Multi-level On-board Data Fusion for 2D Safety Enhanced by 3D Perception for AGVs

Christian Stimming, Annette Krengel, Markus Boehning
Central Unit Research
SICK AG
Hamburg, Germany
{firstname.lastname}@sick.de

Andrei Vatavu, Szilárd Mandici, Sergiu Nedevschi
Computer Science Department
Technical University Cluj-Napoca
Cluj-Napoca, Romania
{firstname.lastname}@cs.utcluj.ro

***Abstract*—Modern AGVs are equipped with several safety laser scanners with a combined 360 deg field of view around the AGV to detect and subsequently avoid collisions with other AGVs, structural elements and, most importantly, workers. This contactless environment perception approach fulfils current safety legislation and safety regulations for driverless industrial trucks. However, obstacle detection is limited to a 2D plane parallel and close to the ground, unable to detect protruding or hanging objects in the path of the AGV.**

In order to avoid collisions with these kinds of objects as well, the idea of PAN-Robots is to enhance the existing 2D safety by a 3D perception system based on an omnidirectional stereo camera. This paper describes the multi-level on-board sensor data fusion strategies implemented in the PAN-Robots project. The fused information of tracked and classified objects is not only used for on-board risk assessment and emergency collision avoidance, but is also communicated to the global control center for advanced fleet coordination and intelligent AGV navigation.

Keywords*—*AGV, environment perception, laser scanner, data fusion, omnidirectional stereo vision, object tracking, object classification, 2D safety, 3D perception, low-level fusion, high-level fusion

I. Introduction

The European safety legislation and safety regulations for driverless industrial trucks [1] require, inter alia, a worker safety system that detects persons or objects in the path of the AGV. Early implementations used mechanical systems that act on physical contact, such as a plastic or foam bumper.

Today, contactless sensor technology based on laser, radar, infrared or ultrasonic that monitors the danger area around the AGV is the state of the art implementation for person protection. Fig. 1 shows safety laser scanners that are predominantly used for this purpose.

Fig. 1. Safety laser scanners for 2D AGV safety

Usually three or four safety laser scanners mounted at the lower perimeter of the AGV ensure a 360 degree field of view around the AGV, as illustrated in Fig. 2.

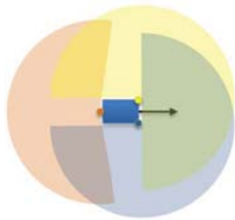

Fig. 2. Bird's eye perspective on an AGV (blue rectangle) and overlapping fields of view of three laser scanners on the perimeter of the AGV: front left and right with 270 degrees (yellow and blue, respectively) and rear with 190 degrees (orange)

However, obstacle detection is limited to a 2D plane parallel and close to the ground, unable to detect protruding or hanging objects in the path of the AGV. In order to avoid collisions with these kinds of objects as well, the idea of PAN-Robots is to enhance the existing 2D safety by a 3D perception system based on an omnidirectional stereo camera mounted on top of the AGV.

The system architecture and the steps of on-board data fusion are illustrated in Fig. 3 and are described in the subsequent sections.

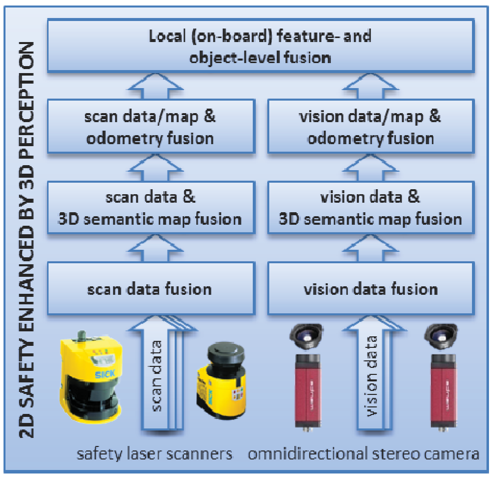

Fig. 3. System architecture

978-1-4673-8200-7/15/$31.00 ©2015 IEEE

II. 2D Safety Laser Scanners

A. Scan data fusion

The laser scanners gather a range profile ('scan') of the environment. As the four scanners have different fields of view into the four directions of the AGV, the complete environment can be observed by fusing all four range profiles of the scanners. Some of the scanners' fields of view are overlapping, but only in part, see Fig. 4. By taking into account the calibrated mounting positions and fields of view of each of the scanners, the scans from the individual scanners can be registered into one common coordinate system around the AGV.

