The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Robotic De-palletizing Using Uncalibrated Vision and 3D Laser-Assisted Image Analysis

Biao Zhang and Steven B. Skaar

Abstract—**In the paper-container industry, bag stacking and un-stacking are 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 robot de-palletizing system is highly desirable. Guiding a robot tool reliably and robustly to insert into the gap on 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 differenting pressure depending upon the number of layers above it, the so-called "variable crunch" factor. In this paper, the method combining an uncalibrated vision and 3D laser-assisted image analysis based on camera-space manipulation (CSM) is developed. The developed prototype system demonstrates the reliable gap insertion in de-palletizing process. It is ready to be installed to a factory floor at the Smurfit-Stone Container Corporation.**

I. Introduction

IN the paper-container industry, at the end of each stage-of-production line, paper bags is stacked layer by layer according to some 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 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.

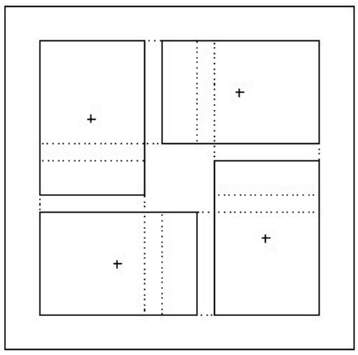

Fig. 1 Pattern of bag stacking

Manuscript received February 28, 2009.

Biao Zhang, Author is with Corporate Research Center of ABB Inc., Windsor, CT 06095, USA. (phone: 860-285-6849; fax: 860-285-6939; e-mail: biao.zhang@ us.abb.com).

Steven B. Skaar is with University of Notre Dame, Notre Dame, IN 46556 USA. (e-mail: SSkaar@nd.edu).

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

Fig. 2 Gaps on paper bags stack

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

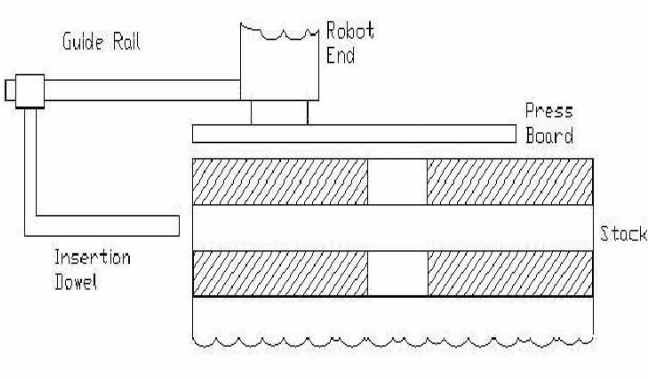

Fig. 3 Gap insertion

The key problem for a robotic de-palletizing 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 robot 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 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 sensor-guided robot system can achieve this task.

II. 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 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 close-loop control approach to drive the positioning error in the image toward zero [3]. 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. 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 in 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.

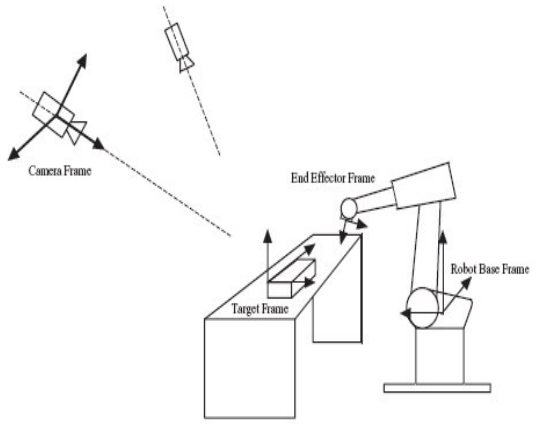

Fig. 4 Coordinate frames of a typical vision system

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 [5]. As indicated in Figure 5, 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 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 Xc-Yc, and the origin is on the system's equivalent focal point.

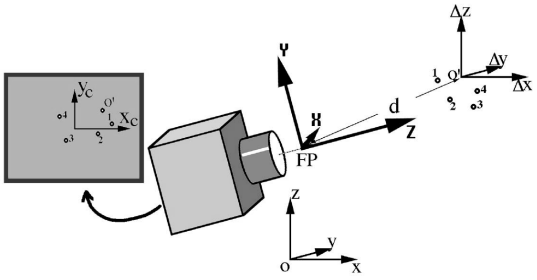

Fig. 5 Coordinate frames of Camera-Space Manipulation vision system

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

$Xc = A11*\triangle x + A12*\triangle y + A13*\triangle z + A14$ (1)

$Yc = A21*\triangle x + A22*\triangle y + A23*\triangle z + A24$ (2)

Where Xc, Yc are with respect to the 2D image frame and $\triangle x$, $\triangle y$, $\triangle z$ are with respect to local frame Δx-Δy-Δz, with origin on the focal axis and where each of A11, A12, … , A24 groups a nonlinear expression dependent upon six view parameters [C1, C2, …, C6] as follows:

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

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

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

$A14 = C5$ (6)

A21= 2(C2C3-C1C4) (7)
A22= C12-C22+C32-C42 (8)
A23= 2(C3C4+C1C2) (9)
A24= C6 (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, 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 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 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: 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 [7]. 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 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 of (1), (2).

