FUSION OF MULTI EXPOSURE STEREO IMAGES AND THERMOGRAPHY FOR OBSTACLE DETECTION ON AGRICULTURAL VEHICLES

M. Nielsen and Kurt Nielsen

Danish Technological Institute
Forskerparken 10F, 5230 Odense, Denmark

ABSTRACT

Autoguided vehicles have been successfully installed in indoor conditions such as hospitals and warehouses. The success can be contributed to the structured environment and LIDAR sensors stopping the vehicle for anything that it not the flat ground. This will not work in an agricultural context, where the vehicles have to move through tall crops. A fusion of sensors with different strengths and scales has to be used. The drawback of a classifier that merges all sensors to give results is timing, whereas a subset of the sensor suite may be given their own rights to make conclusions. This paper investigates a fusion of HDR stereo and thermography that can further be fused with other sensors but is powerful on its own. It is assumed that traversable paths are known and obstacles can be anything unpredictable. Consequently, a scheme that trains traversable paths, while everything else are obstacles is proposed. This way tall maize crops can be trained as traversable. It was found that water is difficult to handle with these sensors. It was often seen the same way as an abrupt decline that would then urge the vehicle to stop, even if it was just a large puddle.

Keywords: Autonomous Vehicles, Sensor fusion, thermography, stereo vision.

INTRODUCTION

Over the years agricultural vehicles become increasingly automated with trajectory row tracking and master-slave vehicle configurations, and autoguided vehicles. Safety is an important aspect. Auto guided vehicles exist in industry, where the surroundings are semistructured and flat. Some cars have collision sensors. But in agriculture the ground is not flat. The vehicles are meant to be driven into crops, and there are certain paths and crops that can be driven in while others cannot. Obstacles include steep hills, lakes, wells and large stones, ditches, wildlife, vehicles and humans. Research has often been done using LIDAR and stereo vision (Kuthirummal et al, 2011, Rouvere et al 2012, Reina et al 2013). We suggest a supplement sensor to the fast reacting LIDAR that fuses of high dynamic range stereo and thermography, which will form a feature vector that can classify traversable and non-traversable paths.

Table 1: Strengths and weaknesses between different sensor modalities.

Sensor	Pros	Cons	Visibility
Stereo	Colour pointclouds or keep the image neighborhood Texture visual classification Can model the terrain, see ditches, holes, positive obstacles incl living positive obstacles	High processing load Occlusion Difficult to see lakes and living things in vegetation. Needs visibility. Short narrow range.	SNR and range is affected by visible particles in the air (fog, dust, smoke, rain, snow)
Thermography	Vehicles and living objects stand out Structures, water, poles also integrate a lot of heat. Works in pixel neighborhood. Response says something about materials.	No 3D representation. Increasing noise that needs to be recalibrated every few minutes which takes a few seconds. Temperatures depend a lot on the weather so everything must be adaptive. Low DR.	Smoke does not affect it. Range is affected in cat 3 fog, but much less than visual cameras for cat1-2 fog and rain.
LIDAR	Easy to setup and get data. low noise point clouds 360 degrees full 3D	Only point clouds Slow rotation at the moment Affected by visibility issues	Affected by rain and fog (less than cameras) but not in smoke
RADAR	Easy setup SLAM maps without GPS 360 degrees, 100 meters radius Can see lakes Unaffected by low visibility environments	Low resolution Only 2D maps	Not affected

METHODS

A Claas Axion was mounted two Basler ace 640-90gc colour cameras in a stereo setup with 80cm baseline and a Flir a615 thermal camera. They were hardware synchronized and took three exposures for high dynamic range images. It took only 40ms to acquire each image set. The Flir camera did not have external trigger but with an internal clock at 50Hz (20ms) the software synchronization was accurate to acquire a thermal image within the first two exposures of the stereo cameras. Calibration was done using an A0 sized checkerboard. The thermal camera saw it inverted when doing the acquisition outside where heat from the sky was integrated in the carbon contained in the black print.

