

# LASER-ASSISTED UNCALIBRATED VISION GUIDED ROBOTIC DE-PALLETIZING

Biao Zhang, Corporate Research Center ABB Inc.; Steven B. Skaar, University of Notre Dame

## Abstract

In the paper-container industry, bag stacking and unstacking is 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 injuries. Therefore, a robotic de-palletizing system is highly desirable. Guiding a robot tool reliably and robustly in order to insert into the gap in the bag stack to pick up a layer of bags without disturbing the stack is highly challenging due to the variation of the gap-center position and gap size under varying pressure depending upon the number of layers above it, the so-called “variable crunch” factor. In this study, a method combining an uncalibrated vision and 3D laser-assisted image analysis based on camera-space manipulation (CSM) was developed. The prototype demonstrated reliable gap insertion in the de-palletizing process and was made ready for installation on a factory floor at the Smurfit-Stone Container Corporation.

## Introduction

In the paper-container industry, at the end of each stage of the production line, paper bags are stacked layer by layer according to a specific pattern, as shown in Figure 1, for storing and transporting. Eventually, the stack of bags needs to be un-stacked layer by layer and fed into a machine for the next procedure in fabrication, or to be packed into a box. This is very labor-intensive work and 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. For these reasons this robotic palletizing and de-palletizing system was developed.

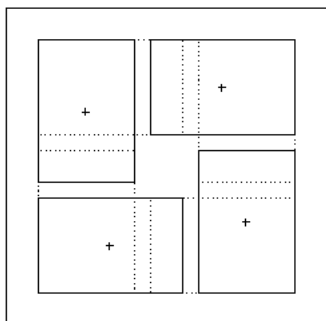


Figure 1. Pattern of Bag Stacking

One automated robotic de-palletizing system would save six human stackers in each paper bag production line in a three-shift operation. The initial investment for installation is recovered in one year. The robotic de-palletizing task is more challenge than is 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 2 shows the gaps.



Figure 2. Gaps on Paper Bags Stack

A robotic de-palletizing system is required, as depicted in Figure 3, to insert a tool into the gap on the stack. Then this portion of bags is lifted up against a press board on the end-effector.

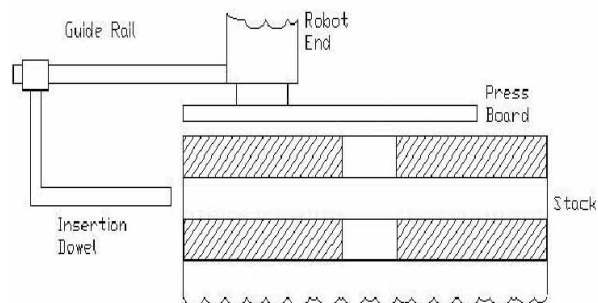


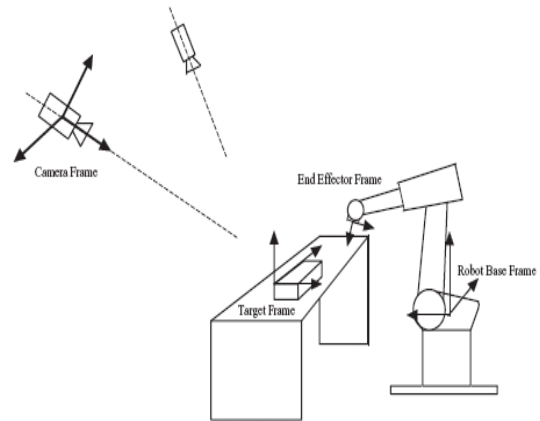
Figure 3. Gap Insertion

The key problem for a robotic de-palletizer 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 existing teach/repeat way to use robots cannot solve the problem in this bag de-palletizing application because the elevation

of the gap-center position and gap size is variable due to varying pressures 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 robotic system individually. Therefore, only a sensor-guided robotic system can achieve this task.