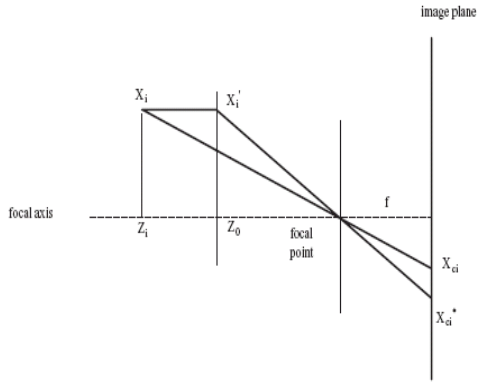

Fig. 6 Projection according to the pinhole camera model

The X coordinate of an ith raw camera-space sample of a particular juncture on the robot end effector is Xci. The flattened sample is determined by $\frac{X_{ci} \times Z_i}{Z_o}$ based on the assumption of a pinhole or perspective lens model, where Zi represents the location of the sample along the optical axis of the camera, and Zo is the location of the origin of the local frame Δx-Δy-Δz with respect to the camera frame. The Y coordinate of the *i*th raw camera-space sample Yci is determined by $\frac{Y_{ci} \times Z_i}{Z_o}$ 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 (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 affect. The estimation procedure is as follows:

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

Compute [C1, C2… C6] 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 (1) and (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 [C1, C2… C6] for each camera. C5 and C6 become the Xc and Yc, the camera-space coordinates of the target point

III. 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 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 [8]. 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 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 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.

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

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 7, 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.

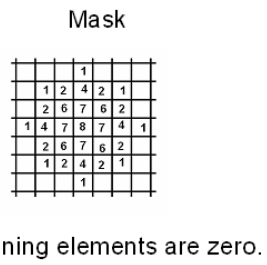

Fig. 7 Applying a mask to each pixel provides data regarding its value as well as surrounding pixel values

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 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.

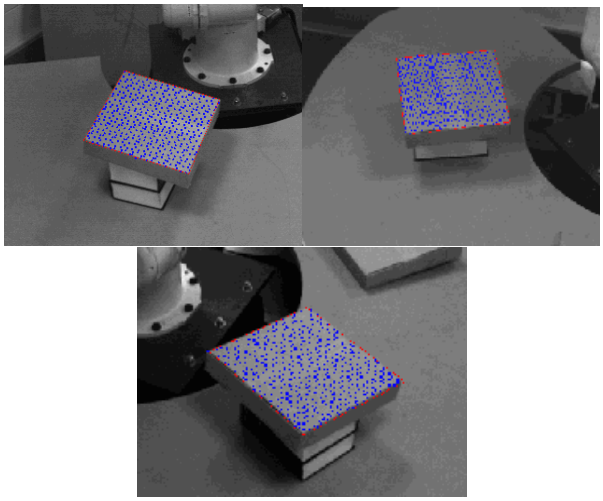

Fig. 8 multiple laser spots are cast on object surface

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.

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.

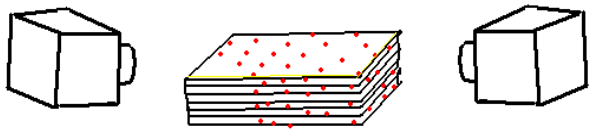

Fig. 9 Three surface meeting

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

IV. IMPLEMENTATION

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

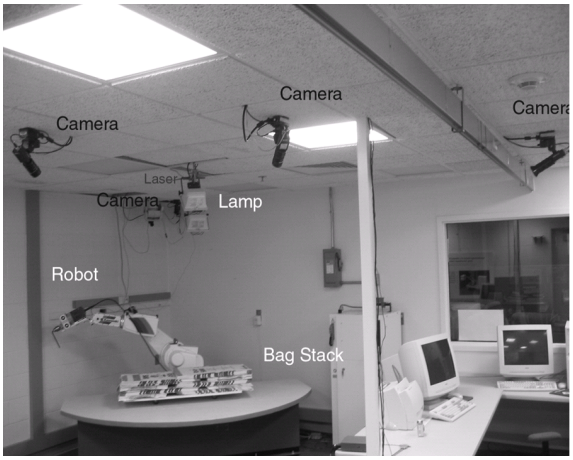

Fig. 10 Vision guided de-palletizing system overview

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 11 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. Then 3D coordinates of the centers are estimated in the nominal-World-frame coordinates.

