-world illumination conditions, which is generally not achieved using traditional 2D-image edge detection.

The detected edge is more precise, because the intersections of fitted surfaces represent the geometric aspects of interest of the physical object. Frayed or damaged edges would not affect these plane intersections.

The edge-detection results directly represent the 3D geometric characteristics of the physical object. Prior knowledge of an object’s geometry can be utilized to falsify the
edge detection results. For example, the three edges of a cuboid-shaped box should be physically perpendicular to each other. By checking angles among three detected edges one can diagnose an incorrect result. This diagnosis makes the system robust. Moreover, the geometric characteristics can be treated as constraints in surface characterization to reduce the number of parameters needed to be fitted in a surface model. A smaller number of parameters of the model needed to be fitted results in the less sensitivity to noise in the data and thereby reduces the required quantity of data.

The example application of edge detection based on 3D image analysis and geometric characteristics is presented in Chapter 6.
CHAPTER 5:

FUNDAMENTAL EXPERIMENTS OF LASER-SPOT-ASSISTED, 3D IMAGE ANALYSIS

Using 3D image analysis based upon use of laser spots, CSM can control the robot in such a way as to address very precisely an arbitrarily placed surface. Two experiments presented in this chapter demonstrate the precision of the laser-spot-assisted, 3D image analysis. The experiments are used to describe the details of how to achieve this high precision. The results of two experiments prove the high-precision ability. Also they lay the foundation of the developed laser-spot-assisted, 3D image analysis for general, real-world application.

5.1 Thickness-Reduction-Gauging Experiment Based on Laser-Spot-assisted, 3D Image Analysis

5.1.1 Objective of Experiment

As described in Chapter 4, laser-spot-assisted, 3D image analysis characterizes the geometry of the object surface based on 3D-coordinate data of the surface points, and subsequently identifies the feature of interest for robot positioning. Thus, the precision of the characterized geometry of the surface limits the precision of CSM to control the robot relative to the surface of the physical object.
The surface-reduction-gauging experiment applies laser-spot-assisted, 3D image analysis to gauge thickness changes of a surface. The physical thickness change after removal of a flat, thin layer of material is measured by caliper and compared with the thickness change estimated using laser-spot measurement.

5.1.2 Experimental Setup

The robot-vision system that we set up for the surface-reduction-gauging experiment consists of a Kawasaki JS5 six DOF robot, a personal computer, three off-the-shelf CCD, monochrome, analog, industrial video cameras (JAI model CV-M50) and one single-dot laser pointer and one laser-grid pointer mounted on a pan/tilt unit. The cameras are connected to a frame grabber board (Data Translation model DT-3152), which is installed in the computer. The laser pointers cast laser spots onto the object surface. On/off of the laser pointers which is controlled by a digital I/O board (CyberResearch model CYDIO-24) installed in the computer. The pan/tilt unit is a computer controlled 2-DOF mechanism. It carries the two laser pointers to illuminate the object surface and accumulates enough density of laser spots on the surface by shifting the projected laser grid.

The system configuration is shown in Figure 5.1. Three static cameras mounted on the ceiling, about 3 meters away from the work space, which is contained within a volume that is approximately a 1000 mm (width) by 1000 mm (length) by 1000 mm (height). The pixel resolution of the camera is 640 by 480 pixels. Each pixel represents about 2mm projected from physical space.
The work piece close-up image is shown in Figure 5.2.

Two pieces of square aluminum plates were stacked on the top of each other and placed inside the workspace volume. The thickness of the top plate was measured both by calipers and laser spots. In order to test the precision of the laser-spot measurement, the
variations that exist in the plate surfaces should be very small. The two 50 by 50 mm square pieces of aluminum were used in the experiment were cut and then machined to have less than 0.01 mm surface variation. The thickness of the plate is 3.20 mm ±0.01 mm.

5.1.3 Experimental Procedure

The procedure of experiment for assessing the accuracy of this approach for gauging surface reduction is briefly described here. As shown in Figure 5.3, two flat plates of aluminum, with known thickness, are placed in the workspace.

Step 1: Laser spots are cast down onto the top surface of the top plate.

![Figure 5.3 Laser spots fall on the top surface of the top plate](image)

Using image differencing, detection and matching process described in Chapter 4, the 2D coordinates of the laser-spot centers on the top surface of the top plate are identified in three CSM cameras.

Step 2: The top plate is removed as illustrated in Figure 5.4. So the thickness change of the surface is the thickness of the top plate. Laser spots are cast onto the top
surface of the remaining lower plate. The 2D coordinates of the laser-spot centers on the
top surface of the lower plate are identified in the same three CSM cameras.

Figure 5.4 Laser spots fall on top surface of lower plate

Step 3: After the two sets of laser-spot 2D-coordinate data are stored in computer
memory, cue-bearing plates mounted on the robot are introduced into the same physical
region as the two stacked plates, as shown in Figure 5.5

Figure 5.5 Cue-bearing plates approach workpiece
Step 4: The newly sampled cues and the joint coordinates of the robot in corresponding poses were applied to update locally the mapping of camera-space end member 2D coordinates to robot joint coordinates as described in Chapter 3. Then, with the new local mapping model and the two sets of laser-spot 2D-coordinate data in each camera space, the 3D coordinates of surface spots were estimated relative to the nominal world frame discussed in Chapter 3.

Step 5: The surface points on the top plate are fit to a plane using least squares. The distance from the plane to the surface points on the lower plate are calculated. The average distance is the estimate of the thickness of the removed plate.

In step 1 and step 2 the number of laser spots cast on the top surface are varied from 100 to 2000 in various tests. The laser-spot-array direction can be shifted slightly using the pan/tilt unit to cast down new surface spots, allowing for accumulation of a virtually unlimited density of points on the surface regions of interest. The result of varying the number of laser spots is given in the next part of this section.

A change in step 2 for a different test was to place multiple paper cues on the top surface of the lower plate in place of the laser spots, as illustrated in Figure 5.6.

![Figure 5.6 Multiple paper cues on surface](image)
The paper cues placed on the surface are the same circular geometry as those borne on the plate mounted on the robot end effector. The 2D coordinates of the cues were identified in three CSM cameras and nominal 3D coordinates of the cues were estimated as the surface-point data, which was applied in same process as step 5, in order to calculate the thickness of the removed plate. The results of this test are also listed in the next part of this section.

5.1.4 Experiment Result and Discussion

The known thickness of the metal plate is 3.20 mm ± 0.01 mm. The calculated average thickness using about 100 spots (accumulated from multiple images) for estimation is listed in table 5.1.

<table>
<thead>
<tr>
<th>Number of Test</th>
<th>Thickness measured (unit mm) by 100 laser spots vs. 100 laser spots</th>
<th>Thickness measured (unit mm) by 100 laser spots vs. 100 cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.27</td>
<td>3.28</td>
</tr>
<tr>
<td>2</td>
<td>3.15</td>
<td>3.18</td>
</tr>
<tr>
<td>3</td>
<td>3.21</td>
<td>3.18</td>
</tr>
<tr>
<td>4</td>
<td>3.16</td>
<td>3.25</td>
</tr>
<tr>
<td>5</td>
<td>3.17</td>
<td>3.10</td>
</tr>
</tbody>
</table>
The precision of thickness gauging is consistently within one tenth of a millimeter. It is one order of magnitude higher than image resolution. This sub-pixel level of accuracy was consistent throughout the robot’s workspace, and for a range of plate orientations.

