Vision-based Autonomous Load Handling for Automated Guided Vehicles

VARGA Robert, NEDEVSCHI Sergiu
Technical University of Cluj Napoca
Email: {robert.varga, sergiu.nedevschi}@cs.utcluj.ro

***Abstract*—The paper presents a method for automatically detecting pallets and estimating their position and orientation. For detection we use a sliding window approach with efficient candidate generation, fast integral features and a boosted classifier. Specific information regarding the detection task such as region of interest, pallet dimensions and pallet structure can be used to speed up and validate the detection process. Stereo reconstruction is employed for depth estimation by applying Semi-Global Matching aggregation with Census descriptors. Offline test results show that successful detection is possible under 0.5 seconds.**

Keywords*—*object recognition; object detection; stereo reconstruction;

I. INTRODUCTION

Integrating Automated Guided Vehicles in industrial environments for transporting materials is becoming more prevalent. Automation leads to cost and time reduction at installation time and during working time. However, this introduces many safety problems and the requirement for precision in localizing the AGV within the facility and other objects around it. Special attention should be paid to the operation of loading and unloading pallets because incorrect positioning could lead to accidents.

The goal of this research is to provide a method for pallet detection and operation point detection. This objective is achieved using a combination of stereo reconstruction methods based on stereo images and object detection from monocular images. The method would serve as a faster and more flexible alternative to laser scanners.

Throughout this paper we will use the following terms: AGV - Automated Guided Vehicle, refers to automated forklifts for logistic operations; operation point - 3D position of the center of the frontal view of the pallet or the future position of the unloaded pallet; load handling - operations pertaining loading or unloading of palletized goods by the AGV.

II. RELATED WORK

The technical literature for the specific task of pallet detection is scarce, almost insexistent. We will discuss related object detection methods and also provide an analysis of current systems in use. The papers [1], [2] discuss solutions for load detection and de-palletizing. Their work focuses on parcel detection and handling using a photonic mixing device (PMD) camera. The authors have identified multiple modalities for automated load detection: no detection at all - inflexible; border detection with laser scanners - slow and cheap; stereo cameras - sensitive to lightning conditions; uncalibrated vision and 3D laser assisted image analysis; 2D range imagery - no orientation avaliable; model based range images; 2D camera and laser scanner.

The first approach we discuss for automatic load handling entails precise positioning of both the AGV and the pallets to be loaded. This eliminates the need to detect the location of the pallet at load time but can lead to positioning errors. Small errors accumulate in time and can lead to accidents due to falling pallets or incorrect gripping. Another alternative is to use laser scanners to estimate the position of the pallet by tracing the contours of objects.

General vision-based object detection methods rely on extracting meaningful and discriminative features that enable the separation of the object from the background. Gradient information is essential for this task and many feature incorporate this by constructing histograms: Histogram of Oriented Gradients [3]; Scale Invariant Feature Transform [4]; Weber Local Descriptors [5]. The calculation of histogram type features can be accelerated by employing integral images [6].

Object detection can be performed with the technique of sliding detection windows. By positioning rectangles at different positions and by changing their size we can investigate whether an object is present or not. For fast classification the two main options are Boosted classifiers [7] with soft cascading [8] and Support Vector Machines with fast kernels such as linear SVM and Histogram Intersection SVM [9].

III. REQUIREMENTS

The automated load handling system is required to perform accurate pallet detection and operation point estimation. It must also provide the orientation of the pallet. The input of the system consists of a pair of images, an operation point position request, information about operation type (the number of pallets, the level, number of reference points, storage type), pallet dimensions, 3D static map. During loading and unloading operations the AGV travels to the operation point and stops at a distance of approximately 2.5 meters. At this position the system must provide the position of the pallets to enable corrections to AGV path. As the AGV approaches the pallets the positions of the pallets are updated online up to a certain distance of approximately 1.8 meters.

For loading and unloading operations the system must detect and provide the 3D position of the pallet or pallets with an accuracy of: 5 cm (@ 1σ) and 1 deg (@ 1σ) at a distance of 2.5 m from the pallets; 1 cm (@ 1σ) and 1 deg (@ 1σ) at a distance of 2 m from the pallets [10]

From the analysis of the requirements we can obtain the minimal requirements for the hardware. In the following we provide details about our hardware components and show that the current setup can satisfy the precision constraints.

