Autonomous Pallet Localization and Picking for Industrial Forklifts Based on the Line Structured Light

Shijun Wang, Aixue Ye, Hao Guo, Jiaojiao Gu, Xiaonan Wang and Kui Yuan

Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing, China

wangshijun2012@ia.ac.cn

Abstract - **This paper proposed an autonomous pallet handing method for industrial forklifts based on the line structured light sensor. The method mainly include two aspects. First, the design of the line structured light sensor based on embedded image processing board that contains a FPGA and a DSP. We solved the problem that Hessian matrix decomposition based light stripe center extraction cannot run in real time. Second, we installed the structured light sensor onto the automatic forklift, and the forklift detected and located the pallet using the geometry character of the pallets, then the controller drived the forklift docking in front of the pallet using position-based visual servoing and picking up the pallet at last. The line structured light sensor was designed using C6000 series visual processing boards developed by our research group. The experiment of the autonomous pallet localization and picking for industrial forklifts verifies the practicability and effectiveness of our proposed method.**

Keywords – Structured light, Pallet, Visual servo, Embedded vision

I. Introduction

Nowadays robots are widely used in many fields, such as industrial robot or named manipulator, family service robot, planetary exploration robot and rescue robot, and so on. As the extension of the application of the robot, the level of automation and intelligence is becoming higher, and one of the forms is the capacity of automatically sensing the environment and automatic processing which is performing much better. In the field of industrial research, there is increasing interest in extending intelligent storage to unstructured environments, which means that the robot has to carry out several specific tasks with great uncertainty, e.g. driving a forklift to dock to a pallet whose pose is not predetermined. The main result of achieving this goal is that the forklift can handle materials fully automated including loading and unloading phases. The main advantages are the use of less structured and much cheaper stocking areas. In this scenario, the forklift must have the capacity of sensing the environment and processing with the uncertainly for completely or partially unknown environments which might introduce more uncertainty and less repetitiveness in handing tasks. Hence reliable but cost-effective sensors together with robust real time algorithms must be employed in automatic pallet handling.

Visual sensors are the most commonly used in robot environment awareness because they can offer a wide range of information. With the advantages of wide range, large field of view, high precision, easy extraction, real-time performance and active controlled etc., the structured light vision measurement method has been widely applied in industrial environment. There have been a lot of products based on structured light vision measurement in the market, such as line structured light 3D laser scanner, cross laser 3D scanner, and coded structured light 3D measuring etc.

This paper presents a method that automatically identifies and locates the pallet using the line structured light sensor and drive the forklift handling the pallet based on position-based visual servoing method. First, we designed a line structured light sensor based on an embedded image processing board, and then applied this senor to search for the pallet and estimate the pose of the pallet if found, and finally drive the forklift to dock in front of the pallet and handle it. This method was proved to be able to complete automatic pallet handling tasks and could be used in intelligent storage system.

II. Related Work

Vision approaches are most commonly used in the automatic identification and location of the pallet. The previous researches are mainly based on the detection of natural or artificial features that include the artificial features, the geometric structure of the pallets, color information and hybrid features. Usually, they used camera, laser scanner and time-of-flight infrared cameras to derive information from the environment.