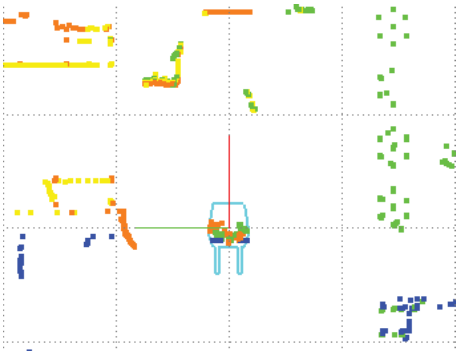

Fig. 4. Scans from four on-board safety laser scanners (colored by device)

The safety scanners that are being used in the PAN-Robots system have no external time synchronization of their laser scanner measurements. Instead, each scanner's timing behavior is controlled by its own single free-running internal clock. Hence, the actual measurement timestamps and the exact duration of one scan are slightly different between the scanners and also over time.

Additionally, the used safety laser scanners (SICK S300 and S3000) have different scan frequencies, with the S3000 having shorter scan durations (30ms) than the S300 (40ms). For this reason, the scans with the shorter interval are filtered in time to use only the closest scan from the S3000 for fusion with the S300 scans. The synchronization is performed in the sensor data processing by interpolating the different scans on one common measurement timestamp. As a result, all scan points from the environment can subsequently be treated as if they had been recorded by one single laser scanner, see Fig. 5.

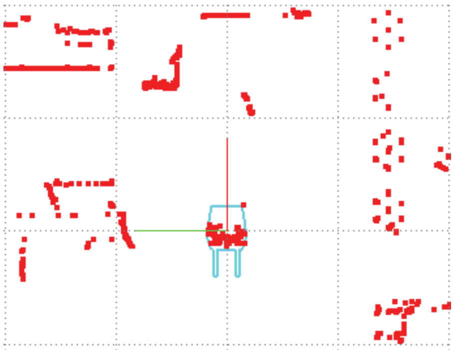

Fig. 5. Fused scans

B. Integration of semantic map information

The semantic map of the warehouse environment is used to identify and extract regions of interest for object detection. The background elimination module is based on a grid map which is obtained from the semantic map. The grid map of the warehouse is shifted and rotated according to the current position and motion of the AGV. This is an efficient use of the information about the AGV odometry which is being determined by the position estimation application anyway. If there is a potential delay in the determined position's timestamp as compared to the current scan timestamp, the position is extrapolated to the current scan timestamp.

The following Fig. 6 exemplifies the background elimination process, where the foreground (red) and background (yellow) scan points are shown, based on the warehouse grid map (greyscale).

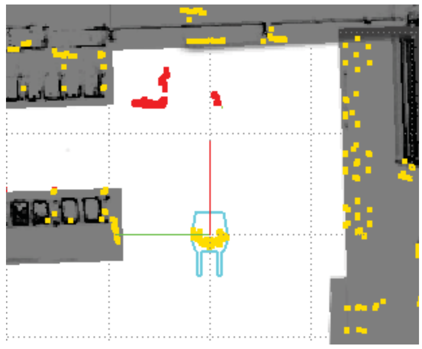

Fig. 6. Background elimination using 3D semantic map

C. Integration of AGV odometry data

Due to the motion of the AGV, the observed relative positions of other objects in the environment will change. This is taken into account by the ego-motion compensation module, where the known AGV ego motion from the odometry information is used to change all relative object positions according to the motion in the time interval between the previous scan's timestamp and the current scan's timestamp. This change in relative position is done for all currently tracked objects in the Kalman filter.

D. Tracking

The remaining scan points that are relevant for the tracking measurements will be clustered into segments according to their euclidian distance, see Fig. 7.