A. Hardware components

The two main components of the system are the sensors and the processing unit. We employ two Manta G-223 NIR cameras mounted in canonical horizontal configuration (displaced horizontally and facing the same way). The resolution of the cameras is 2048px by 2048px. The cameras are equipped with Schneider Cinegon lenses with a focal length of 4.8mm and F number 1.8. An auxiliary light source comprised of several LEDs is positioned between the cameras. The cameras are mounted on the AGV behind the forks and are lowered on demand to grant view of scene in front of the forks.

The processing unit consists of an industrial PC ADLINK MXC-6301 which is a high-performance fanless embedded computer integrating a 3rd generation Intel Core i7 processor and QM77 chipset to provide powerful computing and superior graphic performance. The AGV provides a constant frame rate trigger to acquire images from the cameras. When this was not an option - as in our laboratory - we have used an Arduino Uno microcontroller for synchronous triggering of the two cameras. In the following we demonstrate that our current camera setup can provide the accuracy needed.

B. Horizontal field of view

The field of view for the stereo region must be wider than two pallet widths in the working range of 1.5-2.5m. The horizontal field of view at distance Z has the following formula:

$$X_{fov} = 2 \cdot Z \cdot tan(\theta/2) = \frac{Z \cdot \mu \cdot w}{f} \tag{1}$$

where $\theta = 2atan(\frac{\mu \cdot w}{2f})$ is the field of view angle, $\mu = 0.0055$ is the size a single sensor cell and $w = 2048$ is the horizontal resolution of the camera. We calculate this value at 1.5 m and 2.5 m to obtain 3520 mm and 5866 mm respectively. Considering that the longer side of a single pallet has 1200 mm this lateral field of view can contain two pallets starting at minimal working range. Usually pallets are operated from their 800 mm side, this means that up to 4 pallets can fit in the horizontal field of view.

C. Depth resolution

The depth error must be lower than 1cm in the working range. For this we investigate the depth resolution. The stereo reconstruction returns disparities and the depth is inversely proportional to the disparity by the relation:

$$Z = \frac{B \cdot f}{d} \tag{2}$$

where $B = 195mm$ is the baseline length in metric units (the lateral displacement between the two cameras), and $f = 4.8mm = 872px$ is the focal length, and d is the disparity value. This hyperbolic relation entails that depth errors will be larger at larger distances. Starting from 2 we can express the change in depth by differentiating and arranging the terms:

$$\frac{\Delta Z}{\Delta d} = -\frac{B \cdot f}{d^2} = -\frac{Z^2}{B \cdot f} \Rightarrow |\Delta Z| = \frac{Z^2}{B \cdot f} \cdot \Delta d \tag{3}$$

In the previous equation Δd symbolizes the smallest disparity change and is set to $0.25 * \mu$. Plugging in the constants for B and f results in a depth resolution of 3.30 mm at a distance of 1.5 m and of 9.18 mm at a distance of 2.5 m. Both are lower than the required 10 mm and 50 mm respectively.

D. Orientation angle

The angle of interest is the angle formed by the pallet in the xOz plane (the plane parallel to the floor) with the x axis (the horizontal axis). Starting from the expression of the angle we can express the difference in depth that needs to be observed for a standard 800 mm width pallet that is tilted by 1 degree:

$$tan(\alpha) = \frac{z_2 - z_1}{x_2 - x_1} \Rightarrow z_2 - z_1 = (x_2 - x_1)tan(\alpha) \tag{4}$$

This results in a depth difference of approximately 14mm which is within the stereo estimation precision. The angle and depth estimation can be made more robust by fitting a plane to the region of the detected pallets in order to aggregate the results from a larger area.

E. Detection pixel error

The monocular detector must provide the position of the pallets accurately along the x and y direction. From the requirements the positioning error must be less than 1 cm at 1.5 m. Since the change in the position Δx is related to the change in pixels Δu by :

$$\Delta x = \frac{\Delta u \cdot Z}{f} \Rightarrow \Delta u = \frac{\Delta x \cdot f}{Z} \tag{5}$$

Here, the focal length is expressed in pixels: $f = 872px$. We can express the maximum pixel positioning error for the detector to be approximately 5.81 pixels at 1.5 m and 17.45 pixels at 2.5 m. A similar relation holds for y (mm) and v (pixels) but the height of the pallet does not change, so y is relatively constant.