*Explanation of sub-pixel level of accuracy*

Why is the laser-spot-assisted, 3D image analysis able to estimate the CSM target shift with sub-pixel accuracy despite the large random error from the coarse pixel quantization of camera space together with error associated with the 3D-coordinate estimation of laser spots in CSM nominal world frame? An error consists of two parts, deterministic error and random error. The deterministic error was cancelled because the thickness assessment is from the difference of two sets of laser spots, which have the same deterministic offset due to the proximity in distance and continuous mapping of surfaces points in camera space to the same body’s position and orientation. Therefore, only the random error was left in thickness gauging results and the each laser spot’s position assessment is virtually zero-mean. With the advantage of the effect of averaging to filter out the image-discretization and other noise, a large number of laser spots could achieve high certainty and precision. The above experimental result is the proof. The histogram summary of 10 trials is shown in Figure 5.7. The data clearly show that the error of thickness gauging is random with normal distribution.
The relationship between density of laser spots on the surface and the accuracy of characterization of surface geometry

The normal distribution of the random error brings up the issue of level of certainty. The practical question pertaining to this certainty issue is what a sufficient number of laser spots cast on the surface of an object needs to be in order to guarantee each individual assessment is within a certain prescribed precision. In another words, what is a sufficient density of laser spots cast on the surface, in spots per unit area on a flat surface in order to characterize the geometry of the surface with designated accuracy? The variation of the error is measured by standard deviation (STD $\sigma$) in statistics. It allows one to deduce the relationship between the number of laser spots (samples) and the certainty of the individual assessment (mean $\mu$) for answering the question. Table 5.2 shows STD $\sigma$ of the individual assessments of plate thickness for the five tests.
TABLE 5.2
STANDARD DEVIATION OF THE INDIVIDUAL ASSESSMENTS OF PLATE THICKNESS

<table>
<thead>
<tr>
<th>Number of Test</th>
<th>Mean (μ) of thickness (unit mm) with 100 sample</th>
<th>STD (σ) of individual assessment of thickness (unit mm) with 100 sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.27</td>
<td>0.772</td>
</tr>
<tr>
<td>2</td>
<td>3.11</td>
<td>0.7653</td>
</tr>
<tr>
<td>3</td>
<td>3.21</td>
<td>0.7383</td>
</tr>
<tr>
<td>4</td>
<td>3.12</td>
<td>0.7347</td>
</tr>
<tr>
<td>5</td>
<td>3.14</td>
<td>0.7732</td>
</tr>
<tr>
<td>STD of the mean of the 5 tests</td>
<td>0.068</td>
<td></td>
</tr>
</tbody>
</table>

As illustrated in Figure 5.8, the individual thickness assessments have about a 68% probability of falling within the range between mean (μ)−STD (σ) and mean (μ)+STD(σ), as expected with a normal distribution.

Figure 5.8 Normal distribution of individual thickness assessment
There is the statistical theorem of Standard Deviation of the Mean [29], which is:

“There exists a random process. The standard deviation of the samples from this process is $\sigma_1$. A second random process is derived from the first one by taking out $n$ samples ($n>1$) from the first process and average their values to be a sample for the second process. The samples from the second random process have a standard deviation of $\sigma_n$. The relationship between the $\sigma_n$ and $\sigma_1$ is:

$$\sigma_n^2 = \frac{\sigma_1^2}{n}$$  \hspace{1cm} (5.1)"

Herein, the predicted ($\sigma_{100}$), STD of 100 samples’ mean in the thickness-gauging experiment result, is about 0.07 from this statistical analysis. The actual STD of the mean of the 5 tests with 100 samples is 0.068, which agrees with the statistical analysis. This relationship also was proven using 2000 spots tests.

**Meaning of cues test result**

In the experiment results of table 5.2, the third column shows that the same high precision of thickness gauging occurred in the test of placing multiple paper cues on the surface of the lower plate instead of casting laser spots after the top plate was removed. This result proves that there is no significant detection bias between the laser-spot detection and cue detection in camera space. In other words, there is no bias between 2D position of the detected center of cue and laser spot in camera space if they occupy same physical location. (The “physical location” of the laser-spot center is not, strictly speaking, actually defined. What the tests indicate is that, on average, the software places each camera’s camera-space center of a given laser spot in such a way as to represent the same surface juncture in all three cameras.) This seems unsurprising, but thinking how different are the appearances of the cue and laser spot in physical space, there are enough
reasons to doubt this no-bias result to warrant physical proof. Among the reasons for a possible bias is the fact that laser spots are incident on the surface from a particular angle, and this angle will vary with any shift in the workpiece or laser-pointer base.

The CSM targets established by laser-spot-assisted, 3D image analysis are ultimately for the robot to position the point of interest on a robot’s end effector with high precision. As described in Chapter 3, CSM’s high precision is fundamentally based on the premise that a point or juncture attached on the robot end effector collocates the target point in 3D physical space when these two junctures collocate in at least two cameras’ camera spaces. So the non-bias detection of position between cue and laser spot is a critical premise to extend the CSM high-precision positioning where laser spots are used to establish camera-space targets. The two-component positioning experiment in [2] provides also a solid proof of this ability.

*Reason for sampling new cues on robot end effector*

In step 3 of the thickness-reduction experiment, cue-bearing plates mounted on the robot were introduced into the same physical region as the two stacked plates. The newly sampled cues and the joint coordinates of the robot in corresponding poses were applied to update local mapping between the camera space object is 2D coordinates and robot joint coordinates. Then, with the new local mapping and the two sets of laser-spot 2D camera-space coordinates stored in computer memory, the 3D coordinates of surface spots were estimated relative to nominal world frame. Though, in the thickness-gauging experiment, the robot doesn’t need to position its end effector to the target with high precision, in real world applications described in Chapter 6, the CSM targets established
by laser-spot-assisted 3D image analysis are used for robot positioning. So the experimental studies of step 3 were performed in same situation as real applications.

5.1.5 Summary and Conclusion

The surface-reduction-gauging experiment proves the ability of laser-spot-assisted, 3D image analysis to characterize the geometry of the surface and provide the CSM target with high precision. It also discloses the extent of measurement precision of surface-reduction gauging and reveals the relationship between the density of laser spots cast on the surface, in spots per unit area, to the accuracy of the characterized geometry of surface.

The experimental results also prove the following three premises of the surface-extent application of laser spots using CSM-based 3D nominal World-frame-coordinate estimation:

Though the nominal optical model and nominal robot kinematics model in the CSM system are globally imperfect, the premise is that as long as the laser spots are close enough to each other within the asymptotic-limit region or volume the relative error between them is close to zero-mean. Therefore, any error on the CSM side does not propagate into the measurement of surface reduction relative to an as-located original surface. The experiment repeatedly proves the premise for a range of surface positions and orientations.