The first vision-based pallet detection and localization system mounted on a retrofitted manual forklift can be found in [1]. They solved the problem by identifying two pallet slots and estimating their geometric center in calibrated images. However, that system requires high prior knowledge of the pose of pallet, and only active in the final phase when the pallet is near to the forklift by moving the tines rather than the vehicle itself. [2] described a retrofitted autonomous forklift with the capability of stacking racks and picking up pallets placed with limited uncertainty. The method is based on detection of specific reference lines for concurrent camera calibration and identification, and allows stacking of well–illuminated racks and localization of pallets in front of the vehicle and close to it. A more complex scheme which is on the basis of hierarchical visual features like regions, lines and corners using both raw and template-based detection was proposed in [3]. And the main drawbacks of this method are the complexity of decision trees and the instability of the scale invariance. [4] estimated the pose of the pallets by the extrinsic calibration of the camera relatively to the vehicle and the floor and completely rely on the upper and bottom edges of the front pallet face, therefore the robustness of this method is poor. The method proposed in [5] attempted to estimate pallet pose using structured light method. The main problem is that the accuracy of the measuring method using structured light quickly decreases with distance. The application of fiducials proposed in [6] can solve most of the illumination and calibration problems but it is a big project to stick artificial targets on each pallet and the targets may be fuzzy or broken as time goes by. [7] used the measurement data obtained from LRFs detected and localized the pallet but could not deal with matching problems in the case of multiple targets. [8] used a fast linear program for segment detection, applied to pre-filtered points selected by man gestures on a PDA showing an image from a camera mounted on the forklift. The pallet is identified by the classification of detected segments belonging to its front face and the position of it is then computed. [9] presented two laser scanner based approaches to locate and pick-up pallets which are independent of luminance conditions. [10] presented a combined double-sensor architecture, laser and camera, to solute the problem of identifying and localizing a pallet with large uncertainty prior. [11,12] integrated the vision output with odometry and realized smooth and non-stop transition from glob navigation to visual servoing.

The rest part is organized as following: Chapter III makes a detailed description of the line structured light sensor and the identification and location of the pallet. Chapter IV gives the simulation and pallet handling experiments results, and followed by the conclusions in chapter V.

III. Visual System and State Estimation

In our application, we use the linear structured light sensor that mounted in the forklift to identify and locate the pallet. Because the accuracy of the linear structured light decreases sharply with distance, we dock the vehicle in front of the pallet based on position-based visual servoing method and pick up the pallet precisely. In order to meet the real-time requirements, the traditional line structured light sensor always use sample light stripes center extraction algorithms such as geometrical center method and gray weighted centroid method which are running fast but show poor robustness. The method based on Hessian matrix decomposition proposed in [13] shows high accuracy, anti-noise ability but calculation complexity. With the help of visual processing board based on FPGA and DSP, we successfully developed the line structured light sensor with the features of higher precision, more real-time and stronger robustness.

A. The Visual System

Typical line structured light sensor is mainly consisted of laser projector, camera, processor and so on. As shown in Fig. 1, the narrow laser plane produced by laser projection beam passing a cylindrical lens forms a light stripe when intersect with the surface of the object to be detected. The change of the light stripe shows the fluctuation of the surface. And the task of line structured vision is to obtain the 3D information on the surface of the object from the distorted light stripe image.

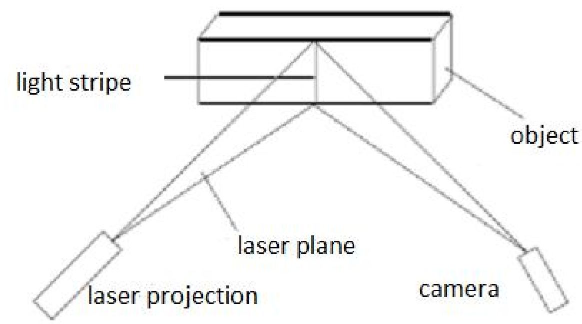

Fig.1 The principle of structured light

The whole algorithm of the structured light sensor is implemented on an embedded system that is shown in Fig.2. The embedded system mainly consists of FPGA, DSP, DDR, SDRAM, FLASH, two video inputs, and various interfaces to upper-computer. The algorithm for structured light can be summarized as the following steps: first, camera catches the light stripe image and transmit it to the FPGA; second image pre-processing, the algorithms of Gaussian convolution and detecting regions of interest are operated in the FPGA, then the FPGA transmits the result to the DSP; and finally the DSP finds the sub-pixel center of the light stripe image based on Hessian matrix decomposition and calculates the 3D coordinates according to calibration result, then the measured data are transmitted to upper-computer through Ethernet or Serial port. The flowchart of the algorithm is depicted in Fig.3.