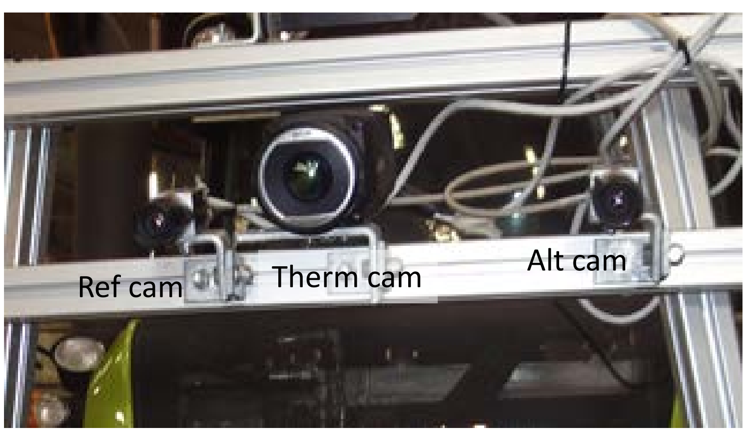

Figure 1. Two Basler 640-90gc cameras and Flir A615 mounted in front of the tractor.

Data was acquired for 3 days and a subset containing 2 fields, a yard surrounded with buildings, a public road and gravel roads was selected for analysis. Obstacles included lakes, hedges, stones, vehicles, fences, ramps, concrete barriers, buildings, people hiding in maize, trees, a dog and game birds.

Figure 2. [left] One exposure [right] Software HDR

The basis for outdoor stereo was (Debevec and Malik, 1997) imaging with the fast Basler cameras. They took 3 exposures in 40ms which was fast enough for moving targets. Tearing was visible, but the cameras were hardware synchronized, which meant that both cameras had the same artifacts, so stereo matching was possible. 3D reconstruction was done real-time using cuda acceleration and the thermal image was warped to corresponding color pixels. Both classic correlations based algorithm (Fusiello et al 2000) and two-four pass dynamic programming (Kim et al 2005) were implemented in CUDA. See figure 3.

Figure 3. from left to right: 1. Reference HDR image. 2. Depth map from dynamic programming. 3. Plane fit and rotated to ground plane to show height from ground and colour coded for red obstacles. 4. Height from ground without the person, based on SMW subpixel correlation depth map.

As (Kuthirummal 2011) suggested, the pixels were partitioned into 600x600mm cells based on the 3D data. A vector of chromaticity, height from a reference plane and temperature was formed for each cell.

Figures 5-6 show an overview of the classification scheme.