Another premise is that error in thickness assessment with laser-spot data is unbiased and random. This means that, provided they are matched, a laser-spot center detected in each camera corresponds to one single physical juncture on the object’s surface. In another words, only error in thickness assessment with laser-spot data can be
averaged out by applying a large amount of laser-spot data on the surfaces. The results of the above experiment repeatedly verify this premise.

There is no bias between 2D position of the detected center of a circular cue and that of a laser spot in camera space if they occupy same physical location. (The “physical location” of the laser-spot center is not, strictly speaking, actually defined. What the tests indicate is that, on average, the software places each camera’s camera-space center of a given laser spot in such a way as to represent the same surface juncture in all three cameras.)

High-precision surface-change gauging extends the vision-guided-robot system based on CSM into a more potent surface-operating system, one that in practice only can be done by humans, and this in an imprecise, non-uniform, and often ineffective way. A number of surface-finishing tasks entail application of force or pressure combined with in-plane motion in order to achieve a scrubbing, polishing or sanding effect. Such tasks often make use of human dexterity and effort, which can result in repetitive-motion injury and incomplete or uneven treatment of the surface. But with high-precision surface-reduction gauging and CSM, our vision-guided robot system can accomplish these tasks efficiently and uniformly. Material removal can be monitored and surface reduction gauged and controlled to within approximately one tenth of a millimeter.

The example applications of the sanding project are presented in Chapter 6.
5.2 Five-Component Positioning Experiment

Previous experimental work [2] has been directed toward verifying the high level of positioning precision across a large span of work space that CSM provides relative to a single flat surface. In experiment of [2] the out-of-plane translational error and one of the two out-of-plane orientation errors were measured. The evidence found in that experiment motivates this five-component positioning experiment, which accesses the out-of-plane and in-plane translational precision as well as both of the two components of out-of-plane orientation as well as in-plane orientation. The experiment is proof that the high level of precision of the surface operation provided by the laser-spot assisted 3D image analysis extends to full 3D rigid-body positioning. The experiment also unveils the multiple-component observability of CSM for high-precision positioning, including orientation.

5.2.1 Experimental Setup

The robot-vision system set up for the five-component positioning experiment consists of a Kawasaki JS5 six DOF robot, a personal computer, three off-the-shelf CCD monochrome analog industrial video cameras (JAI model CV-M50) and one single-dot laser pointer and one laser-grid pointer mounted on a pan/tilt unit. The cameras are connected to a frame grabber board (Data Translation model DT-3152), which is installed in the computer. The laser pointers cast the laser spots onto the object surface, the on/off of which is controlled by a digital I/O board (CyberResearch model CYDIO-24) installed in the computer. The pan/tilt unit is a computer-controlled 2-DOF mechanism. It directs
the two laser pointers to illuminate the object surface and accumulates an adequate
density of laser spots onto the surface by shifting the laser grid.

The system configuration is shown in Figure 5.9. Three static cameras are
mounted on the ceiling about 2 meters away from the work space, which extends
approximately over 1000 mm (width) by 1000 mm (length) by 1000 mm (height). The
resolution of each camera is 640 by 480 pixels. Each pixel represents about 2mm in
physical space. The accuracy of the kinematical calibration of the robot mechanism is
typical. The kinematical error in the workspace is several millimeters.

![System configuration](image)

Figure 5.9 System configuration

The objective of the experiment is to position the two surfaces on an L-shape end
effector parallel to “surface one” and “surface two” on the object at a specified separation
as shown in Figure 5.10. The work piece is two 4 by 4 inch flat metal plates attached to
each other perpendicularly.
Figure 5.10 Position the end member (L-shape) onto surfaces 1 and 2

One aspect critical to achieving high-precision positioning is extraction of the position and orientation of the physical crease between the two flat metal plates, the edge or intersection line shown in Figure 5.11.

Figure 5.11 Intersection line

We characterize the geometry of the two orthogonal surfaces in order to extrapolate to find the edge. The direction of the edge is the cross product of the normal vectors of the two planes.

5.2.2 Experimental Procedure

The 3D image-analysis procedure in the two-surface position experiment consists of the following steps.
Step 1: Point-and-click

Four points in each of two surfaces were selected in the selection camera via a graphical user interface as shown in Figure 5.12. The points demark the two areas of interest on the object surfaces. With the point-and-click method, the surface points are identified in terms of two-dimensional camera-space regions as they would appear in three participant CSM cameras.

![Figure 5.12](image)

Step 2: Illuminate surface junctures and match the spots

Several blasts of the laser grid are cast onto the two surfaces; the pan/tilt unit shifts the laser grid a little between the blasts to make the laser spots distribute, as shown in Figure 5.13. The identified laser-spot centers are accumulated in each participant CSM camera. The detected laser spots in each camera space are matched to produce correspondence of the multiple laser spots in the three cameras.

![Figure 5.13](image)
Step 3: 3D coordinate estimation and spot segmentation

After the spots are matched, 3D coordinates of the spots are estimated in nominal World-frame coordinates. Figure 5.14 shows that the spots are segmented into two groups in order to determine which spots lie on which physical plane of interest, horizontal or vertical. Matched laser spots and local camera-space kinematics combine to allow for a three-dimensional understanding of individual-spot coordinates. This helps with the segmentation problem in two ways: It removes from consideration surface coloration and ambient lighting. And it provides a three-dimensional basis for sorting the spots according to surface region.

Step 4: Fit the edge and orientation to estimate the target

The two groups of 3D surface points are fit to two planes using least square. The intersection line of two planes represents the physical juncture between two flat metal plates. A selected point on the edge provides the position of the target of interest and the normal vectors of two planes provide the orientation of the target of interest, thus
providing the needed pair of coordinates for rigid-body positioning as shown in Figure 5.15.

Figure 5.15 The normal vectors of two planes and intersection line

Step 5: robot roughly positions the end effector to approach target

Three cameras guide the robot such that the L-shape end-effector approaches the target estimated in step 4, as shown in Figure 5.16.

Figure 5.16 The end effector approaches target

Step 6: Sample cues on the end effector en route to the terminus

The L-shape plate bears the cues, which are acquired in the three camera spaces en route to the terminus. The six joint coordinates of the robot are simultaneously sampled in order to synchronize with image acquisition for detecting
cues as shown in Figure 5.17. The cues are judiciously placed on two surfaces on the
L-shape end effector for observability as discussed later.

Figure 5.17 Sample the cues on end effector

Step 7: High precision terminal positioning

The newly sampled cues and corresponding joint coordinates of the robot are
applied in order to update locally the mapping in camera space of the object to robot joint
coordinates. Then, with the new local mapping model, the 3D coordinates of surface
spots are re-estimated relative to the nominal World-frame coordinates. Thus the target of
interest is refined in position and orientation. Then the robot positions the end effector to
the target with high-precision position and orientation as shown in Figure 5.18.