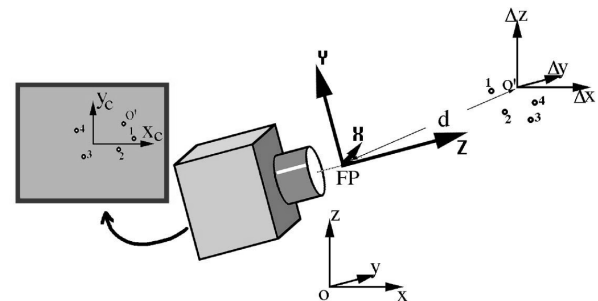
## Camera-Space Manipulation (CSM)

Calibration and visual servoing are two mainstream methods of vision-guided robotics. 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[1],[2]. Calibration relies entirely on an accurate camera model and a 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. Visual servoing takes a closed-loop control approach to drive the positioning error in the image toward zero [3]. One of the biggest drawbacks in visual servoing is that one needs 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. 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. The method of camera-space manipulation (CSM) emerged in the mid-1980s and developed over the past 20 years 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 [4]. 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 4 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.



**Figure 4. Coordinate Frames of a Typical Vision System**

In contrast, CSM uses six parameters to locally identify the mapping relationship between the internal—and directly controllable—robot-joint rotations within the relative workspace and the local 2D camera-space [5]. As indicated in Figure 5, the physical 3D points, which scatter around a local origin (flattening point), are projected onto the 2D image-plane, with  $X_c$ - $Y_c$ , as “camera-space coordinates”. These physical 3D points are designated with respect to a local frame,  $\Delta x$ - $\Delta y$ - $\Delta z$ , the axes 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 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,  $X_c$ - $Y_c$ , and the origin is on the system’s equivalent focal point.



**Figure 5. Coordinate Frames of Camera-Space Manipulation Vision System**

This local mapping relationship is described by equations (1) and (2), which correspond to the assumption of an orthographic camera model

$$X_c = A_{11} \cdot \Delta x + A_{12} \cdot \Delta y + A_{13} \cdot \Delta z + A_{14} \quad (1)$$

$$Y_c = A_{21} \cdot \Delta x + A_{22} \cdot \Delta y + A_{23} \cdot \Delta z + A_{24} \quad (2)$$

where  $X_c$ ,  $Y_c$  represent the 2D image frame and  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  represent the local frame  $\Delta x$ - $\Delta y$ - $\Delta z$ , with origin on the focal axis and where  $A11$  through  $A24$  are groups of nonlinear expressions dependent upon the six view parameters,  $C1$ ,  $C2$ , ...,  $C6$ , as follows:

$$A11 = C12 + C22 - C32 - C42 \quad (3)$$

$$A12 = 2(C2C3 + C1C4) \quad (4)$$

$$A13 = 2(C2C4 - C1C3) \quad (5)$$

$$A14 = C5 \quad (6)$$

$$A21 = 2(C2C3 - C1C4) \quad (7)$$

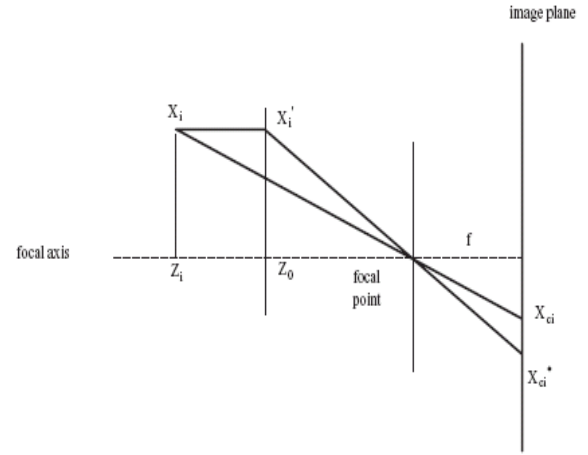
$$A22 = C12 - C22 + C32 - C42 \quad (8)$$

$$A23 = 2(C3C4 + C1C2) \quad (9)$$

$$A24 = C6 \quad (10)$$

The first four parameters,  $C1$ - $C4$ , 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,  $C5$  and  $C6$ , 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 the 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 [6]. 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 the robot. 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: that an adequate number of end-member samples for estimating the view parameters should be acquired within the asymptotic-limit region, and 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 was used [7]. The flattening procedure is based on a presumption of a pinhole projection of physical points onto the 2D image plane, as depicted in Figure 6. 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 given by equations (1), (2).



