

Stereovision Based Obstacle Detection System for Unmanned Surface Vehicle

Han Wang, *Senior Member, IEEE*, and Zhuo Wei

Abstract—This paper presents the stereo based obstacle detection system for unmanned surface vehicle (USV). The system is designed and developed toward the aim of real-time and robust obstacle detection and tracking at sea. Stereo vision methods and techniques have been employed to offer the capacity of detecting, locating and tracking multiple obstacles in the near field. Field test in the real scenes has been conducted, and the obstacle detection system for USV is proven to provide stable and satisfactory performance. The valid range with good accuracy of depth estimation is from 20 to 200 meters for high speed USV.

Index Terms—Stereovision, Obstacle Detection, USV.

I. INTRODUCTION

In the recent decades, one of the largest threats to the national security is attack by sea, either from the pirates or the terrorists. To protect the ships and coast line, unmanned surface vehicles (USV) are expected to act important role in future operations in place of human, to detect, track and even eliminate the potentially-harmful invaders. Functioning on the surface of water, the USV may be required to complete various tasks autonomously so it is important for a practical USV to include situational awareness by sensing the immediate environment to locate obstacles. Hence, to achieve the capability, the stereovision based obstacle detection system for USV is developed.

Stereovision based obstacle detection on the sea for surface vehicle has received much attention in recent years. In the work by J. Larson et al. ([1], [2]), stereovision sensors are used to produce the profile of 3D point data for obstacle detection on the USV platform. The decision on the obstacle is based on its range and height. T. Huntsberger et al. [3] has reported a stereovision system, named the Hammerhead, that detects the objects protrude above the waterline. The Hammerhead system is the second generation version of the system used by J. Larson et al., and has wider field of view by using two sets of stereo cameras. However, there is no evaluation of depth estimation accuracy since ground-truth data are not available.

In our previous work [4], we presented an obstacle detection method based on color salient detection and sparse motion analysis. In practice it worked well in most of the test scenarios.

However, when the visibility at sea was low, the salient detection could fail. Moreover, in the scenario of target following, the stern waves appeared constantly in the camera view that tended to cause false detection. Another drawback was that the method might fail to distinguish occluded obstacles.

To overcome these limitations, we have developed the obstacle detection system by introducing stereovision based detection and tracking methods. The capacity of detecting, locating and tracking multiple obstacles in the near field is offered with good accuracy of depth estimation, in the range from 20 to 200 meters, for high speed USV. Details of the stereovision based obstacle detection system and supporting results on the performance evaluations are in the remaining sections.

II. SYSTEM ARCHITECTURE

The hardware of the obstacle detection system is composed of a pair of Point Grey CCD cameras with wide angle lens. The cameras are mounted parallel 1.7m apart on a firm aluminum alloy bar to construct a stereo rig. A Trimble SPS351 GPS receiver is integrated to provide the ego position and heading of the USV for obstacle localization. Since large volumes of data are to be processed in the system, the software is implemented with multi-threading programming to boost the real-time performance. A user-friendly GUI has been created to display the left and right camera views, the obstacle map, and the navigation device outputs.

In each computational iteration, rectified image pair from the stereo camera is captured to reconstruct the scene in the field of camera view. The sea surface plane is then estimated, and obstacle detection is performed on the plan-view grids integrating height and occupancy information. An image feature based particle filter is used to assist the data association in obstacle tracking. The depth estimation of small obstacles in the camera view is refilled by a multiple window correspondence scheme to enhance accuracy.

A. Stereo Correspondence

In order to reconstruct the scene in 3D, we apply a block-matching based dense stereo correspondence method which is on the basis of work by K. Konolige [5]. The algorithm searches for only strong correspondences and rejects correspondences with low texture confidence and uniqueness ration. It is fast and easy to implement, with acceptable

Manuscript received August 21, 2013. This work was supported in part by Singapore Technologies Electronics Ltd.

The authors are with the School of Electrical & Electronics Engineering, Nanyang Technological University, Singapore 639798 (e-mail: hw@ntu.edu.sg; weizhuo@ntu.edu.sg).