Figure 5.18 Robot positions the end member on two surfaces
5.2.3 Experimental Result

1. Measurement

The cue-bearing L-shape plate positioning was specified to achieve 20-millimeter separation for both horizontal vertical surfaces. Figure 5.19 shows five measurement locations, which are 90 millimeters apart, between the end-effector and work-piece surfaces; they were measured in order to access five positioning components, the error of out-of-plane position, in-plane position and out-of-plane orientation about the X, Y and Z axes, respectively.

![Diagram of measurement setup](image)

Figure 5.19 Measurement

2. Measurement results

In two surface-positioning experiments, the fixture with two target surfaces, which are perpendicular to each other, were placed within the robot’s workspace with arbitrary angles as shown in Figure 5.20.
Table 5.3 shows mean and standard deviation of the error of the five components in 20 tests.

The target point is on the surface 1 (horizontal surface). Figure 5.21 shows the definition of out-of-plane and in-plane component positions and orientations.
**TABLE 5.3**

**EXPERIMENTAL RESULTS**

<table>
<thead>
<tr>
<th></th>
<th>Error of out-of-plane position (unit mm)</th>
<th>Error of in-plane position (unit mm)</th>
<th>Error of out-of-lane orientation about axis X (unit mm)</th>
<th>Error of out-of-plane orientation about axis Y(unit mm)</th>
<th>Error of in-plane orientation about axis Z(unit mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>0.11</td>
<td>0.15</td>
<td>0.12</td>
<td>0.14</td>
<td>0.16</td>
</tr>
<tr>
<td><strong>STD</strong></td>
<td>0.2</td>
<td>0.25</td>
<td>0.21</td>
<td>0.23</td>
<td>0.26</td>
</tr>
</tbody>
</table>
5.2.4 Discussions

The result of this five-component positioning experiment proves the ability of CSM with laser assisted 3D image analysis to deliver the same level positioning precision as the one-surface positioning in [2]. There are a number of details involved in conducting the experiment to achieve such high precision, which are disclosed in the following discussion.

1. Robot kinematical error and observability

The global kinematic accuracy of an industrial robot, such as the Kawasaki JS5, typically is within only 5mm to 25mm due to the deviation of production and assembly tolerances from nominal (CAD model). In the CSM vision-guided robot system the kinematical error is absorbed into CSM’s locally refined camera-space-kinematics parameters, because CSM achieves the local accuracy of positioning the point of interesting on the end effector relative to target surface points by making a mapping relationship directly between robot joint rotations and appearances (cues as shown in Figure 5.22) of the manipulated body in each image space.

Figure 5.22 Cues on end member
For example, the Z (vertical) component of kinematical error of the Kawasaki JS5 was easily observed in a test to manipulate the robot in linear motion with fixed Z coordinate (nominal). With the end effector moving close to the base, the separation between the end effector and a horizontal flat surface was increased. This test clearly shows the effect of kinematical error in this JS5 robot.

In CSM, with the new the cues on the end effector and robot joint coordinates sampled when robot is en rout to the target surface, the nominal coordinates of the surface points are refined, which compensates for the global kinematical error. As shown in Figure 5.23, the values of the Z component of the nominal coordinates of the points on a flat surface placed horizontally are not the same. There is bias between the Z component of the points close to and more distance from the base of the robot. It is consistent with the observation of the drift related to the kinematical error in this JS5 robot. In order to accurately positioning the end effector relative to the surface, the end effector needs to be moved lower in the region close to the base in the nominal World frame to offset the kinematical error, which pulls the end effector higher in the region close to the base.

Figure 5.23 Variation of the values of Z component of the nominal coordinates of the points on a flat surface
The continual changes of the Z components of surface points on a horizontal flat surface in the nominal world frame reflect the continuity of kinematical error over the region. Such gradual, continuous variation is good for precisely controlling the Z component in positioning. However, the continual changes of the Z components resulted in the small angle between the plane fitted by surface points and the X-Y plane of nominal world frame. The inclination of the surface causes the error of out-of-plane orientation. In order to reduce the error of out-of-plane orientation, the plane needs to be fitted by the surface points within a smaller asymptotic-limit region with respect to the target point.

Figure 5.24 Cues on horizontal plate

Also, since the two cues in Figure 5.22 were placed on a vertical plate, the observability of critical components of camera-space kinematics based on the appearances (cues) of the manipulated body in image space is limited. Two cues placed on a vertical plate were only good to control rotation about the axis normal to the vertical plate. More cues are added, therefore, on a horizontal plate (as shown in Figure 5.24) to
increase the observability to control the rotation about the other axis. Therefore, two out-of-plane orientation components were well controlled.

2. Weighting of the surface points

The high precision of the fit of the edge, which is the intersection line between two surfaces, was achieved by a weighted least-squares method. The more distance there is of the surface points to the edge, the less precision of the fit of the edge because of the limited extent of the asymptotic region as described in Chapter 3. In the edge fitting process the surface points closer to the edge were given more weight to estimate the parameters of the intersection line. In the two-surface positioning experiment, the weighting scheme could be varied across a large region, within which the high-precision positioning was consistent. However, too heavy of a weighting of points near the edge applied could cause the precision of the edge fitting to be very sensitive to random noise of the points near the edge with consequent loss of the advantage of averaging out the random noise using a large number of samples.

5.2.5 Summary and Conclusion

It is shown experimentally in this section that the laser-assisted 3D image analysis with camera-space manipulation is able to perform high-precision surface operation. The in-plane and out-of-plane translational error is within 0.3 mm. The mean of error magnitude is under 0.1 mm, which are two orders of magnitude better than robot accuracy and one order of magnitude better than the camera resolution. The in-plane and out-of-plane orientation error is within 0.4 degree, and the mean of angle-error magnitude is under 0.1 degree.
The observability issue was examined through the experimental studies. Newly added samples of cues placed upon a surface with unit normal perpendicular to that of the original plate, in the vicinity of the maneuver terminus, improve the orientation precision. The positioning precision using various weighting formulae to fit the edge based on the distance to the edge were examined. The positioning precision was improved by giving more weight to points near the edge. However, excessive increase of weighting of samples near edge results in sensitivity to random sampling noise.

The conclusions presented in this section were made based on the observation of hundreds of tests across a large span of work space and with various angles of the work piece. They lay the foundation of the extension of the laser assisted 3D image analysis to more general applications, which are presented in Chapter 6.
As discussed in Chapter 4, the difficulties and limitations associated with the conventional 2D image analysis are partially responsible of lack of progress for vision-guided robotics in real-world applications. In this chapter three types of general application, which challenge robotics in industry today, show how the present 3D image analysis method enable high precision and robust robot positioning with CSM.

6.1 Five-Component Peg-In-Hole Assembly

6.1.1 Background of the Peg-In-Hole Assembly Application

The peg-in-hole assembly application represents a generic assembly process of mating two parts with peg-hole couplings. Such a task is one of the most common in manufacturing. For example, in the automotive industry: piston assembly into an engine, assembling of a transmission clutch, torque converter into a transmission housing, etc.