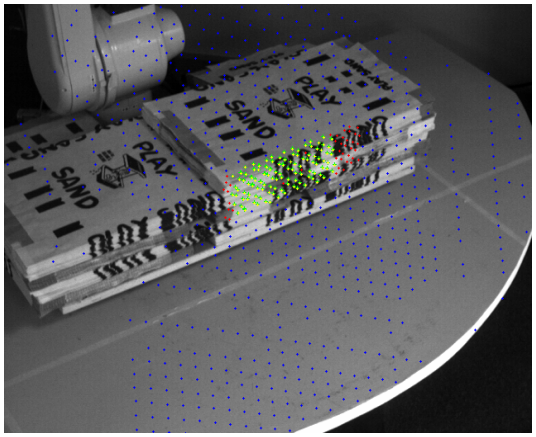

Fig. 11 Multiple laser spots on three surfaces of bag stack

Step 2: The laser spots close to the right upper corner of stack are used to fit three perpendicular planes for intersecting to find the edges and corner, as shown in Figure 12.

Fig. 12 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 13

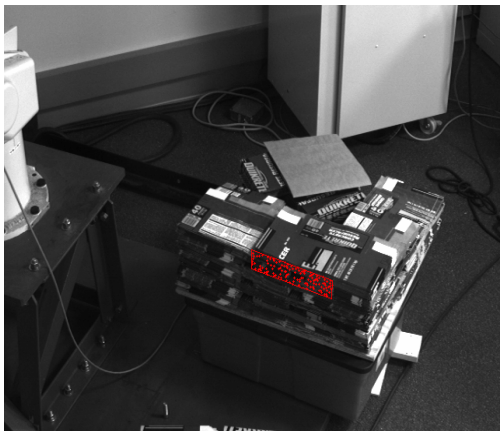

Fig. 13 Rough estimation of the location of the gap

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 14, 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.

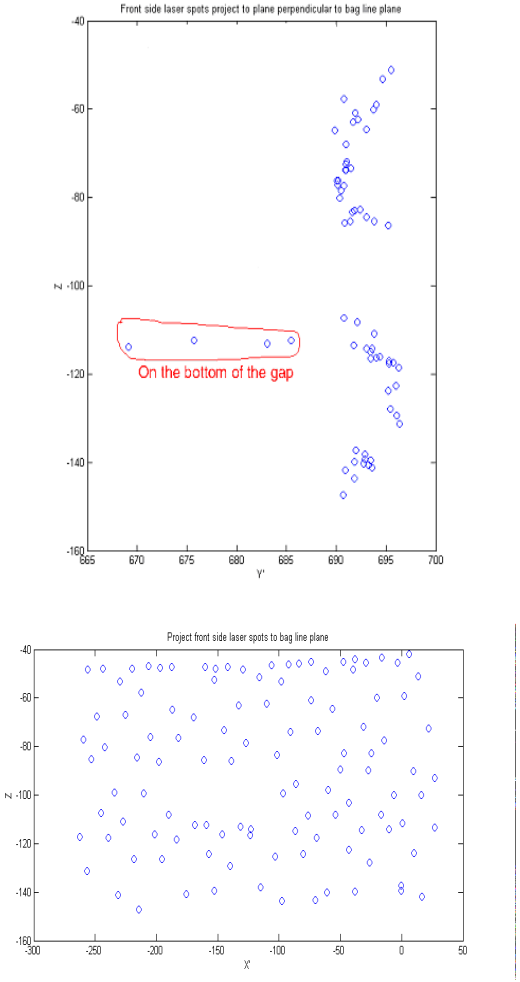

Fig. 14 Laser spots on front surface

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 15.

Fig. 15 The robot inserted the tool into the gap and pick up the bags

V. CONCLUSIONS

The developed prototype of de-palletizing system demonstrates the reliable gap insertion in un-stacking process. It showcased a unique advantage, the robustness of the laser-spot assisted 3D image analysis with CSM. It also demonstrates 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. The developed method can also be used to similar de-palletizing applications.

ACKNOWLEDGMENT

The author thanks 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.* Esgleville, 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. Loredo, "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.