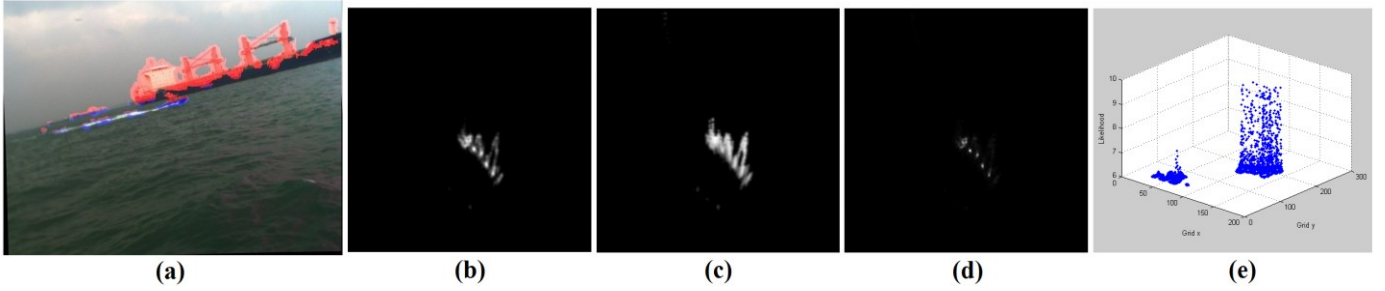


Fig. 1. (a) Scene of obstacle detection. (b) Occupancy grid of the scene. (c) Height grid of the scene. (d) Likelihood grid of the scene. (e) 3D visualization of the likelihood grid.

accuracy. The 3D positions of the matched points are referred to the left camera reference system.

In our system, the correspondence is done on 8-bit gray-level images of 640 by 480 pixels. The maximum disparity search is limited to 64 pixels in consideration of the processing speed and minimum depth constraint.

B. Sea Surface Estimation

Thinking intuitively, the maritime obstacles protrude above the sea surface. Although the sea surface is not exactly planar, it can be considered flat in general. Hence, the sea surface shall be estimated in the first place.

By using the random sample consensus (RANSAC) fitting method, the sea surface plane can be estimated from the 3D point set of the reconstructed scene. To reduce the computational load, the matched points on the image are sub-sampled by a factor for plane fitting. This process shortens the plane fitting time without affecting the performance. The method is simple and straight-forward, yet robust and accurate. When the sea surface plane is extracted, the obstacles can be extracted from the scene by a simple threshold of 3D point distance to the sea surface plane. Figure 1(a) shows the classification results by the plane fitting that the obstacle pixels are labeled in red color and the non-obstacle pixels are labeled in blue.

C. Obstacle Detection

The obstacle detection is performed on plan-view grids integrating occupancy and height information. By using the rotation matrix computed from the sea surface normal, all of the reconstructed 3D points of the obstacles are transformed to a new coordinate system in which the XZ plane is aligned to the sea surface. The XZ plane defines the grid map, and divides a region of the sea surface plane into a set of cells of fixed size. In this case, the cell size is set to 1-by-1 meter, and the grid map size is set to 500-by-500 meters. The points projected in each cell is defined as

$$C_{(x,y)} = \{i | x_i = x, y_i = y, Z_i \in [0, 500]\} \quad (1)$$

where Z_i denotes the depth of the i th point in the point set $C_{(x,y)}$ at location (x, y) on the grid.

The occupancy grid $O(x, y)$ is defined by the amount of 3D points above the plane projected in the cells of the grid. The cell value is calculated as

$$O(x, y) = \text{num of } C_{(x,y)} \quad (2)$$

The basic idea of the occupancy grid representation is that each cell stores a probabilistic estimate of the state which is the presence of an obstacle at the location in the perceived environment.

Another grid is built with the height information from the stereo data. On the height grid, each cell holds the largest height value of the points projected in the cell. The height equals to the distance of the point to the sea surface plane in meter, and the height is thresholded at a maximum value h_{\max} . A vacant cell is assigned with a minimum value h_{\min} .

$$H(x, y) = \begin{cases} \min(\max(\text{dist}_i | i \in C_{(x,y)}), h_{\max}) & \text{if } C_{(x,y)} \neq \phi \\ h_{\min} & \text{if } C_{(x,y)} = \phi \end{cases} \quad (3)$$

where $\text{dist}_i | i \in C_{(x,y)}$ denotes the distance from the i th point in the point set $C_{(x,y)}$ to the sea surface plane at location (x, y) on the grid.

After the occupancy and height grids generated, a Gaussian smoothing operation is performed on the grids respectively in order to decrease the influence of stereo errors. Figure 1(b) and (c) show the resultant occupancy and height grids of the scene in Figure 1(a), respectively. The black cells represent empty regions while grey cells represent occupied areas. The brighter is the color of a cell, the higher value of the cell is.