Those tasks are simple, and any normal human being can perform them. Even for mating some heavy and big parts, some labor-saving mechanisms can bear the weight of the part which is directed into place by the human worker’s remarkable ability to guide closure visually to accomplish the task. However, the peg-in-hole assembly task is very
difficult to have carried out by a robot. As described in Chapter 1 the majority of industrial robots are used in the teach-repeat mode, if a hole is not exactly in the location as robot was taught, a robot would be unable to mate the parts successfully. Many researchers have repeatedly attempted to apply vision-guided robotics to have a solution for it. But none of them is suitable for overcoming the problem in the real world as pointed out in Chapter 2 and Chapter 4. Although there are a few industrial solutions provided by some of the biggest industrial robot manufacturers, the intrinsic problems of the systems limit the range of the applications and performance, and human guidance is most often applied.

6.1.2 Current Alternative Industrial Solutions

1. Vision guided robot system solutions

One of the solutions is single-camera 3D developed by Braintech 3D, which is the exclusive vision-guidance solution provider for ABB, one of the world is largest industrial robot manufacturers. It is a calibration method of vision-guided robotics as introduced in Chapter 2. As shown in Figure 6.1, a CCD video camera mounted on the robot end-effector (the robot hand).

![Figure 6.1 Camera on the robot end-effector](image-url)
The robot first moves the end-effector with the camera to a taught pose which is close to the work piece. Then camera acquires a picture of the surface of the work piece and the position and size of the features in the 2D image are detected. Based on a calibrated perspective camera model, as illustrated in Figure 6.2, the distance from the camera to the surface of the work piece is estimated. Finally the system uses this information to guide the calibrated robot’s hand to insert the part into the hole.

![Figure 6.2 Perspective camera model, as example of depth inference](image)

For example, in the following engine-head assembly application show in Figure 6.3, the gripper on the robot has four pegs which need to be inserted into the holes on the engine head in order to grasp the part. The camera and gripper are moved to the pickup position over the conveyer, where the engine head appears. In an image from the camera the position centers of the holes and diameters of the holes are detected. The distance from the camera to the top surface of the hole is estimated based on the distance between the holes and the diameters of the holes. After that, the x and y coordinates of the holes are estimated based on 2D image calibration.
This single-camera 3D vision-guided system has the common problems of calibration as described in Chapter 2. The first problem is that the robot and camera have to be calibrated frequently in order to maintain accuracy of positioning. This need affects the cycle time of the assembly process, which is a critical issue in the automotive industry. Second, the parts must have several features, which have previously known sizes or relative separation, in order to estimate depth using the camera model. This requirement is prohibitive in more general cases. Even for the present engine-head assembly case, if the engine head is made by founding, the variation of the size of the hole during the cooling process reduces the precision of position. Also lighting and texture raise the reliability issue of 2D image-feature detection.

Another vision-guided system is a 3D laser vision system developed by Fanuc, the second largest industrial robot manufacturer. A 3D laser sensor is mounted on the robot end effector as shown in Figure 6.4. The 3D laser sensor based on the triangulation method as described in Chapter 2 to estimate the distance of the point on the laser line to the camera. The orientation of the work piece surface is estimated by the four intersection points of two laser lines with the detected feature, such as circle.
This solution is still a calibration method. It has all of the common problems of the calibration method. On the target side, though it applies a laser line, which helps eliminate the problems from variation of lighting environment, the needs of 2D image feature detection are still vulnerable to changes of the lighting environment.

Besides the robustness and reliability problems discussed above, the precision of both vision-guided systems is limited by the image resolution and robot kinematic error. They cannot generally perform assembly tasks that require high precision, and are seldom used.

2. Force control system solution

An alternative means of vision-guided robotics in assembly is where the robot is guided by force control. One force control system is developed by ABB Robotics. As shown in Figure 6.5, a force sensor is installed between the robot and the end-effector.
After calibration, the gravity load of the end-effector is compensated for in the force sensor frame, which means the robot only responds to the contact force applied on the end-effector in this force control system. In the peg-in-hole assembly process, the peg is moved to the taught start position as illustrated in Figure 6.6. Then control of the robot is switched from teach/repeat to force control. The peg is moved down along the Z axis slowly with a spiral search motion in the X-Y plane. If the peg touches the edge of the hole and contact force registers over a certain designated value, the movement along the Z axis stops and the spiral search motion finds the hole. Once the peg moves along the Z axis beyond the pre-set engage distance, which means the peg has been successfully inserted into the hole, the peg quickly moves to the bottom of the hole to complete the insertion process.
Force-control guidance applies active and passive compliance such that the manipulator is able to insert the peg into a hole, the exact location of which is unknown. However, the performance of the assembly process is not competent. Due to inefficient force-based searching, the variation of the cycle time is large. Also the response of the force control system to the different contact properties of the surface limits maximum assembly speed. All of these reduce the utility.

6.1.3 Experimental Objective

The peg-in-hole assembly task requires robot positioning of one part relative to the surface of other part. And CSM is an excellent alternative for positioning a rigid body relative to as-located surfaces. The two-surface positioning experiment of Chapter 5 demonstrates this. The key problem in the present peg-in-hole experiment (a cylinder was inserted into a hole in a cylinder receptacle; Figure 6.7 shows the two parts, which are to be mated) is to provide precise position and orientation of the target from a non-flat surface (cylindrical surface here).

Figure 6.7 Two insertion parts
Therefore, the peg-in-hole assembly-task experiment is used herein to examine the laser-assisted 3D imaging with camera-space manipulation. It also demonstrates that the high level of positioning precision extends to non-flat surfaces.

6.1.4 Experimental Setup

In the peg-in-hole test, components of the vision guidance system have the same configuration as the two-surface positioning experiment. Figure 6.8 shows the peg, or cylinder part, mounted on the robot end-effector. The cylindrical work piece containing the hole is placed on a flat surface, as shown in Figure 6.8.

![Figure 6.8 Peg on robot end-effector](image)

6.1.5 Experiment Procedure

Step 1: Laser spots are cast onto the surfaces of the cylindrical receptacle and flat supporting surface as shown in Figure 6.9. The laser spots are identified and matched in three cameras. The nominal 3D coordinates of all matched surface points are estimated.
Step 2: The laser spots that fall on the flat plate surface as well as the outer wall and inner wall of the hole are segmented as shown in Figure 6.10. The plane is fitted using least squares application of the points on the flat plate. The norm vector of the plane provides the target out-of-plane orientation. The out-of-plane position was calculated by the known height of the cylinder and the norm vector. They control the robot’s terminal depth of insert action and insert-part direction.
Step 3: The nominal 3D surface points on the outer surface and inner surface of the cylindrical receptacle are projected onto the flat surface plane x-y as shown in Figure 6.11.