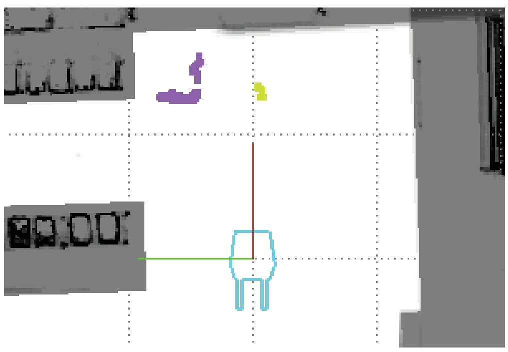

Fig. 7. Remaining scan points as segments for tracking

The list of segments resulting from the object segmentation task is the input data to the Kalman tracking algorithm. The list of segments is compared to the already tracked objects in an association step based on Euclidean distance of the centers of gravity or distinctive contour points [2]. Associated objects will get their state vectors updated according to the current segment measurements. For segments without associated objects, a new object track will be initialized. And for objects without observed segments, the object tracks will be deleted from the tracking after some suitable waiting time.

Tracked objects are passed onto the object classification module for object output, and will also be predicted for the future time steps. Using a constant velocity motion model, tracked objects are predicted for the subsequent association step. Each object has a (locally) unique identifier (ID) and parameters such as position, size, and velocity of each object are tracked.

Fig. 8 shows the end result of detected, tracked and classified objects in the foreground area of the 2D safety laser scanner perception system.

Fig. 8. Tracked Objects (a pedestrian and two AGVs). Video is for reference only.

III. OMNIDIRECTIONAL STEREO CAMERA

Stereovision based environment perception has been used successfully in various applications and systems but most of them are related to advanced driving assistance [13]. Within the PAN-Robots project a solution is proposed for representing the surrounding environment using a stereovision based system. Because of the mounting position and the usage of omnidirectional fisheye lenses [8], a 360 degree field of view is achieved around the AGV. The main advantage of this sensor is that it provides a large quantity of data that can be interpreted by taking into account the sensor model. It provides information in 3D as opposed to the 2D information delivered by laser scanners.

The stereovision system consists of 2 cameras, 2 fisheye lenses and an ECU (Electronic Control Unit) which performs the actual processing of the fisheye images. A preliminary version of the system was presented in [6] focusing on the initial parts of the processing: multi-channel rectification using the construction of virtual imagers. The outline of the set of modules and tasks performed within the omnidirectional stereovision based environment perception system is described in Fig. 9.

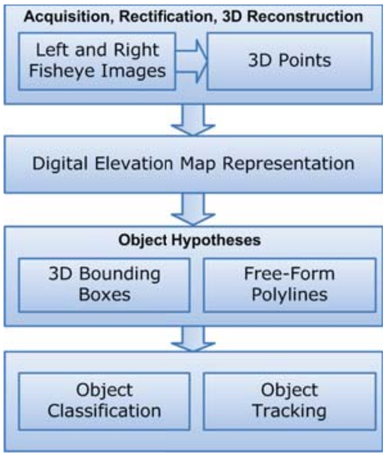

Fig. 9. Stereo-vision based environment perception

In order to address the real-time requirements of the system multiple modules have been implemented in CUDA on the GPU. Additionally, these modules are organized in a way that facilitates the implementation of a GPU/CPU pipelined execution.

A. *Dense Stereo Reconstruction*

An important part of the system is the dense stereo reconstruction implemented on the GPU. The objective of this module is to calculate the 3D points based on each rectified image pairs. The 3D scene depicted by the input images is reconstructed by estimating the depth from the disparity values for each pixel. This process results in the 3D points which are connected based on proximity and colored to obtain the rendering of the scene (Fig. 10). The implemented algorithm is a modified version of semi-global matching (SGM) with subpixel accuracy [9, 10]. system.