IV. Proposed approach

The proposed approach relies on combining detection from monocular images and stereo reconstruction to estimate the position of the operation point. Stereo depth estimation requires two calibrated cameras and can perform accurate reconstruction under 1 second of the desired scene. By combining intensity and depth features the pallets and other objects can be identified.

In order to provide the operation point the system must perform the following processing steps: stereo image rectification; stereo matching; pallet detection from the left image; stereo reconstruction and plane fitting on the pallet regions that results in the operation point in 3D. Rectification and reconstruction requires the intrinsic and extrinsic camera parameters from a calibration procedure. For pallet detection we follow the standard pipeline for object detection: preprocessing, candidate generation, feature extraction, classification, refinement and verification. The stereo matching algorithm relies on the work of Hirschmuller et. at from [11]. In the following we describe each step of the detection algorithm. The most important parameter settings are given in Table I.

A. Stereo Matching and Reconstruction

For stereo matching we calculate the Census transform of the rectified image pairs [12], [13]. The local cost volume will be the Hamming distance of the Census descriptors. Hirschmuller's Semi-Global Matching [11] aggregation method is applied to the cost volume. The energy function is minimized by considering 4 main propagation directions: to the left, to the right, downwards and upwards. The four propagation steps are performed simultaneously on four separate threads. The aggregated cost volume is checked for consistency, quadratic subpixel interpolation is applied and lastly the disparity image is filtered with a median filter. We are also experimenting with other stereo matching methods: standard Sum of Absolute Differences; fast normalised cross-correlation [14], scanline dynamic programming matching [15], [16].

B. Pallet detection steps

We model the pallet by three legs separated by two pockets, see Figure 1. The relative position of the legs and the aspect ratio are given by the pallet dimension specifications and also from experiments. This model corresponds to the frontal view of a standard Euro pallet with two pockets. In theory, this model is simple and one would expect that regions A, C and E will have the same visual features. However, in practice there is a lot of variability in appearance both for pallet legs (due to material hanging from pallet load, different colorings of pallets) and especially for pallet pockets. Pocket regions can be pitch dark when loading from the ground floor, but can also be the brightest zone from the image when the light is coming from behind the pallet (loading from first floor with the windows behind the rack). Because we are using grayscale images for detection in some cases the background and the pallet may have similar intensity and texture. In the following we describe each phase of the detection method.

In the preprocessing step the image is filtered by a gaussian kernel and histogram equalization is performed. Histogram equalization was found to be useful, even when the exposure time of the camera was properly set. Since the upper half of the input image is covered by forks, and also, this part does not contain useful information we disregard it from the processing steps. For similar reasons the lower part and the left and right extremities are not useful. It follows that all processing is done on a central stripe. Even though we use only a restricted region of the large image, high image resolution is important for precise localization.

Candidate generation is the operation that provides candidate bounding boxes for the classifier. Exhaustive checking of every possible bounding box would be unfeasible and it is also not necessary. The main cue for candidate is the image gradient. To obtain the gradient we filter the image with Sobel filters, one for each axis and threshold the images at $T = 10$. This low threshold is necessary to capture all pallets. Since we always work with a near frontal view of the pallet, only horizontal and vertical edge detection is necessary. Sobel filter was found to perform best compared to other filters. Canny edge detection could be used, but the parameters need to be changed adaptively and it also takes more time. The next step is to find horizontal guidelines that restrict the search space drastically. For this we accumulate vertical gradient values (horizontal lines) along each line of the image. The local maxima of this will represent dominant horizontal lines. By controlling the number of neighbors that are considered (vnh) for local maximum detection we can adjust the number of guidelines that will be generated. After the horizontal guidelines are detected vertical lines are searched only between two horizontal lines. Vertical lines are detected when the sum of horizontal gradient values along the line are above a threshold $P = 0.1$ percent. We also restrict the length of these lines - which corresponds to the pallet height - based on empirical data. After all vertical lines are detected we consider only rectangle candidates that have lines at extremities. It is also required for a candidate to have vertical lines inside corresponding to the pallet pockets as described by the pallet model. The position of the interior lines is permitted to change in a small interval. Even after enforcing the before mentioned minimal requirements on the possible candidates still a large number of candidates remain (more than 300000 for a single image).