![Diagram](image)

Figure 6.11 Laser spots on outer and inner surface of the cylinder receptacle

The physical dimensions of the cylinder are known. The outer-surface points and inner-surface points are concentric. The center of the hole is the same as the center of the outer-cylinder surface. Let coordinate $X_0$ and $Y_0$ be the center of the hole. The radius of the hole and the radius of the outer cylinder are known. The surface points on inner and outer surfaces are constrained according to the following:

\[(x^i - X_0)^2 + (y^i - Y_0)^2 = r^2 \quad (6.1)\]

\[(X^i - X_0)^2 + (Y^i - Y_0)^2 = R^2 \quad (6.2)\]

Where $x^i$ and $y^i$ are the coordinates of the inner surface points, $X^i$ and $Y^i$ are the coordinates of the outer surface points, $r$ is the known radius of the hole, $R$ is the known radius of the outer cylinder.
The coordinates of the center were estimated by minimizing the following function $J$.

$$J(X_0, Y_0) = \sum_i [(x^i - X_0)^2 + (y^i - Y_0)^2 - r^2] + \sum_j [(X^j - X_0)^2 + (Y^j - Y_0)^2 - R^2]$$ (6.3)

Once the coordinates of the center are estimated, the 3D coordinates of these spot centers relative to the aforementioned nominal physical frame can be fit precisely to a geometric model of the feature, thus providing the needed pair of coordinates for rigid-body positioning. (For this 5DOF task, in-plane orientation need not be controlled.)

Step 4: The peg with the cues on end-effector approaches the center of the hole. The local model mapping the 2D image coordinates to joint coordinates is refined using the newest, most local samples of the end-of-arm cues and joint coordinates of the robot. Then the 3D coordinates of the surface points are refined and the Step 2 and Step 3 repeated to refine the precision center of the hole in the nominal World frame. This coordinate system has an important attribute attained through the use of CSM: as described in Chapter 3, it is the coordinate system relative to which the nominal kinematics model – the same model given by the manufacturer – is locally virtually perfect. The word “locally” here means the local region of joint space, physical space and time near which the maneuver culminates.

Step 5: Robot was controlled to guide the insertion peg to align with the center of the hole and a precise depth with precise orientation relative to axis of the hole. The three components of position and two components of orientation or five degrees of robot freedom were controlled.
6.1.6 Experimental Result

The insertion tests have been successfully conducted across a large subspace of the robot’s workspace as shown in Figure 6.12. It is in front of the JS5 robot. The size of subspace is approximately 500 mm (width) by 500 mm (length) by 250 mm (height). The out-of-plane angle between the flat surface, where the cylindrical receptacle is placed, and the horizontal table (i.e., the axial tilt relative to the horizontal) varies in the $20^\circ$ conic range. The pointing direction ranges across the full $360^\circ$ from test to test.

The rim of diameter of the outer surface of the peg is 89.8 mm and the diameter of the inner surface of the hole is 90.3 mm. The system consistently demonstrated adequate precision to perform this tight tolerance assembly task, which means the poisoning errors are smaller than the gap between outer surface of the peg and inner surface of the hole.
The positioning results are reliably and without exception consistent with the same level sub-pixel positioning accuracy of the two-surface positioning experiment.

6.1.7 Conclusions

The successful insertion in this tight tolerance peg-in-hole assembly experiment proves the extendibility of high-precision surface operation based on the laser-spot assisted 3D image analysis with CSM to one particular non-flat surface. The result also establishes a robust and practical vision-based solution for a typical peg-in-hole type of robotic assembly application. The high precision, reliability, robustness and high efficiency provided by the present system are much better suited to real-world application than are existing industrial solutions.

6.2 Non-Flat Surface-Finishing

6.2.1 Background of the Surface-Finishing Task

A number of surface-finishing tasks entail application of force or pressure combined with in-plane motion in order to achieve an abrasive, scrubbing, polishing or sanding effect.

Figure 6.13 At various scales, from a small mold to a ship’s hull, human action is often needed in order to effect the required finishing process.
As indicated in Figure 6.13, the motion could be applied to restoration of a ship hull, polishing of a mold or form, and sanding of a building façade. Such tasks often make use of human dexterity and effort, which can result in repetitive-motion injury and incomplete or uneven treatment of the surface.

This control problem can be separated into two parts: in-plane position and orientation control and out-of-plane force control and orientation control. The latter is based upon the need for both even contact and a reasonably well-controlled normal component of force. The in-plane piece entails considerations of evenness of coverage across the desired area or region.

6.2.2 Objective of the Sanding Experiment

The present experiment is based on the laser-assisted 3D image analysis and camera-space manipulation. The sanding task is a general application extended from the thickness-reduction-gauging experiment and two-surface positioning experiment. It shows how robotic motion is used to bring about the desired results using the robot’s artificial dexterity in lieu of direct human action.

6.2.3 Setup of the Sanding Experiment

The sanding experiment has a setup to similar the thickness-reduction-gauging experiment. A dexterous robot, as indicated in Figure 6.14, represents an alternative to direct human application of force normal to the surface combined with tangent-plane movement parallel to the surface. A sanding sponge is attached to the end of the robot. The wood board clamped on a fixed table in the workspace is the object being sanded.
6.2.4 Experimental Procedure

This six-axis robot has the dexterity required to apply a sanding sponge evenly across the surface within its workspace. The key problem is how to control the in-plane position and orientation of the sander as well as the out-of-plane force and orientation. The latter requirement is based upon the need for both even contact and a reasonably well-controlled normal component of force. The in-plane part entails considerations of evenness of coverage across the desired area or region. To achieve these goals, the geometric characteristics of the object surface need to be understood. We achieve this understanding in a manner that is synergistic with operation of the vision-guided robot. The method makes use of 3D image analysis using laser spots. Results from this surface characterization determine the robot motion for bringing about the desired results using the robot’s artificial dexterity in lieu of direct human action.

The 3D image analysis procedure in the sanding task works synergistically with the robot control in the following way.

Step 1: Figure 6.15 shows that a human operator may designate using a Graphical User Interface the surface regions that must be treated. With the point-and-click
technology and laser spots, the surface-region information is understood by the vision-guided robot system. Basically the surface information that results is made accessible to the participant cameras through automatic pan/tilt control of a laser pointer that is turned on and off in order to synchronize with cameras’ image acquisition in order to allow for robust laser-spot detection via image differencing.

Step 2: Figure 6.16 shows multiple laser spots that are automatically cast onto the selected surface regions. Laser-spot centers are identified and nominal World frame coordinates of these spots are estimated. The surface regions are characterized.

Figure 6.15 Select the points to specify the region of interest, using computer monitor

Figure 6.16 Multiple lasers are cast onto the surface and identified
Step 3: Based on this surface information, the initial robot motion is planned, and the robot acts on the surface with both in-plane and out-of-plane components well controlled. There is no use of active force control. Force magnitudes normal to the surface are controlled indirectly through computer-vision-determined “interference” – specified in millimeters – between finishing-tool pre-contact surface and work piece pre-contact surface. In Figure 6.17 the left-most image represents the un-deformed sanding tool approaching the un-deformed surface. The middle image indicates how, in its full-contact pose, the tool would appear relative to the surface absent deformation due to contact. “Interference” is defined from this idealization. The final picture shows deformation of both tool and surface. An assumed stiffness of the combination transforms “interference” in mm into force in Newtons.