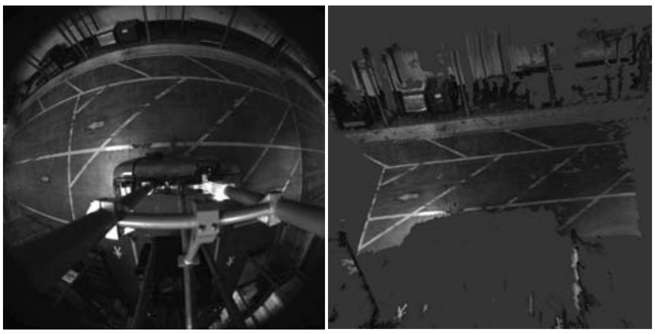

Fig. 10. Fisheye image (left) and 3D reconstruction (right)

B. Digital Elevation Map representation

Based on the 3D points obtained from the dense stereo reconstruction an intermediate representation in the form of a Digital Elevation Map (DEM) is used. The DEM is constructed probabilistically [7], taking into account the direct and inverse sensor models of the omnidirectional stereovision sensor calculated experimentally.

In the classical version each DEM cell contains a single height corresponding to the maximum value. This is an essential choice which is necessary to address the problem of bimodal cells. In order to be able to detect and represent hanging and protruding objects, each DEM cell contains 2 different heights: a minimum and a maximum height. For the lower height the smallest value above the ground plane which has a consistent weight is considered. Using this representation 2 vertical height profiles are obtained.

Additionally, each elevation map cell calculates the mean and the standard deviation of the intensity value corresponding to that cell. After obtaining the height and intensity values the gradient is also calculated in each cell. By considering these new aspects a more accurate DEM cell grouping mechanism can be implemented for detecting objects. system.

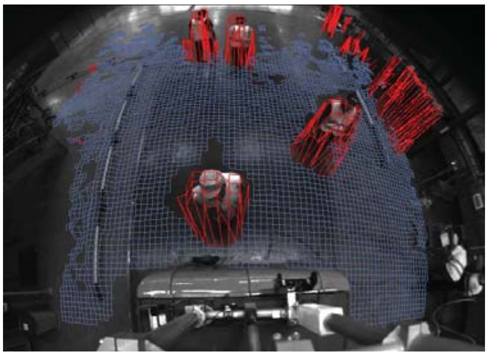

Fig. 11. The obtained probabilistic inverse-sensor model based DEM

C. Ground Plane Temporal Filtering

To increase the robustness and accuracy of the elevation map cell classification, a temporal filtering is applied to the detected ground plane. As usually the road profile in the warehouse environment is considered to be flat, it can be represented by the following general equation:

$$Ax + By + Cz - D = 0 \tag{1}$$

The four parameters A, B, C and D are updated in time by using a standard Kalman filter mechanism. system.

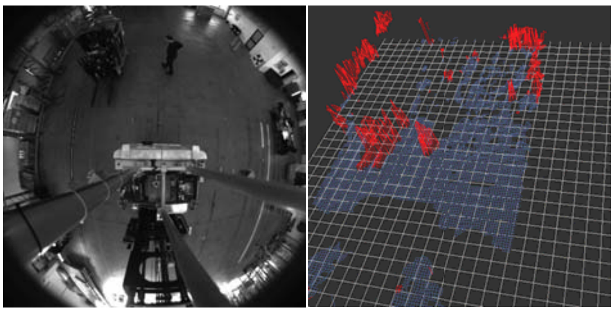

Fig. 12. The estimated ground plane (right) for agiven scenario (left).

D. Object Detection

The elevation map measurements are used to extract a set of obstacle hypotheses. First, the DEM cells that are labeled as "obstacle" are grouped into connected entities (blobs). Then, for each connected group of DEM cells, two different models are determined: a 3D cuboidal model that is used to define the region of interest for the object classification step, and a free-form polygonal model used to estimate the object motion [11].

One of the main advantages of the vision-based object detection solution is the ability to detect hanging objects (see Fig 16. a). The basic idea consists in determining, for each DEM cell, two different height values: the minimum height of the object that does not belong to the ground plane and the maximum height of 3D points that are projected in the same cell. These height values are used in the grouping stage to define the lower and upper object boundaries.