Feature extraction module extracts the same feature vector for each candidate. We use 4 types of features: difference of mean intensity, standard deviation of intensity, mean edge strength along x/y directions, mean disparity and disparity difference. Each feature is calculated from a rectangular region efficiently using integral images. For disparity features we ignore invalid disparities and adjust the region area accordingly. Each feature is normalised by the area of the rectangle. The resulting feature vector consists of $M = 24$ features (see Figure 1): channel 0: mean intensity on region A; channels 1-4: mean intensity difference between regions B-E and A; channels 5-9: standard deviation of intensity on regions A-E; channels 10-18: edge strength for each left, lower and right border of the regions A, C and E (green rectangles); The edge strength is the area normalised sum of vertical or horizontal gradient values; channel 19: mean disparity on region A; channels 20-23: mean disparity difference between regions B-E and A.

The classifier scores each candidate rectangle by applying the boosted classifier. AdaBoost learning is performed with $N = 1000$ two-level decision trees. Cascaded prediction can be used for up to 10-100 times faster execution time but we have found that in practice it is difficult to estimate the rejection thresholds. All candidates with scores less than a given threshold θ are discarded. On the remaining rectangles we perform non-maximum suppression based on the percentage of overlap and classification score. This step is necessary because multiple detections may correspond to a single object and we want to retain only the detection with the highest score. The overlap criteria is $\frac{R_1 \cap R_2}{R_1 \cup R_2}$, where the numerator is the intersection and the denominator is the union of the two rectangles in question.

To train the classifier we rely on a manually labeled dataset with pallets as positive examples and any other regions as negative examples. For this we have labeled video sequences of loading and unloading operations from the Elettric80 Viano warehouse. Each pallet is indicated by a manually drawn rectangle. During the classifier training we generate candidates from each image that has ground truth information. The rectangles that are sufficiently close to actual pallets are used as positive examples while all other regions are considered negative examples. We select a random 50-5000 negative examples from each image. We can perform bootstrapping rounds by evaluating the classifier on images that do not contain pallets for additional hard negative samples and retrain the classifier.

C. Detection postprocessing

There are many correction methods that can be applied on the generated detections to increase the robustness of the method.

First, the expected positions of the pallets are known in advance. For normal loading operation the pallets occupy a central position and should also have the same position along the y axis because moving towards the pallets does not change their vertical position. This information can be used to favor rectangles that are close to these central positions. We penalize the scores of detected pallets by a factor that is proportional to the distance from the expected locations.

Second, since information about the number of pallets from the scene is available, this can be used to eliminate false positives. Let this number be n, then we set the classifier threshold to a low value and we only retain the first n rectangles according to score values.

Third, we can perform detection on both the left and right image and find pair correspondences to validate correct detections. This could also help in estimating the distance to the pallet. Pallet pair should appear at the same height and the displacement should coincide with disparity values.

Fourth, we apply tracking and temporal integration to smooth out the positioning errors. In the case of temporal integration we perform detection on multiple frames then choose the mean values for overlapping detections and eliminate detections that do not have temporal continuity. For tracking we apply a Kalman filter on 4 dimensions (x,y, width, height).

D. 3D pallet position and orientation

Once detections are available in the form of rectangles the 3D position of the pallet can be estimated by making use of stereo information. The plane of the frontal view of the pallet is approximated by Least Squares or RANSAC from the disparity image. This plane is fitted only to the pallet points from the legs of the pallet. It provides a more accurate estimation for the distance to the pallet. The orientation angle can be extracted by sampling the plane at extremities and calculating the angle.

E. Implementation details and optimization

All processing modules are implemented in C++, compiled with Visual Studio 2010 compiler with OpenMP multithreading features enabled. Other settings include: fast code optimization enabled, fast floating point model, omit frame pointers. OpenCV 2.4.5 is the chosen library for image processing functions. Essential for fast execution is the reliance on integral images to compute feature sums, predicting with a boosted classifier and code parallelization.

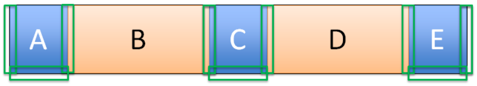


Fig. 1: Model of the pallet employed for detection - the three regions A,C and E correspond to pallet legs, while B and D are pockets

TABLE I: Relevant parameters of the detection algorithm

Parameter	Description	Value
N	number of weak learners	1000
M	number of features	24
T	Sobel edge threshold	10
P	threshold percent for line detection	0.1
vnh	vertical neighborhood size for NMS	5
hnh	horizontal neighborhood size for NMS	3
roi	region of interest (x,y,x2,y2)	(300,980,1748,1225)
pos	relative vertical line positions	{0,0.125,0.42,0.58,0.875,1}
B	Baseline length	195 mm
f	focal length	4.8 mm
P_1	penalty for small disparity change	5
P_2	penalty for large disparity change	100

V. Experimental results

In this section we present results from detection tests and distance estimation tests. These are the two essential tasks that need to be resolved.