![Interference Illustration](image)

**Figure 6.17 Illustration of interference between a sanding block and workpiece surface.**

Step 4: Multiple laser spots are automatically cast onto the selected surface regions again. The surface regions after first sanding are characterized geometrically. Surface reduction that has resulted from application of previous action on the surface is calculated typically to within tenths of a millimeter relative to the prior surface contour prior to the start of the finishing procedure surface reduction is gauged using the 3D image analysis approach of the previous chapter. By monitoring the progress of this
surface reduction, future robot-delivered action can be planned automatically and the interference is judiciously adjusted region by region across the surface. The nominal kinematics equations of the robot are characterized and used, but precision of the robot’s actual positioning must be much more accurate than these global kinematics relative to the as-located surface. Material removal is monitored and surface reduction gauged and controlled to within approximately one tenth of a millimeter until the target surface reduction is reached everywhere.

6.2.5 Experimental Results

Actual sanding tests were performed on a variety of flat surfaces located across the full extent of the domain of the robot’s workspace. None of the tests involve careful positioning of system elements of cameras, workpiece, robot or laser-pointer-bearing pan/tilt unit. And none of these elements’ physical position were assessed or calibrated. Therefore the system well represents the case where a robot is introduced in a convenient way into a setting where the surfaces of interest cannot be moved. The control scheme should perform robustly with a level of precision consistent with the tests described herein. Actual surface reduction due to abrasion of the sanding process is assessed using uncalibrated system elements, including uncalibrated cameras, and a sanding action to compensate for departure of the most recently sensed extent of surface removal from a user-specified reference.

The thickness of the original wood board is about 11 mm. The desired thickness of removed material is 1.5 mm. After the robot conducts the sanding operation and laser-spot assessment reveals that the 1.5 mm extend has been reached, the wood board is cut
into several pieces in order to measure the new, reduced thickness in the broad central region of sanding of the board using caliper as shown in Figure 6.18.

![Figure 6.18 White crosses on wood board indicate where measurement occurred](image)

The new thickness of the wood board is 9.5±0.1 mm consistently over the 350 mm (length) by 90 mm (width) selected area of interest. The out-of-plane angle between the wood board and the horizontal table varies in 20° conic range. Considering the original 11 mm board width, that represents material removal of 1.5±0.1 mm.

6.2.6 Conclusions

The present experiment describes a system and scheme whereby robotic motion is used to bring about the desired results using the robot’s artificial dexterity in lieu of direct human action. The system has the following unique characteristics:

1. It is visually guided; however, neither the robot nor separate cameras require calibration. This offers the advantage that larger-scale operations can be accomplished by introducing cameras and robot into the setting without a calibration event for either.
2. Once the cameras have been directed toward the surface regions of interest a human operator may designate by Graphical User Interface the surface subdomains that must be treated, as those regions appear in any given one of the minimum of two introduced cameras. The surface information that results is transferred automatically to the participant cameras through automatic pan/tilt control of a laser pointer which is turned on and off in order to synchronize with cameras’ image acquisition to allow for robust laser-spot detection via image differencing. Prior positioning or locating of the pan/tilt unit that bears the laser pointer, as with the cameras, need not be characterized or measured – as long as the pointer has direct line of access to pertinent surface junctures.

3. The out-of-plane objectives of the task may be stipulated in terms of the amount of material to be removed. Provided neither camera nor workpiece positions change between the start and the finish of the operation, material-removal precision may be realized down to a tenth of a pixel of camera resolution. In the experiments reported herein this translates to about a third of a millimeter.

4. There is no use of active force control. Force magnitudes normal to the surface are controlled indirectly through visually determined “interference” – specified in millimeters – between finishing-tool pre-contact surface and workpiece pre-contact surface. The effective stiffness, for example in Newtons per unit of interference in mm, can be stipulated apriori or estimated using a single force sensor as it changes gradually over varying robot pose and/or workpiece region.
The present sanding application represents various surface machining applications. The experimental results show that the present laser-assisted 3D image analysis with CSM is competent to this surfacing category of task which requires gauging of surface reduction without of prior knowledge of the original surface contour geometry. This should permit needed robot motion for surface treatment such as deburring, cleaning, painting, polishing, sanding and drilling.

6.3 Palletizing and De-palletizing Applications

6.3.1 Background of Palletizing and De-palletizing Project

In the paper-container industry, at the end of each stage-of-production line, paper bags need to be stacked layer by layer according to some pattern, as shown in Figure 6.19 for storing and transporting. Eventually, the stack of bags needs to be un-stacked layer by layer and fed into a machine to undergo the next procedure in fabrication, or to be packed into a box. This is very labor-intensive work. It is hard for companies to find enough people to fill these positions. Also the repetitive stack and un-stack work can easily cause back and waist injury. Therefore, the robot palletizing and de-palletizing system were developed.

Figure 6.19 Pattern of bag stacking
6.3.2 Palletizing System

Each hand of bags is required to drop onto the current (and building) top of the stack from a certain distance for good stacking quality. However when the bags are stacked onto each other, the thickness of each hand of bags depends upon the weight of the bags above it. With this unknown elevation of a stack, the robot cannot perform the stacking task in teach-and-repeat mode. A vision-guided robot palletizing system was therefore installed based on the laser-assisted camera-space manipulation method. As shown in Figure 6.20 a bank of laser pointers, which was mounted on the ceiling of the factory over the bags stack, projected the laser spots straight down onto the top surface of the stack. The elevation of the stack was estimated after each laser spot was detected and the 3D coordinates of the center were estimated. Then the robot, guided by CSM, moves the gripper to the required drop position in order to stack bags layer by layer.

Figure 6.20 Bank of laser pointers

6.3.3 De-palletizing System

The de-palletizing task is more challenging to automate than is to the palletizing work and only could be done, previously, by a human worker by inserting fingers into the gap (hole) formed by the stacking pattern on the stack and taking off each group of bags layer by layer. Figure 6.21 shows the gaps.
A robot de-palletizing system is required, as depicted in Figure 6.22, to insert a tool into the gap on the stack, then this portion is lifted up to a press board on the end-effector.

The key problem for a de-palletizing system is how to reliably and robustly achieve gap-center insertion of the mechanical finger without touching or disturbing the stack. Limited by the thickness and size of bags, there is only a small tolerance for engagement-positioning error. The elevation of the gap-center position and gap size is variable due to differenting pressure depending upon the number of layers above it, the so-called “variable crunch” factor. Also after storage and transportation, the stack might rotate slightly relative to the pallet. All of these variations make it impossible to teach the robot every gap-center position and orientation in advance and just repeat the same action to un-stack the bags. Every gap should be located by the robot system individually.
Therefore, only a vision-guided robot system can achieve this task. Figure 6.23 shows the overview of a vision-guided de-palletizing demonstration system.

![Vision guided de-palletizing system overview](image)

Figure 6.23 Vision guided de-palletizing system overview

Three ceiling cameras view the gaps together with three near-planar surfaces of the stack. One single laser pointer and one multiple laser pointer are mounted on the pan/tile unit. A six-axis robot is controlled by a computer based on the visual information acquired from the cameras.

Reliable and robust gap-center location and orientation is critical. Traditional 2D image analysis to extract the gap center would be ineffective under the varying illumination and complex coloration of bags that typify the company’s product.