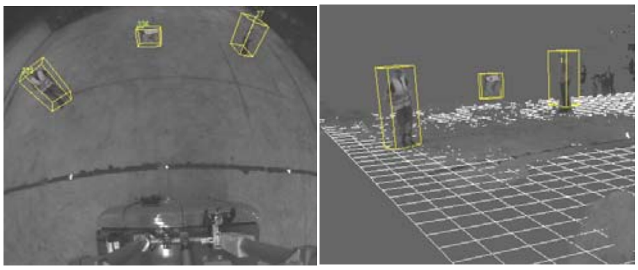

Fig. 13. Detecting hanging objects. A scenario containing two pedestrians and a hanging box. Left: fisheye image. Right: 3D view.

E. Object Tracking

The proposed tracking solution relies on two separate steps: *motion estimation* and *filtering*. In order to estimate the obstacle motion we use an Iterative Closest Point based method, which aims to estimate an optimal rotation and translation between the associated polygonal models that describe the same target in different frames [12]. The resulted transformation parameters are subjected to a Kalman filtering technique.

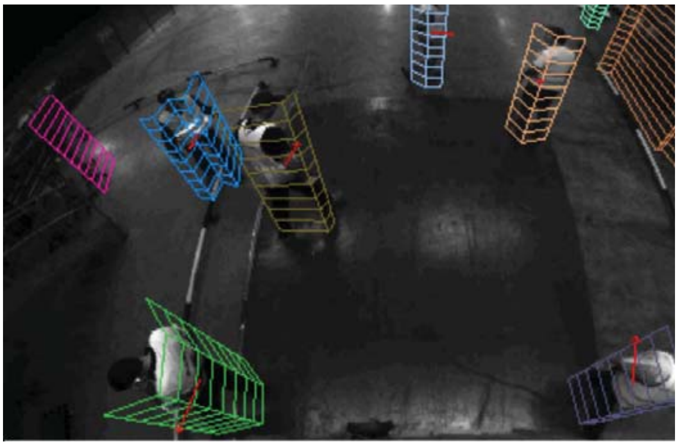

Fig. 14. Object Tracking.

F. Object Classification

Having a set of obstacle hypotheses represented as 3D oriented boxes, each of them is classified as "AGV", "Pedestrian" or "Other obstacle".

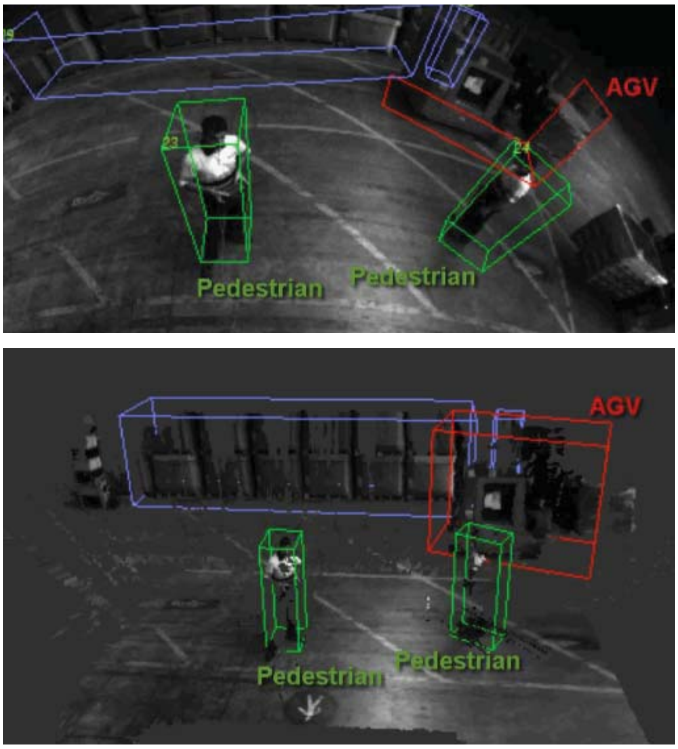

Fig. 15. Object classification

For object classification we use Ada-boost [4] where each obstacle is classified based on the visual codebook based features [5].

IV. Local On-board Object-level Fusion

The data resulting from the two threads of 2D perception and 3D omnidirectional vision are fused on-board the AGV to enhance object detection, tracking and classification performance. By fusing all these data, a more precise representation of the environment is possible. All the sensors have advantages and disadvantages and by taking into account these facts a truly authentic view of the surroundings of the AGV can be acquired.