**Figure 6. Projection According to the Pinhole Camera Model**

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

$$\frac{X_{ci} \times Z_i}{Z_o} \quad (11)$$

Which is 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

$$\frac{Y_{ci} \times Z_i}{Z_o} \quad (12)$$

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 are established for each camera in the CSM vision system, one gets separate camera-specific expressions for equations (1) and (2). 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 effect. The estimation procedure is as follows:

1. Choose an origin of the local frame, the closer to the target, the better.
2. Compute  $[C1, C2... C6]$  for each camera using samples flattened about this local frame's origin.
3. Estimate the relative position of the target point with respect to the local frame by solving the non-linear equations (1) and (2).
4. Shift the origin of the local frame to the newly estimated target position.
5. Repeat steps 2 through 4 until the shift of the 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  $[C1, C2... C6]$  for each camera.  $C5$  and  $C6$  become  $X_c$  and  $Y_c$  for the camera-space coordinates of the target point.

### 3D Laser-Assisted 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 visualization 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 a new image analysis in three-dimensions using an approach that complements CSM technology [8]. The target information from the 3D 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 three-dimensional 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 onto a 2D image plane. This 3D information, for example from a CAD file, would facilitate reliability and robustness, and enhance the utility of results gained from 3D image analysis.

For detecting the location of the center of the laser spot in each camera space, the laser-spot identification procedure is as follows [6]:

1. 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.
2. Turn off the laser pointer and acquire the image of the object surface with the camera.

3. Take the image difference between these two images to make only the laser spot stand out.
4. Apply a "mask", as indicated in Figure 7, in order to condition the differenced image, replacing all pixel values except those in the rightmost, leftmost, uppermost, and lowermost three 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

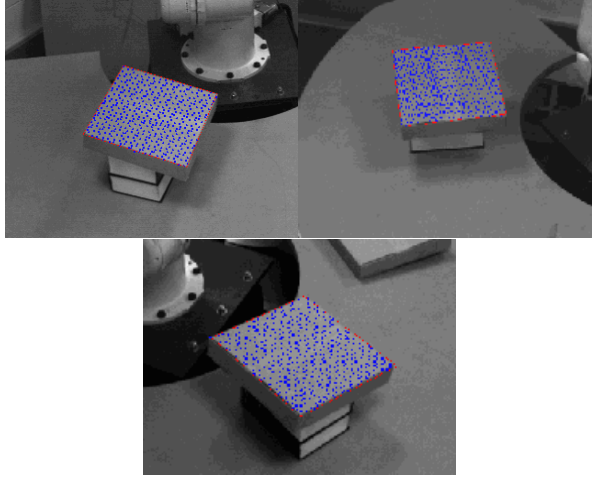
			1		
	1	2	4	2	1
	2	6	7	6	2
1	4	7	8	7	4
	2	6	7	6	2
	1	2	4	2	1
			1		

Remaining elements are zero.

**Figure 7. 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.

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, the 3D shape-measurement approach and the ambient-illumination independence of using laser-spot identification, 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 8 [9]. Then, the laser-spot 3D coordinates in the nominal World-frame are estimated.

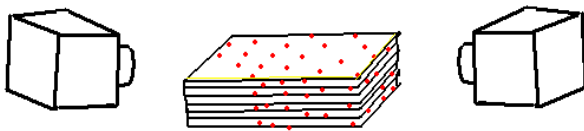


**Figure 8. Multiple Laser Spots Casted on Object Surface (Views from Three Cameras)**

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 with 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 image discretization and other noise. This characterization is applied either to a previously known model of the object's surface geometry or to a quadratic or other polynomial geometry 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. 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 9. 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.

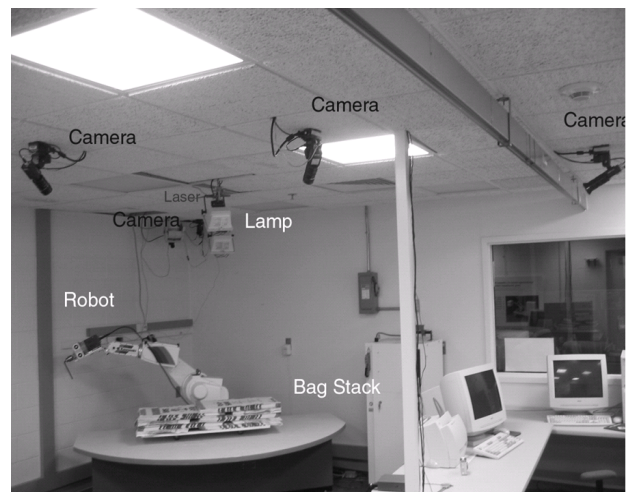


**Figure 9. Three surfaces meeting**

There are three advantages of edge detection based on 3D image analysis. First, the edge-identification procedure is independent of variations in illumination and reflective properties of various materials 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. Second, 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. Third, 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 into a surface model. A smaller number of parameters of the model needed to be fitted results in less sensitivity to noise in the data and, thereby, reduces the required quantity of data.