Fig. 2 The embedded image processing board

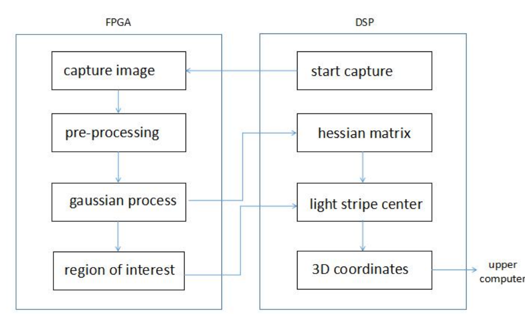

Fig. 3 The flowchart of the algorithm

B. Extracting Light Stripes Center

The algorithm for extracting light stripes center used in this article can be found in [13]. The light stripes image is processed by convolving with Gaussian function, illustrated in equation (1) and (2), and get the result at *(x, y)*, marked as r_x, r_y, r_{xx}, r_{xy}, r_{yy}.

$$g_\delta(x) = \frac{1}{\sqrt{2\pi}\delta}\exp\left(-\frac{x^2}{2\delta^2}\right) \quad (1)$$

$$\begin{cases} g_{x,\delta}(x,y) = g'_\delta(x) g_\delta(y) \\ g_{y,\delta}(x,y) = g_\delta(x) g'_\delta(y) \\ g_{xx,\delta}(x,y) = g''_\delta(x) g_\delta(y) \\ g_{yy,\delta}(x,y) = g_\delta(x) g''_\delta(y) \\ g_{xy,\delta}(x,y) = g'_\delta(x) g'_\delta(y) \end{cases} \quad (2)$$

Where δ refers to the standard deviation of Gaussian function. As a result of the separability of the convolution operation, two-dimensional Gaussian convolution can be decomposed into one dimensional convolution with rows and columns respectively. This property is fit for running in FPGA, because the image is introduced into FPGA line by line.

As a result of the narrow-band light filter, the gray of the image captured by camera is low except the light stripes part. In order to save running time, we detect regions of interest first and process these regions only. Theoretically speaking, the center of the light stripe should be located on the regions with strongest brightness, so we can distinguish the regions of interest through the gray value. This method works well upon most occasions, but if there are strong light disturbances such as incandescent lamp, the result will be with large error. As shown in the middle of Fig. 4, some detected center points are in the region of incandescent lamp. This paper use the synthesis of gray value and the extremum of the second order Gaussian filtering as the evaluation indicator to locate the regions of interest. As shown in the right of Fig. 4, majority points of the detection error are eliminated. The result of regions of interest and the corresponding Gaussian convolution is transmitted from FPGA to DSP.

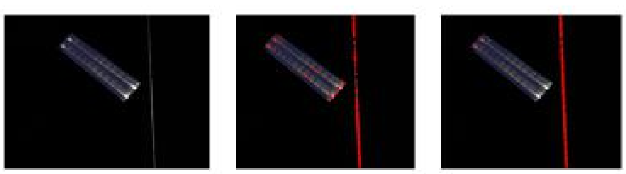

Fig.4 Left: original image. Middle: method of gray. Right: proposed method

The Hessian matrix of image coordinate at *(x, y)* is shown in equation (3), while the elements of the matrix were computed previously.

$$H(x,y) = \begin{bmatrix} r_{xx} & r_{xy} \\ r_{xy} & r_{yy} \end{bmatrix} \quad (3)$$

The second order gradient in the direction of the normal of this point is the eigenvalue of the Hessian matrix, and the direction of the normal is the corresponding eigenvalue. Then the subpixel coordinate adhere to the reference point *(x, y)* is (p_x, p_y).

$$(p_x, p_y) = (x + tn_x, y + tn_y) \quad (4)$$

$$t = -\frac{n_x r_x + n_y r_y}{n_x^2 r_{xx} + 2n_x n_y r_{xy} + n_y^2 r_{yy}} \quad (5)$$

If $(tn_x, tn_y) \in [-0.5, 0.5] \times [-0.5, 0.5]$ and the eigenvalue is greater than the threshold, (p_x, p_y) is considered as the central coordinate of the light stripe.