We implement the likelihood measure of grid cells proposed by R. Muñoz-Salinas [6]. The measure combines the occupancy and height information in determining the likelihood of a grid cell to be the center of an obstacle. The method searches for regions in which the occupancy and height values fall into the certain normally distributed ranges. The likelihood measure $L(x, y)$ at location (x, y) s on the grid is defined as

$$L(x, y) = \frac{\exp\left(-\left(\frac{(O(x, y) - \mu_o)^2}{2\sigma_o^2} + \frac{(H(x, y) - \mu_h)^2}{2\sigma_h^2}\right)\right)}{2\pi\sigma_o\sigma_h} \quad (4)$$

where μ_o and σ_o denote the expected mean and standard derivation of the occupancy grid $O(x, y)$ respectively; μ_h and σ_h denote the expected mean and standard derivation of the height grid $H(x, y)$ respectively. Figure 1(d) is an example likelihood grid of the scene in Figure 1(a). The black pixels represent cells whose likelihood is zero. The brighter is the color of a cell, the higher is its likelihood value. For the sake of

clarity, 3D visualization of the likelihood grid in Figure 1(d) is shown in Figure 1(e). The regions with high likelihood values indicate the presence of the boat target and the containership. Local peaks in the likelihood grid suggest the center positions of the obstacles.

When the likelihood grid is obtained, the obstacle detection can be done by thresholding operations and morphological operations on the likelihood grid. With proper settings of the processing parameters, the obstacle detection turns out to be precise and stable. Rectangular bounding box of an obstacle on the grid and its corresponding bounding box on the left image frame are given as the obstacle detection results.

D. Tracking by Data Association

The obstacle tracking mainly relies on the data association of detected obstacles by using their location on the grid map and image color information.

The detected obstacles are arranged in the detect list and fed to the obstacle tracking module. A detected obstacle will be firstly put into the wait list other than immediately confirmed as tracked. An item in the wait list will be constantly associated with the new incoming items in the detect list in the following computational time steps. If the item fails to associate with any detected obstacles within the certain period of time, it will be dropped from the wait list. This criterion helps to exclude the false detection.

We use the size and location of the detected obstacle on the grid map as measures for association. The constraint parameters vary in a non-linear manner depending on the depth estimation of the obstacle. Furthermore, a Hue-Saturation-Value (HSV) color histogram [7] is also employed to associate the items in the lists, which helps to minimize the association errors. The HSV histogram is populated with $(N_h N_s + N_v)$ bins. When the saturation and value of HSV are low, the color cannot be considered a good descriptor. Thus, the pixels with saturation and value lower than two thresholds set to 0.1 and 0.2 respectively, are ruled out in the $N_h N_s$ -bin calculation, but included in populating additional value-only N_v bins.

Once an item in the wait list is confirmed as tracked, it will be relocated in the track list. An item in the track list is to be associated with the detected obstacles in the similar way. The procedures of data association work well in obstacle tracking with stable detection data. However, tracking by data association could fail when it comes to the partial occlusion of obstacles or incorrect stereo correspondence. An image feature based particle filter is applied to resolve the issue.

E. Image Feature Based Particle Filter

A sequential importance resampling particle filter is implemented. The particle filter tracker works with the left camera image using two image features.

1) Transition model

A second-order autoregressive model is used to model the

obstacle dynamics transition, including the coordinates and scaling of the bounding box of the obstacle in the image frame. It generates the prediction of the state with previous tracking results with the noise. The noise in the transition model is assumed to be Gaussian.

In practice, we limit the scaling factor within a narrow range. The variance in horizontal direction is considered to be smaller than that in the vertical direction due to the effects of wave motion. At each time step, 200 particles are initialized.

2) Observation model

The observation model is built by utilizing two image features. One is the HSV color histogram, which is also used in the data association. The other is an edge orientation histogram, which is a simplified variant to the SIFT feature [8]. It is simple and fast to compute, and invariant to rotation in some amount.

The edges in the gray-level image are extracted using the Sobel operators in the horizontal and vertical direction: The magnitude and the orientation of the edges are then calculated. Only the edges with adequately large magnitudes are included in populating the orientation histogram in order to remove weak noisy edges. The bounding box that defines the obstacle in the image frame is divided to 4 sub-regions, aligned in a 2-by-2 grid. For each sub-region, an edge orientation histogram is built with 8 orientation bins. Thus, the descriptor is an orientation histogram with 32 bins in total.