## Implementation

Figure 10 shows the overview of a vision-guided depalletizing demonstration system. 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/tilt unit. A six-axis robot is controlled by a computer based on the visual information acquired from the cameras.

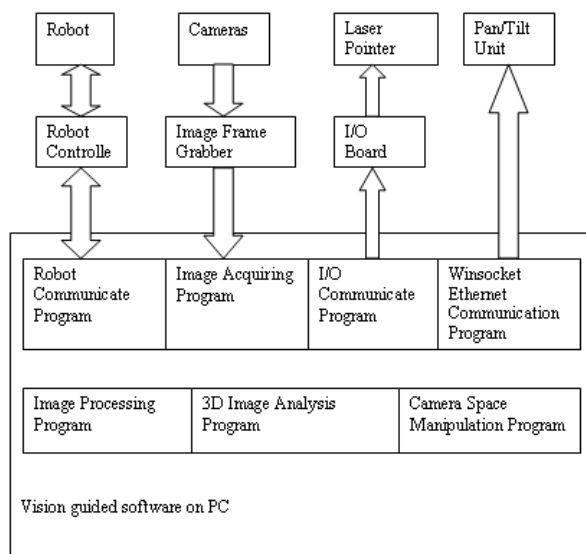


**Figure 10. Vision Guided De-palletizing System Overview**



The configuration of the fixed cameras is selected due to the cycle-time requirement of the de-palletizing system for keeping pace with paper-bag production lines. Compared to the eye-on-hand configuration, fixed cameras can acquire images while the robot is placing the bags into the feeder of the paper-bag production line. Also, this configuration avoids the problem of the robot blocking the laser projection in the eye-on-hand configuration, which really simplifies the robot path planning and task sequencing.

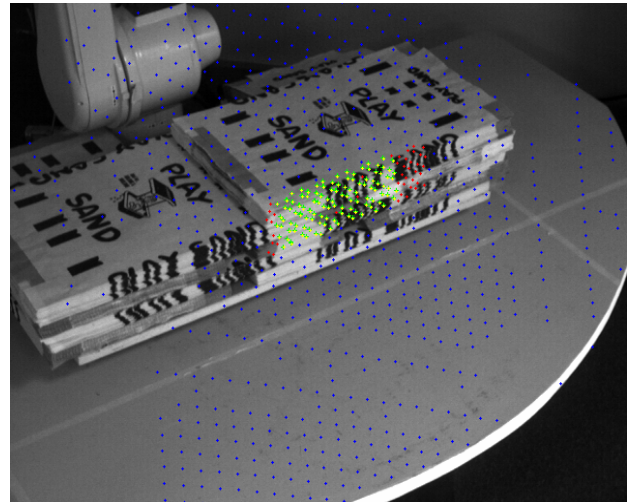
The vision-guided software written in C++ runs on a PC. It reads and writes the robot joint coordinates into the robot controller through the serial port. The pan/tilt unit is controlled by the PC C++ program through an Ethernet port. The image is acquired from cameras through a DT3150 frame grabber. An overall diagram of the system is shown in Figure 11.



**Figure 11. Overall Diagram of the System**

Reliable and robust gap-center location and orientation is critical. Traditional 2D image analysis to extract the gap center would be ineffective under varying illumination and given the 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 12 shows the center, which is superimposed on the image with laser projection off, of detected multiple laser spots cast onto the top, front and side surfaces of the stack. Spot centers are detected and matched among cameras. Then, 3D coordinates of the centers are estimated in the nominal-World-frame coordinates.



**Figure 12. Center of Detected Multiple Laser Spots on Three Surfaces of Bag Stack**

Step 2: The laser spots close to the right upper corner of stack are used to fit three perpendicular planes for intersecting in order to find the edges and corner, as shown in Figure 13.