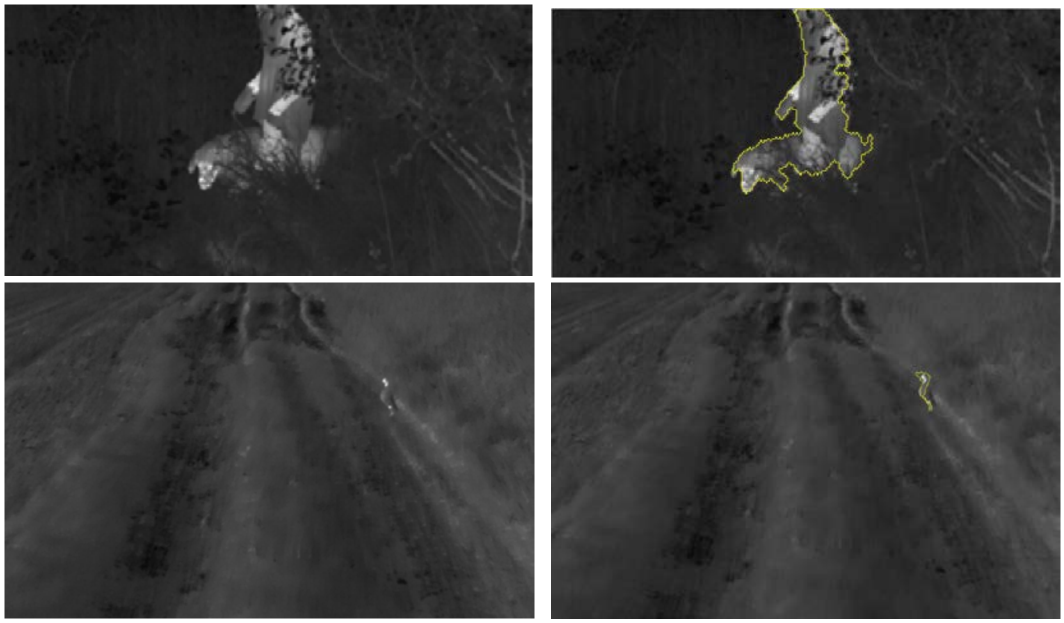

Figure 4. Thermal Camera. Segmentation done by region growing. A dog with girl in the bushes and a bird near the maize crop line is detected.

Classification of each cell was based on the distribution of these vectors within a cell. The distribution was tested in two ways: 1. Accumulative Histogram Carving. 2. Expectation Maximization fitting with Gaussian Mixture Models. The score was defined as the intersection of a cell histogram and the trained histogram (1), and the Bhattacharyya distance of GMMs (2). It is important to note that using E-M GMM for classifying the distribution of features is different from using normal GMM classification which only classifies features.

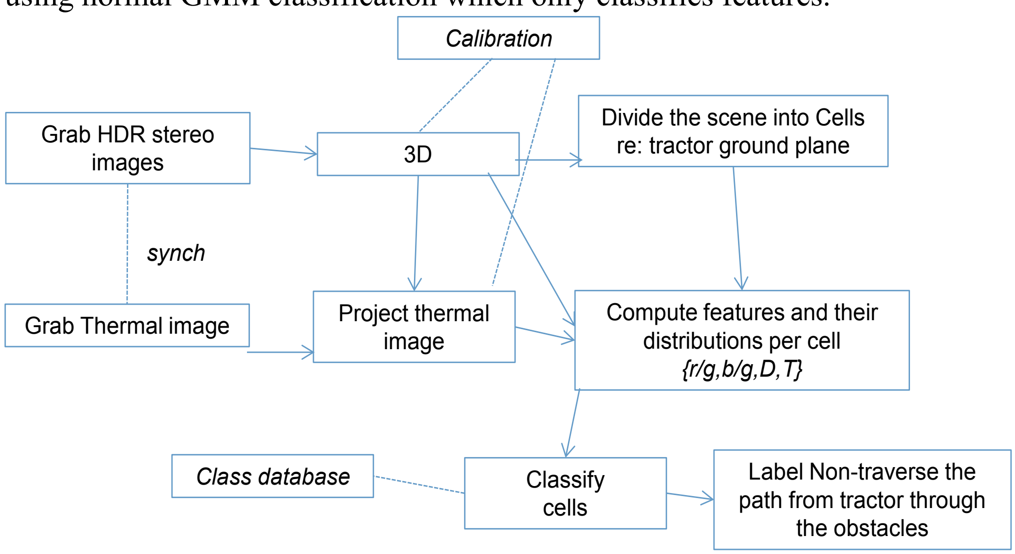

Figure 5. Fusion between thermography and stereo. Stereo cameras are calibrated intrinsically and extrinsically. The thermal camera is calibrated extrinsically in relation to the reference camera. Epipolar geometry allows easy computation of the displacement of pixels in the thermal image so the thermal iamge can be

warped to overlay the reference image. Intrinsic distortion was ignore because the lens was narrow angle and has next to no distortion.