One important characteristic that needs to be taken into account when performing sensor data fusion is the fact that usually the laser scanners and the static 3D digital maps offer a better accuracy and detection stability when compared to stereo reconstruction. On the other hand, the data obtained from the stereo sensor is much richer and represents objects in 3D in 360 degrees around the AGV, while the laser scanners provide only 2D information.

Apart from these sensor specific characteristics, the sensors and the data need to be synchronized as well. The computation units of 2D perception and 3D vision have different clocks. In order to merge the lists of tracked objects provided by each device, a global time reference is necessary and each data source must provide a precise timestamp. This is achieved by network time protocol (NTP) synchronization where both the 2D perception system and the 3D vision system synchronize with the reference clock of the PAN-Robots global control center.

Both sensors need to be registered to a common reference frame. In our case this was performed by perceiving a reference object, at the same time, with both 3D vision and laser sensors and then estimating the optimal transformation parameters between the two coordinate systems by using a least squares technique.

A. Probabilistic object-level fusion

In modern data fusion applications probabilistic methods are considered to be the standard approaches [3]. Probabilistic fusion expresses the data uncertainty using a probability distribution for each information source. In the PAN-Robots system, each device provides an object list containing the position, size, orientation, velocity, and classification of each detected object along with the uncertainty of these measurements.

The fusion algorithm gathers information about objects in the surrounding of the AGV from the different devices and incorporates all available information which has been accumulated up to that point in time. It keeps track of a fused object list in order to provide as much information as possible about the surrounding at the current time.

Whenever a new object list arrives from either 2D perception or 3D vision, the fused object list is updated according to the newly gathered measurements. The fusion process itself then consists of two major phases:

1) Object association phase

For each detected object, it is checked if it can be associated to an object in the fused object list. The fusion module incorporates the tracking information from each individual device, which means that if an object has been detected by the same device before this association is preserved. Afterwards, remaining objects are associated to the closest fused object within a certain range or, if such a fused object is not found, induce the creation of a new fused object. For each fused object the algorithm keeps track of its detecting devices and deletes the fused object if it has not been detected in the most recent object list of at least one device.

2) Object update phase

For each object in the fused object list, a Kalman filter based prediction is applied to object center, object box size, heading, and object velocity. Afterwards, if there is an association to a current measurement, the respective Kalman states are corrected accordingly. Classification fusion is based on the classification uncertainty which is provided by each device and indicates the reliability of the assigned class label.

As the laser scanners do not provide height information, fusion is performed based on 2D data and the height information obtained from stereo reconstruction is assigned to the respective fused objects.

Due to differing frame rates and processing times in the different modules, the object lists do not necessarily arrive at the fusion module in chronological order. Specifially, the 3D vision output can be delayed with regard to the 2D perception. In order to cope with this situation, the fusion algorithm keeps track of the history of the fused object list and, if necessary, goes back in time, incorporates the new measurements, and afterwards updates the list according to the more recent measurements from 2D perception again.

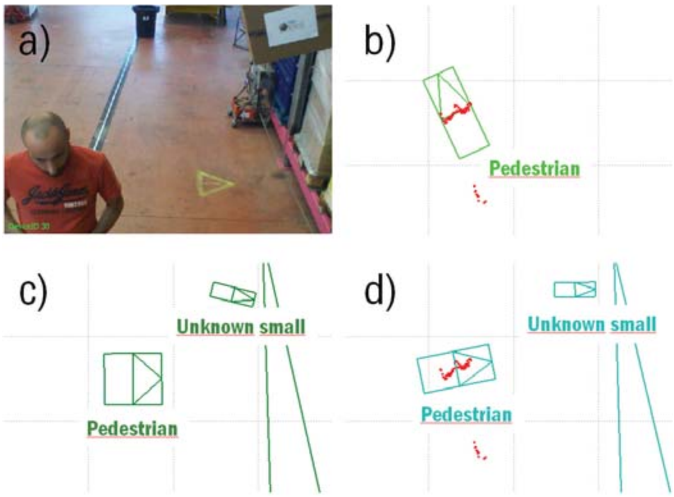

Fig. 16. On-board object fusion for a scene with a protruding object. a) Image of the scene with hanging object (upper right corner), b) Objects detected by 2D perception, c) Objects detected by 3D vision, d) Fused objects.