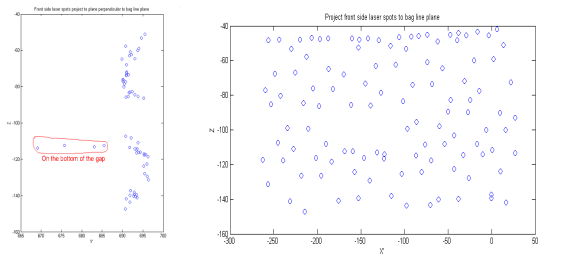
**Figure 13. Edges and corner of the bags stack**

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 14.



**Figure 14. Rough Estimation of the Location of the Gap**

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 15, 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 allows the system 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.



**Figure 15. Laser Spots on Front Surface**

Step 5: The robot inserts the tool into the gap and a linear actuator pushes the upper board to grasp the bags, as shown in Figure 16.



**Figure 16. The Robot Inserted the Tool Into the Gap and Pick Up the Bags**

## Conclusions

The prototype for the de-palletizing system developed in this study demonstrated reliable gap insertion in an un-stacking process. It showcased a unique advantage, the robustness of the laser-spot-assisted 3D image analysis with CSM. It also demonstrated the flexibility of the new method to guide the robot to perform the less complex 2.5 D tasks. It is ready to be transferred to a factory floor to un-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. This method can also be used in similar de-palletizing applications.

## Acknowledgments

The authors are thankful to Bill Aman and Wayne Schumm, Smurfit-Stone Container Corporation, for their support.

## References

- [1] H. Zhuang, "Simultaneous Calibration of a Robot and a Hand-Mounted Cameras," IEEE Trans. on Robotics and Automation, vol. 11, No.5, October 1995.
- [2] F. Dornaika and R. Horaud. "Simultaneous Robot-World and Hand-Eye Calibration," IEEE Trans. on Robotics and Automation, vol. 14, No.4, August 1998.
- [3] L.E. Weiss, "Dynamic Visual Servo Control of Robots: an Adaptive Image-Based Approach," Ph.D. dissertation, Robotics Institute, Carnegie-Mellon University, Pittsburgh, PA, 1984.
- [4] S. B. Skaar, W. H. Brockman and R. Hanson, "Camera space manipulation," International Journal of Robotics Research, vol. 6, No.4, pp. 20-32, Winter 1987.
- [5] S. B. Skaar and G. Delcastillo, Revisualizing Robotics: New DNA for Surviving a World of Cheap Labor. Esgeville, PA: DNA Press, LLC, 2006, pp. 107-140.

- 
- [6] M. J. Seelinger, "Point and click Camera-space Manipulation, Mobile Camera-space manipulation, and Some Fundamental Issues Regarding the Control of Robots Using Vision," Ph.D. dissertation, Dept. Aerospace and Mechanical Eng., University of Notre Dame, Notre Dame, IN, 1999.
  - [7] S. B. Skaar, W. H. Brockman and W. S. Jang, "Three dimensional camera space manipulation," *International Journal of Robotics Research*, vol. 9, No.4, pp. 22-39, August 1990.
  - [8] B. Zhang, E. J. Gonzalez-Galvan, J. Batsche, S. B. Skaar, L. A. Raygoza and A. Lored, "Precise and Robust Large-Shape Formation using Uncalibrated Vision for a Virtual Mold," in *Computer Vision*, Z. Xiong, Ed. Vienna, Austria: I-Tech, 2008, pp. 111-124.
  - [9] Z. Fan, "Industrial Applications of Camera-Space Manipulation with Structured Lights," Ph.D. dissertation, Dept. Aerospace and Mechanical Eng., University of Notre Dame, Notre Dame, IN, 2003.

## Biographies

**BIAO ZHANG** is currently a research scientist in Mechatronics and Robotics Automation department, US Corporate Research Center of ABB Inc. He received his Ph.D. in mechanical engineering from University of Notre Dame. His research interests include vision-guided robotics, robotic force control assembly, design of experiment optimization, automation on material handling, and human robot interface (phone: 860-687-4939; fax: 860-285-0273; e-mail: biao.zhang@us.abb.com).

**STEVEN B. SKAAR** is professor of Aerospace and Mechanical Engineering at the University of Notre Dame, co-inventor of robust, new visually guided robot solutions <http://www.nd.edu/~sskaar/> and editor of the American Institute of Aeronautics and Astronautics volume *Teleoperation and Robotics in Space*. He draws heavily on NASA's experience with robot use together with a range of other application types to illustrate the problems and great future prospects of robotics.