For pallet detection we evaluate the detection rate and false positive rate of different classifiers on the acquired datasets. We will refer to the two dataset as Viano1 and Viano2. These consist of image sequences of loading scenes from the warehouse of Elettric80 at Viano. Detections generated by the automatic pallet detector are matched to the manually labeled pallets from each image. Scoring is based on the intersection and the union of the two rectangles. Let A be the area of the detected rectangle, B be the area of the ground-truth rectangle representing the pallet, let C and D be the area of their intersection and union respectively ($D = A + B - C$). The following evaluation criteria are used:

- strong positive match - corresponds to a high absolute overlap in the x direction, and a high relative overlap in the y direction (z axis for the AGV):

$$|D.width - C.width| < 10px \tag{6}$$

$$C.height/D.height > 0.7 \tag{7}$$

- weak positive match - corresponds to a moderate/high absolute overlap in the x direction, and a moderate relative overlap in the y direction (z axis for the AGV):

$$|D.width - C.width| < 15px \tag{8}$$

$$C.height/D.height > 0.5 \qquad (9)$$

- strong false positive - corresponds to a low overall overlap:

$$C.width/D.width \cdot C.height/D.height < 0.4 \quad (10)$$

- weak false positive: otherwise

We opt for the absolute difference in the x direction since the precision is crucial on the x axis. The 10 pixel threshold corresponds to approximately 2.1 cm at 2 m with our current setup. We then calculate the detection rate for the strong positives and weak positives. The false positive rate takes into consideration only the strong false positives since the weak false positives give the correct position of the pallet at all times but fail the precision constraint.

Using these criteria we have obtained on the Viano1 test set a detection rate for weak positives of **94%** (5007 out of 5333); a detection rate for strong positives of **84%** (4476 out of 5333); a false positive rate of **1.5%** (80 out of 5333).

Table V shows the test results on the Viano2 test set, which contains 6707 labeled pallets from 37 different scenes. The majority of misses are due to pallets at non-standard distances, reflections from plastic covering the pallets, other objects similar to pallets present in the images. The influence of each parameter was tested. It can be observed that histogram equalization drastically improves the detection performance. Introducing stereo features and enforcing correct aspect ratios help with more exact localization (higher strong true positive rate).

For distance estimation precision we first perform online pallet detection on a dummy pallet. The distance to the detected pallet is estimated at given positions in the working range (1000-2500 mm). Each distance is measured at least 100 times in slightly varying lighting conditions to evaluate the standard deviation of the measurements. This test shows (see Table II) that all errors are below 1.5 cm at 1 standard deviation. An increase in precision can be obtained by further tuning the parameters from the stereo reconstruction. We can also correct the distance values by estimating the model of the error as a function of the stereo disparity value.

Additional tests were performed to determine whether or not the selected classifier is suitable for the detection task. Several classifiers were considered starting with a simple manually tuned decision stumps which are constraints on the features (such as the distance of the left and right legs must be the same). This classifier can be used as a baseline and it is the fastest one. We consider several variants of boosted classifiers with 100-1000 weak learners and cascaded prediction. SVM is also applied (using the *libsvm* library) with linear kernel function, even though the execution time is large. Table III compares the classifiers on a smaller dataset Viano2. The execution time provided is the time required to make 300,000 predictions, which is the typical number of candidates for a thorough and reliable detection. A comparison of the DET (Detection Error Tradeoff) curve for sevaral classifiers is illustrated in Figure 4. We provide the area under the curve up to the 10^0 mark as a single performance metric - the lower the value the better the classifier. Even though the cascaded version performs best, in practice the rejection thresholds may cause premature the elimination of true positives. Table IV compares the execution time of the major processing steps in two configurations: 4 directional stereo aggregation with boosted 1000 decision trees; and 2 directional stereo aggregation with 1000 cascaded decision trees.