Finally, the fused object list is used in the PAN-Robots local on-board path planning module for obstacle cirumnavigation. Furthermore, the consolidated object list is communicated to the PAN-Robots control center, which fuses the objects lists of all PAN-Robots AGVs and infrastructure systems into a coherent global live view of the warehouse. Ultimately, this allows an efficient and flexible AGV fleet management.

V. Summary and Conclusions

In mixed traffic warehouses, where AGVs, manual forklifts and workers share the working space, on-board safety laser scanners efficiently ensure worker safety.

However, there are situations where structural elements in the path of the AGV can lead to accidents and material damage because the obstacles, such as hanging or protruding objects, are not detectable by the 2D safety laser scanners. These situations can be covered by enhancing the 2D safety concept by 3D perception using an omnidirectional stereo camera. This way, the accuracy and robustness of the first technology is combined with the higher classification accuracy in a much larger observation area of the second.

Smart fusion of the tracked and classified objects of both complementary environment perception systems as well as the additional available information, such as the 3D map of static objects in the warehouse, leads to a highly robust and reliable detection of obstacles. This supports the "vision zero", a vision of zero AGV related accidents in modern warehouses.

Acknowledgments

PAN-Robots has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 314193.

References

[1] DIN EN 1525:1997-12 "Safety of industrial trucks – Driverless trucks and their systems", 1997.

[2] Christian Stimming, Florian Ahlers, "Object Contour Tracking in Vehicle and Infrastructure Laserscanner Systems", Workshop on Intelligent Transportation (WIT), 2010.

[3] Bahador Khaleghi, Alaa Khamis, Fakhreddine O. Karray, Saiedeh N. Razavi, Multisensor data fusion: A review of the state-of-the-art, Information Fusion, v.14 n.1, p.28-44, January, 2013.

[4] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," The Annals of Statistics, vol. 38, no. 2, pp. 337–374, 2000.

[5] A. Costea, S. Nedevschi, Multi-class segmentation for traffic scenarios at over 50 FPS, Proceedings of 2014 IV Symposium, Dearborn, USA, 2014, pp. 1390-1395.

[6] M. Drulea, I. Szakats, A. Vatavu, S. Nedevschi, "Omnidirectional stereo vision using fisheye lenses", Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference, Sept. 2014.

[7] S. Mandici, S. Nedevschi : "Aggregate Road Surface based Environment Representation using Digital Elevation Maps" Intelligent Computer Communication and Processing (ICCP), 2014 IEEE International Conference, Sept. 2014.

[8] A. Mäyrä, M. Aikio, M. Kumpulainen, "Fisheye optics for omnidirectional perception", Intelligent Computer Communication and Processing (ICCP), Sept. 2014.

[9] I. Haller, S. Nedevschi, "GPU optimization of the SGM stereo algorithm", 2010 IEEE In-ternational Conference on Intelligent Computer Communication and Processing (ICCP), pp.197, 202, 26-28 Aug. 2010.

[10] H. Hirschmuller. "Accurate and efficient stereo processing by semi-global matching and mutual information", Proceedings of CVRP, pages 807-814, 2005.

[11] A. D. Costea, A. Vatavu and Sergiu Nedevschi, "Obstacle Localization and Recognition for Autonomous Forklifts using Omnidirectional Stereovision", to be published at 2015 IEEE Intelligent Vehicles Symposium, June 28 – July 1, 2015.

[12] A. Vatavu, S. Nedevschi, "Real-time modeling of dynamic environments in traffic scenarios using a stereo-vision system", 2012 15th International IEEE Conference on Intelligent Transportation Systems (ITSC 2012), pp.722-727, Sept. 16-19, Anchorage, Alaska, USA, 2012.

[13] S. Sivaraman, M. M. Trivedi, "Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis," IEEE Transactions on Intelligent Transportation Systems,vol.14, no.4, pp.1773-1795, Dec. 2013.