C. Parameters Calibration

The parameters calibration of the line structured light sensor consists of two parts, the calibration of camera intrinsic parameters and the calibration of laser plane. The 3D point (x_w, y_w, z_w) in the world coordinate system is mapped to (u, v) in the image coordinate system by equation (6).

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K * \begin{bmatrix} R & t \end{bmatrix} * \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (6)$$

where K is the intrinsic parameter of the camera. Because the points mapped to the center of light stripe are in the laser plane, they satisfy the equation (7).

$$aX_w + bW_w + cZ_w + d = 0 \quad (7)$$

In order to acquire the laser plane equation, at least three non-collinear points on the laser plane should be known. Parameters of laser planes are calibrated through planar target of unknown orientations [14]. During the calibration process, the camera captures a number of images from different orientations, and every image contains the planar target and light stripe. The intersections between the planar target and light stripe in every image are calculated, and the 3D coordinates of these intersections are obtained from invariance of cross-ratio. These intersections are the non-collinear calibration points for laser plane calibration. Then they are used to fit the laser plane by the least square method.

D. Identify and Locate the Pallet

The measured points acquired from the light structured sensor are continuous or isolated. The isolated points indicating the narrow plane or representing noise should be removed first. This means the center points acquired from hessian matrix decomposition will not be used in next step if they are too far away from the neighbors in the image coordinate system. Then the method calculates the remaining points' 3D coordinates in the camera coordinate system, fits the lines using the least squares method and label the collinear lines.

The approach for pallet recognition resets upon the well-known ICP algorithm. ICP is an iterative algorithm that matches two scans in a point-to-point manner. This means it takes the scan-points one-by-one, in the reference scan and tries to find a corresponding point in the current scan. In our method, we registered the current scan with a model predefined instead of two scans being registered for the ICP algorithm. The algorithm can be operated in two-dimensional space for all the points lying in the same plane namely laser plane and the model of the plane is an actual 2D scan.

The standard pallets are the object with special geometry structured, as shown in the left of Fig. 5. There are nine rectangular units used for supporting the body on the bottom which can be used as feature. With the help of the expectancy value of the pallet position and orientation, the model can be chosen as shown in the right of Fig. 5. We can only get partial information of the nine units, two lines for every unit at most because of the effect of shade. The lines are paralleled or vertical and some lines only include a few points. We also find that the three lines lying in the front face of the pallet include more points than others and they are collinear. So we search for line group with three lines collinear first, and if there are more than one group satisfy the conditions, we will sort the group according to the number of points it contained. Then the algorithm estimates the position and the posture of the point group relative to the camera, and then uses the result as the initial registration of the ICP algorithm. If the number of points that can be registered is more than a certain threshold, the scan can be identified as pallet and the pose of the pallet is the result of ICP algorithm. Otherwise, the remaining points group should be regarded as the front face of the pallet and identified using ICP algorithm until find the pallet or all the group are tested.

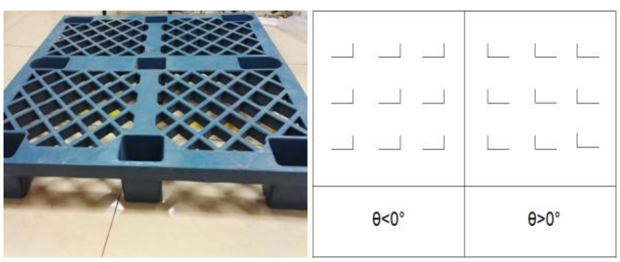

Fig.5 Pallet and pallet models

Because the linear structured light's measurement precision will decrease sharply as the distance increases, there will be of higher error rate if the first measurement results are applied as goal planning the path of the forklift. Hence, in order to increase the accuracy, we are trying to use position-based visual servoing method to dock the forklift in front of pallet. The motion control method based on the position-based visual servoing needs a real-time measurement of the offset between the current pose and target pose, and the motion control rate is designed according to it. It is time-consuming using the ICP algorithm to calculate the pose of the pallet for every frame which could not meet the requirement of real-time visual servoing control. In our method, we mark the center points of the pallet onto the image, and then using the visual tracking method to follow the pallet location in the image. And when we are calculating the pose of pallet, only the points of the three lines on the front face of the pallet are used. If camera misses the tracking target, for example, the pallet being out of vision field, the method applies the pose of last frame to design the forklift's motion track and start the pallet's searching program simultaneously.