TABLE II: Distance measurements and stereo depth estimation - all values are in millimeters

Real	**Avg. absolute**	**Max. absolute**	**Standard**	**Error**
distance	**error**	**error**	**deviation**	**at 1 σ**
1000	1.84	4.11	0.91	2.74
1500	2.59	13.15	2.14	4.73
1800	3.00	10.18	3.32	6.32
2000	6.18	8.87	0.90	7.08
2100	2.40	14.28	2.93	5.33
2100	2.40	14.28	2.93	5.33
2200	4.85	20.69	5.67	10.53
2300	10.27	24.77	3.89	14.17
2400	3.96	15.02	4.46	8.42
2500	6.74	24.40	8.25	14.99

TABLE III: Detection rate, false positive rate and execution time for 300k instances of different classifiers on the Viano2 training set

Classifier	**Detection**	**False positive**	**Execution**
tpye	**rate [%]**	**rate [%]**	**time [ms]**
Boosted 1000 DTs	99.81	0.18	3140
Cascaded 1000 DTs	99.07	0.19	140
Boosted 100 DTs	98.69	0.74	180
Linear SVM	58.69	39.25	55000
Manual Decision Stumps	51.40	37.57	4

TABLE IV: Typical execution time of each processing step - rectification is performed on the full 2048px by 2048px image while other steps only on the region of interest. Total time includes other minor processing.

Step	**Boosted1000 [ms]**	**Cascaded1000 [ms]**
Rectification	35	35
Stereo matching	338	190
Pallet detection	3788	195
Total	4180	430

VI. Conclusion

This paper presented a solution for automatic pallet detection by combining stereo reconstruction and object detection from monocular images. The selected method can perform detection under 1 second and the tests show that it can provide the desired accuracy. We demonstrate the performance of the method by evaluating it on recorded sequences from a real warehouse. The optimal classifier based on experimental results is an ensemble classifier with 1000 boosted decision trees, with cascaded prediction that relies on both stereo and intensity features.

Future work will involve improving the robustness of the detection method by automatically setting the exposure time for the cameras, by improving the classifier's precision and by

TABLE V: Viano2 test set (0 means not used): blur - Gussian blur; h. eq. - histogram equalization; stereo - stereo disparity features; negs - number of negative samples per each training image; center - penalize detections that are far from the center; aspect - enforce good aspect ratio; wtp - weak true positive rate; stp - strong true positive rate

blur	h. eq.	stereo	negs	center	aspect	wtp	stp
1	1	1	5k	0	0	95.14%	76.29%
1	1	1	5k	1	1	94.50%	78.70%
1	1	1	5k	0	1	94.88%	**79.28%**
1	1	1	5k	1	0	94.70%	75.53%
0	1	1	5k	0	0	91.36%	75.91%
1	0	1	5k	0	0	72.83%	56.85%
1	1	0	5k	1	0	**96.07%**	78.18%
1	1	1	500	0	0	92.12%	73.61%

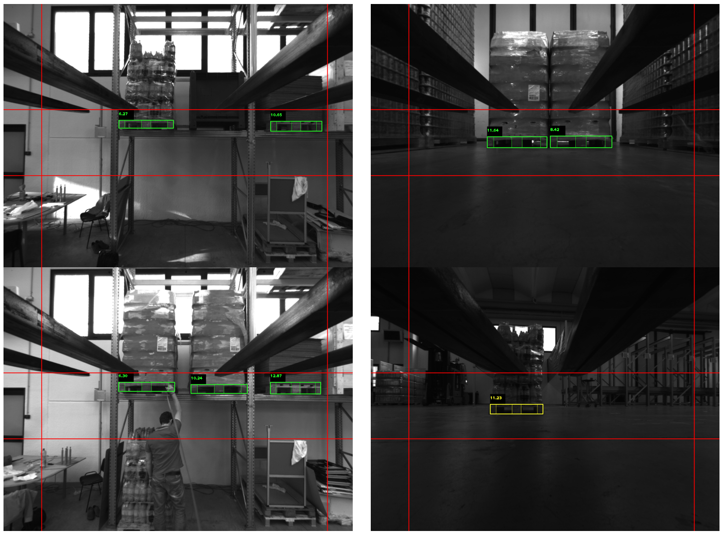
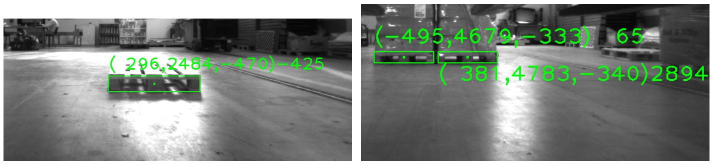


Fig. 2: Sample detections from the Viano1 dataset. The numbers indicate the score (confidence) of each detection.