3) Likelihood measure

Both of the image features are in the form of one-dimension histogram. The similarity between the obstacle feature, k' and the candidate feature of the particle, k_i can be measured by the Euclidean distance between them. The likelihood distribution can be defined by

$$L(z|x) \propto \exp\left(-\frac{\|k'-k_i\|_2^2}{\sigma^2}\right) \quad (5)$$

where $\|\Delta\|_2$ denotes the L2 norm; σ denotes the standard deviation of the Gaussian noise in the measurement; z denotes the measurement vector.

Under the assumption that the two features are independent to each other, the overall likelihood of a particle is computed as a product of the likelihoods of the two features [9].

$$L(z_{overall}|x) = L(z_{color}|x) \cdot L(z_{edge}|x) \quad (6)$$

4) Tracking with particle filter

The image feature based particle filter tracker is capable of handling the situations in which the data association tends to fail. When the association restores correctly from failure, the particle filter disengages from tracking and let the data association take over since data association is more computationally efficient. Nevertheless, if the data association fails to restore, the obstacle will be tracked by the particle filter for a while before dropped from the track list and considered as track lost. This combined tracker works robust and efficient in real time.

F. Multiple Window Correspondence

Ares-based correspondence method using fixed size window suffers from some drawbacks that, if the window size is too small and does not include enough intensity variation, it gives poor estimate; on the other hand, if the window size is too large and contains large disparity variation, it may not represent correct matching due to different projection distortions [10]. To overcome the disadvantages, a multiple window correspondence method is introduced to compute the disparity of the small tracked obstacle. Multiple windows for correspondence are generated by expanding the original rectangular bounding box of the obstacle in the vertical direction, horizontal direction, or both. By repeating the expanding several times, we can get a set of windows of different sizes and size ratios.

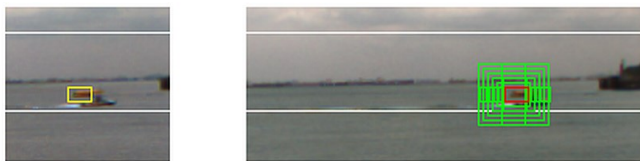


Fig. 2. Multiple window correspondence.

In the left image of Figure 2, the region of interest of the obstacle is given by a yellow rectangular bounding box. Multiple windows of different sizes and size ratios are generated based on the yellow box as the template window. Each window slides along the horizontal epipolar line through the line search range, and the normalized cross-correlation is calculated. The search range is narrowed down by the disparity data from the dense stereo correspondence done previously. There are a set of arrays of matching scores calculated. The maximum in each array is picked as the candidate match (the green boxes in Figure 2 indicate the corresponding candidate matches in the right image). Populating the candidate values in a histogram, the best match can be suggested (the red bounding box in Figure 2). The method helps to reject false matches, and offers sub-pixel accuracy at the same time.

III. EXPERIMENTS

The experiments in the actual scenes have been done to determine the performance and accuracy of the obstacle detection system in detecting and tracking obstacles, and estimating depth. A speed boat cruising at different speed and azimuth is as the target. The positions of the boat target and the USV are logged respectively by onboard Differential GPS with the same time reference as the stereovision based obstacle detection system.

Shown in Figure 3(a)-(d), a boat target traveled from right to left in the camera view, and passed behind a floating buoy. Partial occlusion occurred when the boat was running behind the buoy (Figure 3(b)-(c)). Both of them have remained tracked with unique IDs unchanged. The partial occlusion scene is successfully handled by the system.

The system has been tested in other scenarios as well. Figure 4(a) shows that the boat target was moving forward ahead in the distance while the USV is following. The scenario demonstrated in Figure 4(b) is that the boat target was running from left to right in the camera view while a containership was in the background. The tracking of the boat target was not affected by the background. Even the boat target running at high speed close to the USV can be detected and tracked properly (Figure 4(c)). White waves and splashes generated by the speed boat did not influence the performance. Moreover, the system can respond quickly to high speed incoming target in the field of view (Figure 4(d)).