The pseudo code version of the method is reported in Algorithm 1.

Algorithm 1 pallet identification

1. acquire scan data
2. filtering by pixel position
3. find lines in 3D space
4. grouping collinear lines
5. **for** all group of points **do**
6. estimate position and posture
7. compute ICP
8. **If** model match successful **then**
9. **goto** 13
10. **end if**
11. **end for**
12. **go to** 1
13. pallet tracking
14. **do**
15. estimate the pose
16. position-based visual servoing control
17. **while** pallet in sight && not reach the destination
18. **if** reach the destination **then**
19. pick the pallet
20. return
21. **else**
22. control the vehicle using odometer
23. **go to** 1
24. **end if**

D. Motion control based on position-based visual servoing

Set the target location as origin and establish a coordinate system which takes the target direction as x axis, the forklift's pose in this coordinate system is shown in Fig.6. Where *r* is the distance between the center of the drive shaft and the target position, α represents the angle between forklift's current orientation and the line connecting the origin of vehicle coordinate system with the target position, θ is the angle between current direction and target direction.

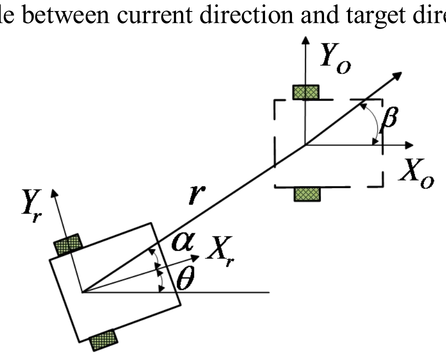

Fig.6 Offset between vehicle and target

Suppose robot's linear velocity is v, angle speed is w, and then the kinematic model can be set as following:

$$\begin{pmatrix} \dot{r} \\ \dot{\alpha} \\ \dot{\beta} \end{pmatrix} = \begin{pmatrix} -\cos\alpha & 0 \\ \frac{1}{r}\sin\theta & -1 \\ -\frac{1}{r}\sin\theta & 0 \end{pmatrix} \begin{pmatrix} v \\ \omega \end{pmatrix} \quad (8)$$

Reference [15] designed the following control rate, and tested that robot could get to the target pose in limited time.

$$\begin{cases} v_{\mathrm{cal}} = k_\rho r \cos\alpha \\ \omega_{cal} = k_\beta \beta + k_\theta \theta \end{cases} \quad (9)$$

Considering that the linear velocity and angle speed are limited in the actual system, we assumed that the highest linear and angle speed of robot to be V_{max} and W_{max}, then:

$$\begin{cases} k_\upsilon = | v_{\mathrm{cal}} / \upsilon_{\max} | \\ k_\omega = | \omega_{cal} / \omega_{\max} | \end{cases} \quad (10)$$

And finally the designed control rate should be:

$$\begin{cases} \upsilon = v_{cal} / \max(k_\upsilon, k_\omega) \\ \omega = \omega_{cal} / \max(k_\upsilon, k_\omega) \end{cases} \quad (11)$$

IV. Experiment And Result

In order to verify the method proposed above, we designed the line structured light sensor based on the embedded visual processing board developed by our lab and mounted this sensor onto a modified automatic forklift, and accomplished the experiment of automatic pallet handling.

A. Line structured light sensor design

The developed line structured light sensor is shown in Fig. 7. The laser projector employs 850nm infrared wavelengths to decrease the effect of visible light, and the back of the narrowband light filter is a digital CCD camera with a distinguishability of 640*480.