(a) overexposed image of a bare pallet

(b) material covering the right pallet

Fig. 3: Successful detections from Viano2 in difficult cases. 3D point and orientation angle (in centidegrees) is indicated for each pallet.

validating the detections using multiple methods. Several of the validation methods still need to refined and implemented. Further online tests with actual AGV loading operations are required to validate the experimental results.

VII. Acknowledgment

PAN-Robots is funded by the European Commission under the 7th Framework Programme Grant Agreement no. 314193.

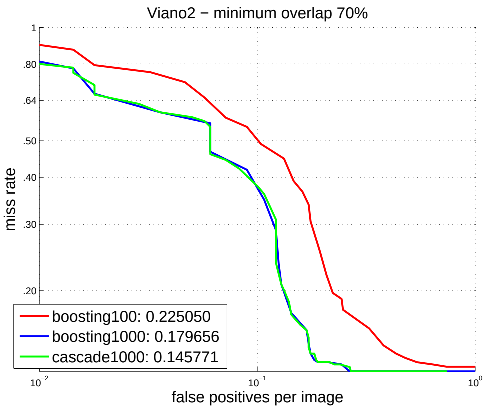


Fig. 4: DET curve on the Viano2 dataset using different classifiers and the areas under each curve

The partners of the consortium thank the European Commission for supporting the work of this project. We would like to thank our partners from Elettric80 for the opportunity to record sequences at their warehouse from Viano.

References

[1] F. Weichert, S. Skibinski, J. Stenzel, C. Prasse, A. Kamagaew, B. Rudak, and M. ten Hompel, “Automated detection of euro pallet loads by interpreting PMD camera depth images,” *Logistics Research*, vol. 6, no. 2-3, pp. 99–118, 2013.

[2] C. Prasse, S. Skibinski, F. Weichert, J. Stenzel, H. Müller, and M. T. Hompel, “Concept of automated load detection for de-palletizing using depth images and RFID data,” pp. 249–254, Nov. 2011.

[3] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *CVPR*, 2005, pp. I: 886–893.

[4] D. G. Lowe, “Object recognition from local scale-invariant features,” in *ICCV*, 1999, pp. 1150–1157.

[5] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, “Wld: A robust local image descriptor,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 32, no. 9, pp. 1705–1720, 2010.

[6] F. M. Porikli, “Integral histogram: A fast way to extract histograms in cartesian spaces,” in *CVPR*, 2005, pp. I: 829–836.

[7] Schapire, “The strength of weak learnability,” *MACHLEARN: Machine Learning*, vol. 5, 1990.

[8] L. Bourdev and J. Brandt, “Robust object detection via soft cascade,” in *CVPR*, 2005, pp. II: 236–243.

[9] S. Maji, A. C. Berg, and J. Malik, “Classification using intersection kernel support vector machines is efficient,” in *CVPR*, 2008, pp. 1–8.

[10] “Pan-robots plug and navigate robots for smart factories: User needs and requirements,” 2013.

[11] H. Hirschmuller, “Accurate and efficient stereo processing by semi-global matching and mutual information,” in *CVPR*, 2005, pp. II: 807–814.

[12] R. Zabih and J. Woodfill, “Non-parametric local transforms for computing visual correspondence,” in *ECCV*, 1994, pp. B:151–158.

[13] C. D. Pantilie and S. Nedevschi, “SORT-SGM: Subpixel optimized real-time semiglobal matching for intelligent vehicles,” *IEEE T. Vehicular Technology*, vol. 61, no. 3, pp. 1032–1042, 2012.

[14] M. J. Hannah, “Computer matching of areas in stereo images.” Ph.D. dissertation, Dept. of Computer Science, Stanford University, 1974.

[15] Y. Ohta and T. Kanade, “Stereo by intra- and interscanline search using dynamic programing,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-7, no. 2, pp. 139–154, Mar. 1985.

[16] J. C. Kim, K. M. Lee, B. T. Choi, and S. U. Lee, “A dense stereo matching using two-pass dynamic programming with generalized ground control points,” in *CVPR*, 2005, pp. II: 1075–1082.