Only the laser-spot-assisted 3D image analysis can extract the reliable gap target for the robot. The procedure includes these steps.

Step 1: Figure 6.24 shows the multiple laser spots were cast onto the top, front and side surfaces of the stack. Spot centers are detected and matched among cameras, as described in Chapter 4. Then 3D coordinates of the centers are estimated in the nominal-World-frame coordinates.
Step 2: The laser spots close to the right upper corner of stack are used to fit three perpendicular planes for intersecting to find the edges and corner, as shown in Figure 6.25.

Step 3: With the 3D coordinates of the corner in the nominal World frame, and a known size and thickness of the bags, the center of whichever gap is closest to the corner is roughly estimated in the 3D nominal World frame, as shown in Figure 6.26.
Step 4: Analysis of the distribution of spots on the front surface in the 3D nominal World frame, which represents the geometric characteristics of the front surface and gap, will also identify the gap center. As illustrated in Figure 6.27, the spots on the bottom can be identified by the distance between them and spots falling on the front surface. Therefore, fitting the front plane of the stack with the spots around the gap and checking the distance of spots to the plane can identify the bottom-gap spots. Also the front plane provides the orientation of gap insertion. With knowledge of the gap size, the elevation of the gap center is estimated. Investigating the pattern, and particularly the absence, of laser spots is able to verify the gap center and identify its size in 3D nominal World frame. This use of a redundant gap-center position and orientation determination provides reliable and robust targeting to insert the metal finger into the gap and grasp the bags.
Step 5: The robot inserts the tool into the gap and linear actuator push the upper board to grasp the bags, as shown in Figure 6.28.

![Figure 6.28 The robot inserted the tool into the gap and pick up the bags](image)

6.3.4 Summary of Palletizing and De-palletizing Systems

Though the palletizing and depalletizing tasks didn’t require the high level precision and six-components positioning as the assembly and sanding tasks, these tasks showcased a unique advantage, the robustness of the laser-spot assisted 3D image analysis with CSM. They also demonstrate the flexibility of the new method to guide the robot to perform the less complex 2.5 D tasks. The developed palletizing systems have been running in factories over several years to stack various types of bags, which have different color, material, size, etc, under a variable ambient lighting environment on the floor and vibration on the ceiling, where the vision system is mounted. The performance of the systems satisfies all the requirements from the customer. The developed prototype of de-palletizing system also demonstrates the reliable gap insertion in un-stacking process. It is ready to be transferred to a factory floor.

The success of the implementation of CSM and the absence of workable solution from other vision guided methods in palletizing and de-palletizing applications prove again the statement in Chapter 1. Even the simplest task for a human being using hand-eye coordination, such as stacking the bags, cannot be accomplished by a robot under
either teach/repeat or vision-guided methods based on calibration or visual servoing. That is because they do not take advantage of the robot’s artificial dexterity insofar as the way to construe and pursue the maneuver in the reference frame of the visual sensor in order to retain the robustness and reliability of what CSM and humans do.
CHAPTER 7:
SUMMARY AND FUTURE WORK

7.1 Summary

The dissertation presents a laser-spot-assisted, three-dimensional image analysis, which work together with camera-space manipulation, for robot surface operation and assembly. The discussion of a pervasively missing present-day element, in actual industrial robot applications which are discussed in Chapter 1, the review of current vision-based robotics approaches of Chapter 2, and the examination of intrinsic problems in the conventional 2D image analysis led to the study and development of the present approach.

Three-dimensional image analysis in a vision-guided robot system with the use of laser spots extends the application of camera-space manipulation. The dissertation addresses the approach through illustrative and proof-of-concept experiments and applications. Three related applications and two fundamental experiments have been presented as part of this dissertation.

The two fundamental experiments, thickness reduction gauging and two-surface positioning, prove experimentally that the approach is capable of positioning with high precision and robustness relative to object surfaces that are arbitrarily located within the robot’s workspace.
A comparison of current industrial solutions for robotic assembly with the present, high-precision peg-in-hole assembly experiment reveals the intrinsic problem with current vision-based robotics and proves the unique advantage of the present method. The second, surface-finishing task entails using the laser spots for the combination of gauging the surface reduction as well as positioning the tool with high precision. It entails both the thickness gauging and high-precision poisoning relative to a surface, as with the five-component positioning experiment. The third palletizing and de-palletizing applications manifest the robustness of the present method in real-world application over time, and the flexibility of the method with various levels applications complexity.

The method has the unique advantage of casting laser spots onto the surfaces as well as the high positioning precision within an asymptotic-limit volume or region for both optics and kinematics, and “averaging out” the zero-mean image noise using a large amount of cue data.

These advantages of three-dimensional image analysis remove the illumination and surface-coloration vulnerabilities of traditional two-dimensional image analysis. That has been one of the big obstacles for implementing a vision-guided robot system in the real world, despite the fact that the robot has ample mechanical ability to complete a vastly larger range of tasks compared with actual present-day use.

These advantages of three-dimensional image analysis also improve the precision of the surface operating without paying the high cost of elevated global precision of the mechanism. These advantages also permit exploitation of prior understanding of the local underlying 3D geometry of the object.
7.2 Future Work

1. Pursue the opportunity to implement the laser-assisted 3D image analysis with CSM with industrial applications which require high-precision positioning and reliability, such as machining and assembly.

2. Investigate the integration of general commercial 3D data processing software with the present laser-spot method in order to automatically identify the features of parts using a CAD model and characterize the geometry of the surfaces to provide high precision targets for robot positioning using CSM.

3. Investigate the scalability of the system, especially on precision. Study the ability of the present method to position a rigid body relative to surface with precision that exceeds the precision of robot repeatability, using high resolution cameras zoom lenses and a smaller work space.

4. Implement the present approach to guide wide range of mechanism other than robots. Many mechanism guided by human hand-eye coordination maybe able to be guided using some variation of this control approach.

5. Integrate the force control approach to complement the present method.
APPENDIX A:
THE DEFINITION OF ASYMPTOTIC-LIMIT REGION

There is a gradually shifting “nominal World frame” to which target points or manipulation objectives are referred. The nominal World frame $x^N$-$y^N$-$z^N$ is generally a translation and rotation of the World frame fixed to the robot based, $x$-$y$-$z$, because of the difference between the nominal forward kinematics and real forward kinematics of the robot. The nominal World frame shifts gradually because the rigid-body accommodation of actual kinematics is local to any particular region of joint space. Also there is a nominal local frame corresponding to the nominal World frame. The nominal local frame shifts in a same way as the nominal World frame does.

![Figure A.1 Coordinate frames of Camera Space Manipulation vision system](image)

There are six parameters of the homogenous coordinate transformation between the world frame and nominal World frame. If the infinite number of no noise-error samples within a maximum size of region to the origin of local frame, $\Delta x$-$\Delta y$-$\Delta z$, do the
best to identify these all six parameters. If the residual of this fitting is under the order of magnitude of error in kinematics and optical model, this region is the asymptotic-limit region.
BIBLIOGRAPHY


[22] A. Ardeshir Goshtasby, 2-D AND 3-D Image Registration, 2005