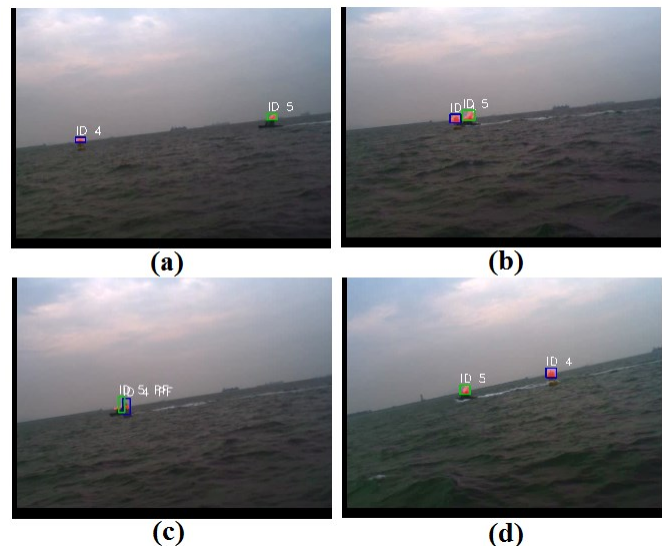


Fig. 3. Boat partially occluded by buoy. (a) Before partial occlusion. (b)-(c) Partial occlusion occurred. (d) After partial occlusion, both targets are still tracked.

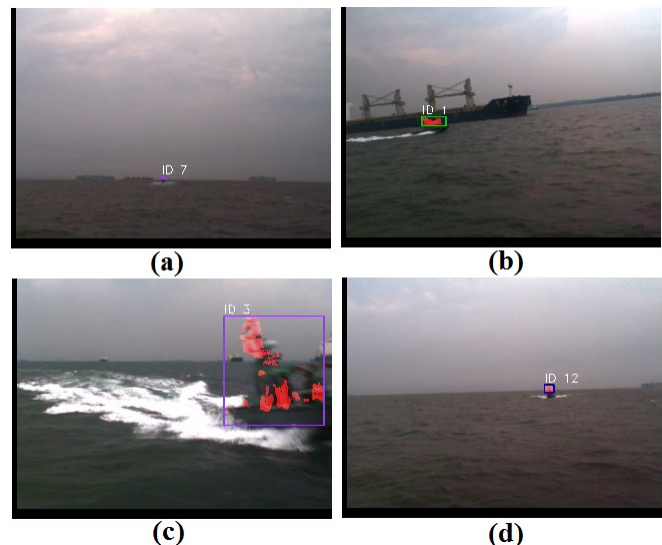


Fig. 4. More test scenarios of obstacle detection on the sea.