Fig. 7 The developed line structured light sensor

We use the Bouguet method implemented in the Caltech Camera Calibration Toolbox to calibrate the intrinsic parameters of the camera. The intrinsic parameters are concluded as (12).

$$K = \begin{bmatrix} 554.87996 & 0 & 320.24167 \\ 0 & 554.48512 & 208.41173 \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

As stated above, the parameters of the laser planes are calibrated through planar target of unknown orientations. Fig. 8 shows one of the calibration images and the fitting results of laser lines corresponding to it. Fig.9 shows the fitting result of laser plane using LS and RANSAC method respectively.

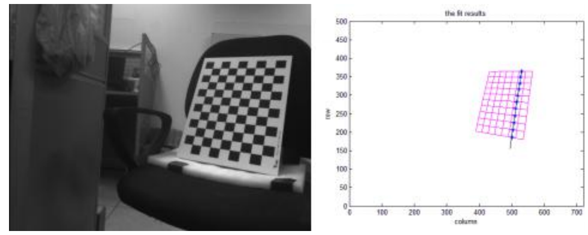

Fig.8 The calibration image and the fitting result of laser line

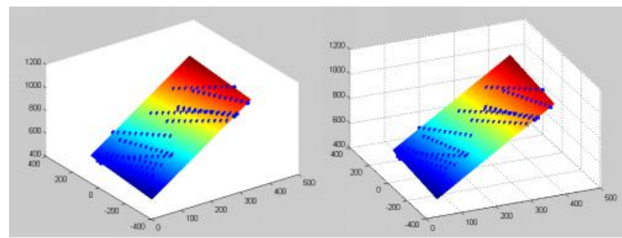

Fig.9 The fitting result of laser plane

The calibration results of laser plane using different method are shown as follows:

$$\begin{aligned} P_r &= [298.6586 \quad 1.8047 \quad 0.1352] \\ P_l &= [299.1866 \quad 1.8030 \quad 0.1326] \end{aligned} \quad (13)$$

A. Automatically Pellet Handling

In order to test the effectiveness of the proposed method, we carried out simulation and visual stabilization experiments based on line structured light sensor. In order to show that the forklift still can dock in the front of pallet in ideal pose although there are errors with the measurement of the structured light, we add random noise to the offset between current pose and target in the simulation experiment. Same as actual vehicle, we set the maximum linear and angle speed of the robot to be *Vmax=1m/s, Wmax=1rad/s,* and set control rate to be $k = (k_\rho, k_\beta, k_\theta) = (0.1, 0.6, -0.4)$ in the simulation experiment. The experiment result is shown in Fig.10 and we can see that the controller realized posture stabilization finally.

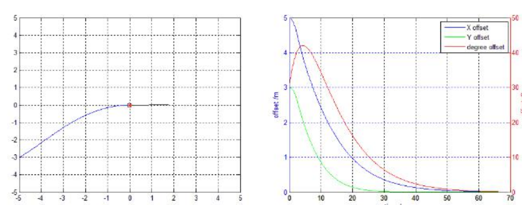

Fig. 10 The simulation experiment result

Then, we mounted the line structured light in the AGV and tested the method through the visual posture stabilization of AGV. The transformation matrix between the coordinate system of vehicle and sensor should be calibrated first. In order to use the sensor easily in other equipment and improve the accuracy of calibration at the same time, we defined a coordinate system on the shell of the sensor which was different from the camera coordinate system. The coordinate system was defined using the center of the shell as original, X axis parallel to the direction of laser, and Y axis parallel to the line connecting between the camera and the laser. Then we calibrate the coordinate system using eye-in-hand calibration method by manipulator. The scan data described in camera coordinate system and plane coordinates are shown in Fig. 11.

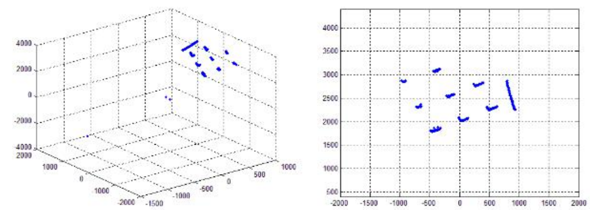

Fig. 11 Scan points in 3 d-coordinate and 2 d-coordinate

V. CONCLUSION

This study proposed a method of pallet automatically handling based on the line structured light sensor, which included the development of line structured light sensor, pallet identification and localization and vehicle visual docking. Unlike other structured light sensor, we realized Hessian matrix decomposition based light stripe center detection in real-time by embedded visual processing board. Besides, we identified and localized the pallet using the geometry structure of it based on model match method, and use position based visual servoing method driving the vehicle approach the pallet.