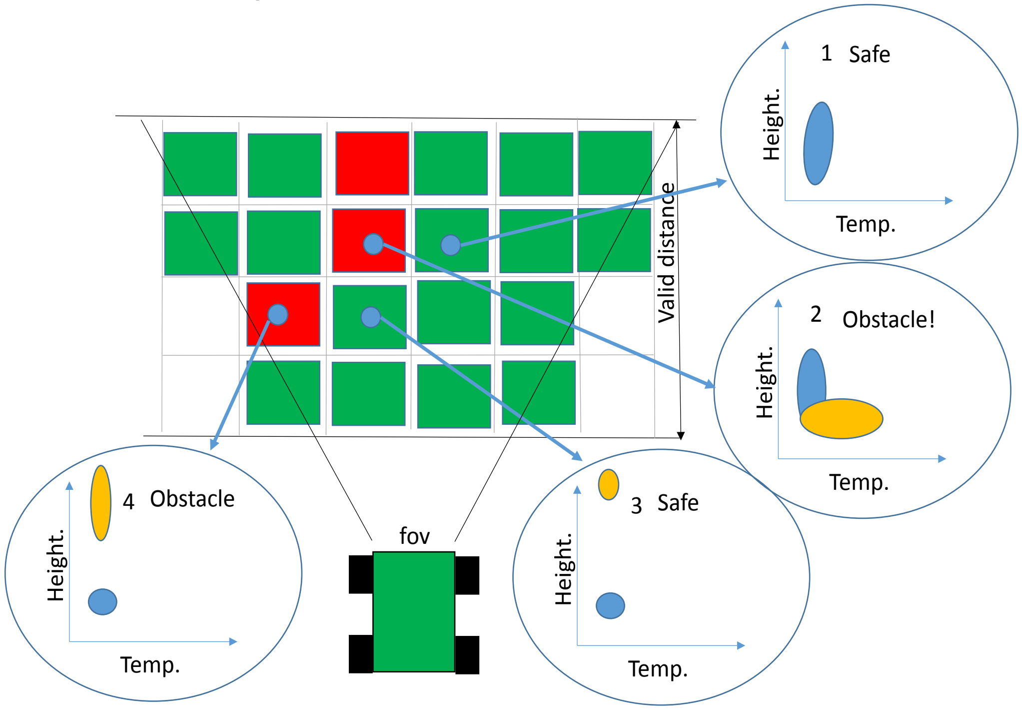

Figure 6. The FOV in front of the tractor was divided into a grid and the classification was based on the "distribution of feature values", rather than absolute values. Cells behind an obstacle was marked as not traversal, too. Four examples of cell distribution are shown in 2D plots for Height and Temperature: 1. Typical tall crop situation. Many heights exist. It is safe. 2. In the same tall crops, there is now a small animal. STOP! 3. Flat ground and high hanging branch above the tractor. It is safe. 4. Now the branches are too low. STOP!

RESULTS

The acquisition and synchronization was fast enough that cars driving 80-100 km/h while the tractor was waiting to cross the road showed up on all 3 cameras at matching pixels. Figure 7 shows the tractor approaching a main road with a car passing at 80km/h. The stones on each side of the entrance were classified as obstacles while the slight inclination to the road was seen as traversable. Suddenly, a car passed, and it was seen as an obstacle in 1 frame. The position of the car matches in the stereo images and thermal image. As the tractor entered the main road, a gravel side road appeared as traversable.

Figure 7. Results shown in the following way: RGB Image / Classified Cells / Thermal Image that has a smaller FOV that limited amount of annotated cells. Green cells: Traversable. Red: Not traversable. Purple/Cyan: Selected cell for showing the GMM classes (when applicable).

Figure 8 shows examples of a farm courtyard. A distant barn wall was detected, as well as people chatting in front of it. A ramp was seen as traversable while the vertical walls around it are not. As the tractor drive up the ramp, the door that appear was seen as an obstacle.