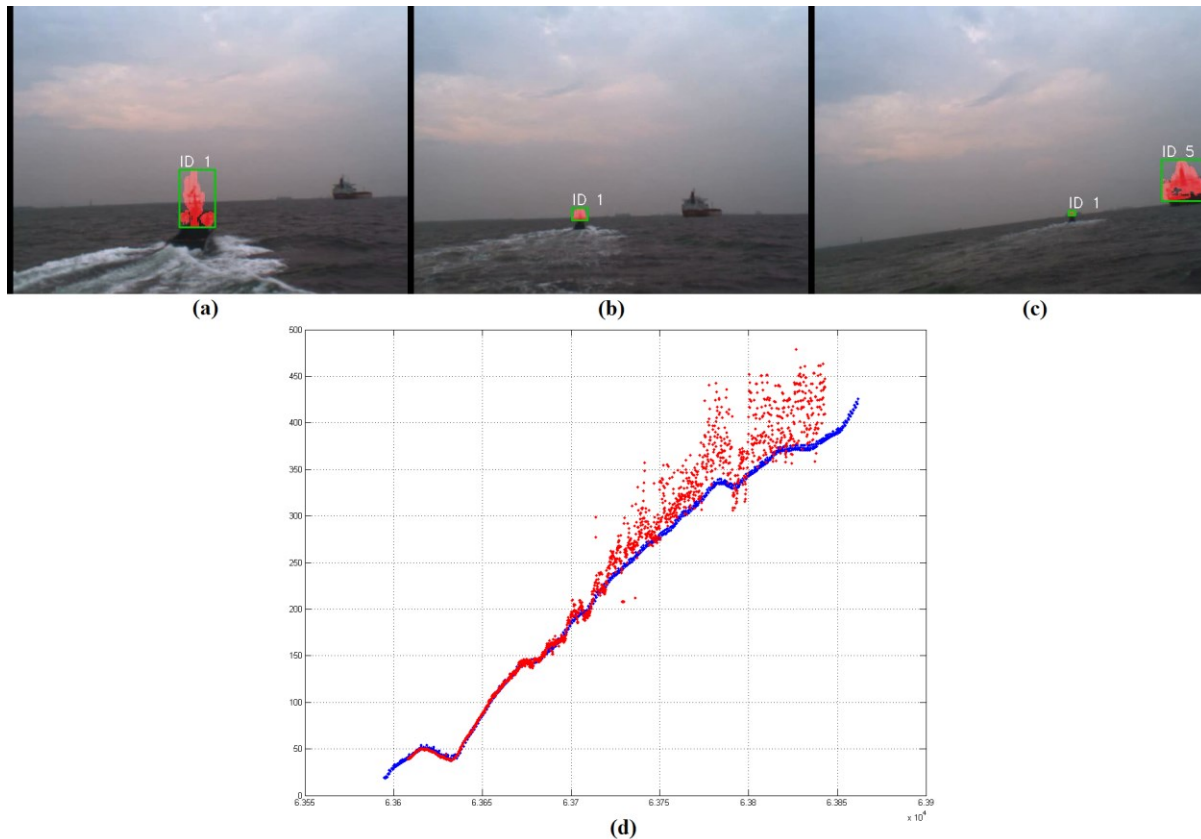


Fig. 5. (a) Boat target cruising in front of the USV in the close distance. (b) Boat target followed by the USV. (c) Boat target travelling ahead faster than the USV. Another ship in the valid range is detected as well. (d) Depth-time plot of the target following scene, measures by Differential GPS (blue) and the obstacle detection system (red).

To illustrate the accuracy of the depth estimation, the depth-time plot of a target following scenario (Figure 5(a)-(c)) is shown in Figure 5(d). The boat target was traveling ahead of the USV at a higher speed. It was stably detected and tracked. Another large ship was then detected and tracked when it was in the valid range of the system (Figure 5(c)). The estimated depth data (red) over time is compared to the ground truth GPS data (blue) as shown in Figure 5(d). The depth estimation is quite decent in the range from 40 to 200 meters, but it starts to deteriorate rapidly from 200 meters since the baseline of the stereo camera is about 1.7m.

IV. CONCLUSION

This paper describes a stereovision based obstacle detection system for USV. The system is capable of detecting, locating, and tracking multiple obstacles in real time, and provides accurate depth estimation in the range from 20 to 200 meters. Various situations can be handled, including partial occlusion of obstacles. We have plans to extend the field of view by adding another two sets of the stereo cameras. The future work also includes testing the system in more complex scenarios.

ACKNOWLEDGMENT

The authors would like to thank Chek Seng Ow, Kah Tong Ho, Benjamin Feng, and Junjie Huang for their help and support. This work was funded by Singapore Technologies

Electronics Ltd.

REFERENCES

- [1] J. Larson, M. Bruch, and J. Ebken, "Autonomous Navigation and Obstacle Avoidance for Unmanned Surface Vehicles," in *Proceedings of SPIE*, vol. 6230, 2006, pp. 17-20.
- [2] J. Larson, M. Bruch, R. Halterman, J. Rogers, and R. Webster, "Advances in Autonomous Obstacle Avoidance for Unmanned Surface Vehicles," in *Proceedings of the AUVSI Unmanned Systems North America*, 2007.
- [3] T. Huntsberger, H. Aghazarian, A. Howard, and D. Trotz, "Stereo Vision-based Navigation for Autonomous Surface Vessels," *Journal of Field Robotics*, vol. 28, no. 1, pp. 3-18, 2011.
- [4] H. Wang and Z. Wei, "A vision-based obstacle detection system for Unmanned Surface Vehicle," in *IEEE 5th International Conference on Robotics, Automation and Mechatronics*, Qingdao, China, September 2011.
- [5] K. Konolige, "Small Vision Systems: Hardware and Implementation," in *International Symposium of Robotics Research*, 1997.
- [6] R. Muñoz-Salinas, "A Bayesian plan-view map based approach for multiple-person detection and tracking," *Pattern Recognition*, 31(12):3665-3676, December 2008.
- [7] R. Hess and A. Fern, "Discriminatively trained particle filters for complex multi-object tracking," *Computer Vision and Pattern Recognition*, pp. 240-247, June 2009.
- [8] D. Low, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60(2):91-110, 2004.
- [9] P. Brasnett, L. Mihaylova, B. David, and N. Canagarajah, "Sequential Monte Carlo tracking by fusing multiple cues in video sequences," *Image and Vision Computing*, 25(8): 1217-1227, 2007.
- [10] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment," in *International Conference on Robotics and Automation*, 1991, pp. 1088-1095.