The method proposed here for identifying and localizing is suitable for the less complicated scenarios. In our future works, we will make further improvement to deal with more complicated scenarios and multiple pallets through the fusion of the camera and line structured light sensor.

REFERENCES

[1] G. Garibotto, S. Masciangelo, M. Ilic, and P. Bassino, "Service robotics in logistic automation: Robolift: vision based autonomous navigation of a conventional fork-lift for pallet handling," in 1997 IEEE 8th International Conference on Advanced Robotics, pp. 781–786, July 1997.

[2] A. Kelly, "Model-based object pose refinement for terrestrial and space autonomy," in 2001 International Symposium on Artificial Intelligence, Robotics and Automation in Space (ISAIRAS01), Montreal, Quebec, Canada, June 2001.

[3] R. Cucchiara, M. Piccardi, and A. Prati, "Focus based feature extraction for pallets recognition," in BMVC, 2000.

[4] S. Byun and M. Kim. "Real-time positioning and orienting of pallets based on monocular vision," in 2008 IEEE International Conference on Tools with Artificial Intelligence - Volume 02, pp. 505–508, 2008.

[5] J. Nygards, T. Hogstrom, and A. Wernersson, "Docking to pallets with feedback from a sheet-oflight range camera," in 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000), pp. 1853–1859, 2000.

[6] M. Seelinger and J. D. Yoder, "Automatic visual guidance of a forklift engaging a pallet," Robotics and Autonomous Systems, 54(12):1026 – 1038, 2006.

[7] L. Baglivo, N. Bellomo, G. Miori, E. Marcuzzi, M. Pertile, and M. De Cecco, "An object localization and reaching method for wheeled mobile robots using laser rangefinder," in 2008 IEEE International Conference on Intelligent Systems, pp. 5–11, 2008.

[8] M.R. Walter, S. Karaman, E. Frazzoli, and S. Teller, "Closed-loop pallet manipulation in unstructured environments," in 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010), pp.5119-5126, 2010.

[9] D. Lecking, O. Wulf, and B. Wagner, "Variable pallet pick-up for automatic guided vehicles in industrial environments," in IEEE Conference on Emerging Technologies and Factory Automation, pp.1169-1174, 2006.

[10] L. Baglivo, N. Biasi, F. Biral, N. Bellomo, E. Bertolazzi, M. Da Lio, and M. De Cecco, “Autonomous pallet localization and picking for industrial forklifts: a robust range and look method,” Measurement Science and Technology, vol. 22, no. 8, p. 085502, 2011.

[11] M. M. Aref, R. Ghabcheloo, A. Kolu, M. Hyvönen, K. Huhtala, and J. Mattila, "Position-based visual servoing for pallet picking by an articulated-frame-steering hydraulic mobile machine machine," in IEEE nternational Conference on Automation and Mechatronics, pp.218-224, 2013.

[12] M. M. Aref, R. Ghabcheloo, and J. Mattila, "A Macro-Micro Controller for Pallet Picking by an Articulated-frame-steering Hydraulic Mobile Machine", in 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 6816-6822, 2014.

[13] C. Steger, "An unbiased detector of curvilinear structures," in IEEE Transactions on Pattern Analysis and Machine Intelligence, pp.113-125, 1998.

[14] F. Zhou and G. Zhang, "Complete calibration of a structured light stripe vision sensor through planar target of unknown orientations," Image and Vision Computing, vol. 23, pp. 59-67, 2005.

[15] R. Siegwart, and R. Nourkhsh, Introduction to Autonomous Mobile Robots: MIT Press, 2004.