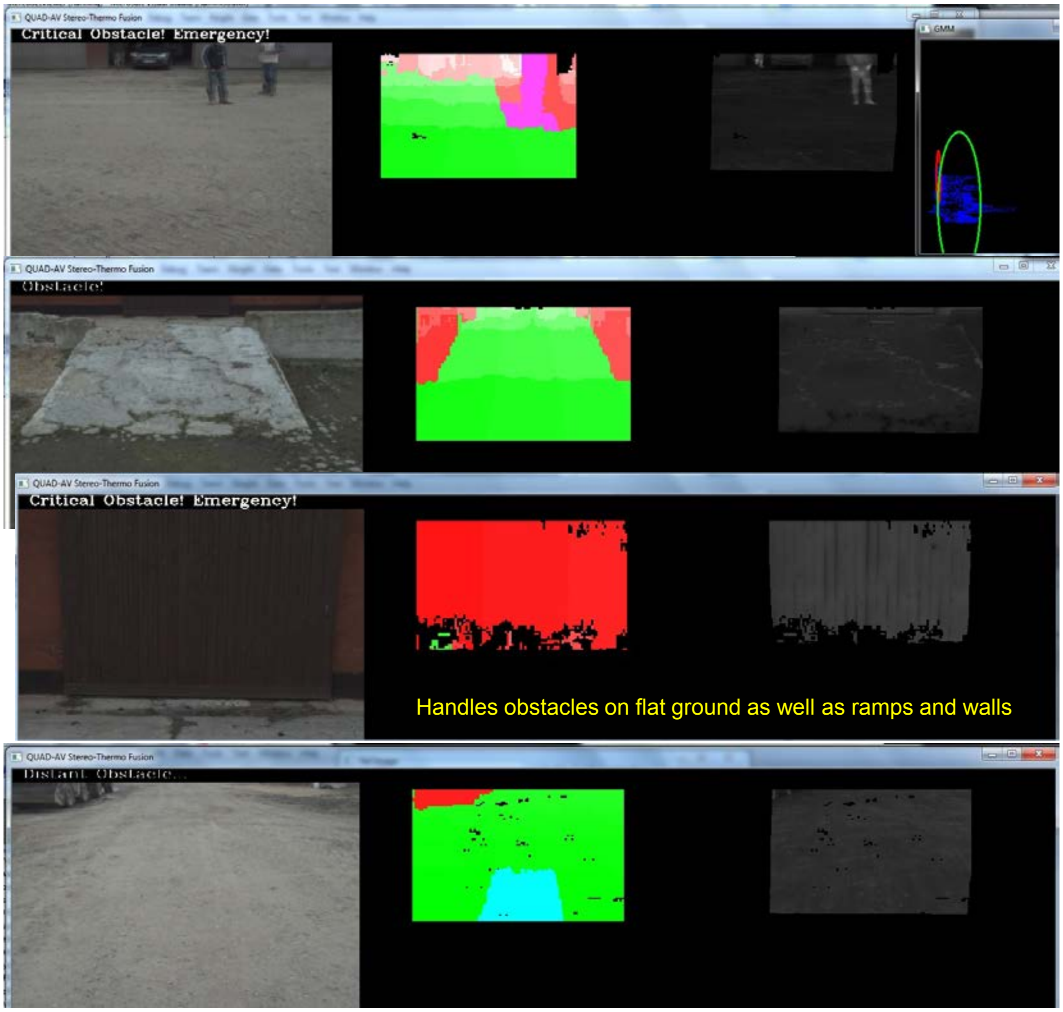

Figure 8. various situations of a farm courtyard. People, vehicles, ramps and doors.

Figure 9 shows the most interesting cases, where a tall crop can be classified a traversable. In the top most row, the left cell is selected (cyan = Traversable + selected) and it is traversable. To the right the GMM space is shown. As tall vertical ellipsoids. The middle cell is red (obstacle). 2nd row selects that cell. Now the GMM shows 2 unique components. The RGB image shows maize. But the thermal image shows a person behind the maize. 3rd and 4th row shows tall grass that's traversable, but then a person shows up and it is an obstacle. The last row shows a person crouching inside the maize and that cell is highlighted.

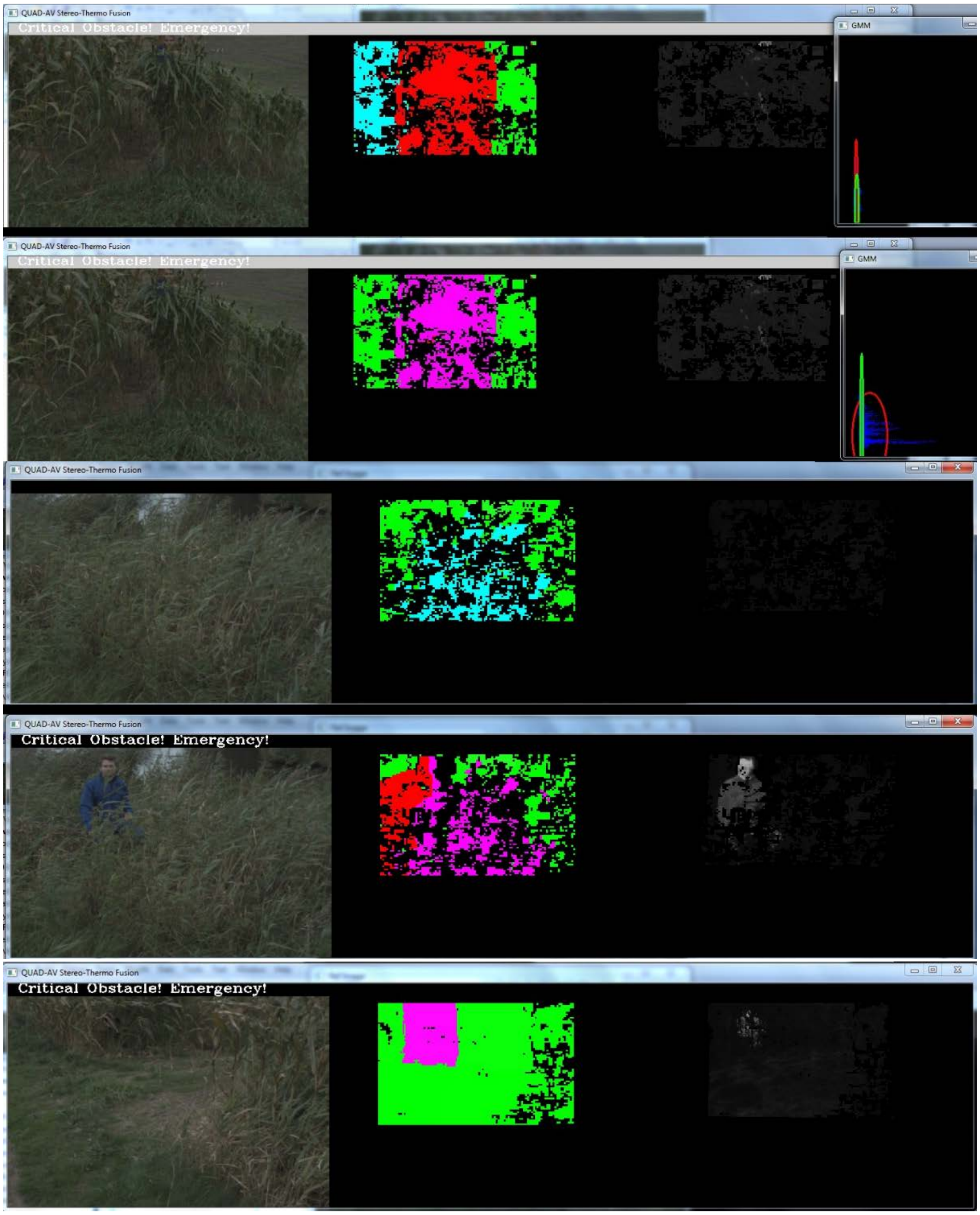

Figure 9. Can learn that the tall crops are ok to drive through, except if a living obstacle is inside

The histogram approach was very fast, but it required more training than the GMM approach because it does not generalize the data well. Execution speed did not increase with training, because it was carving out distributions that were considered traversable in the same data space. GMM classification was easier to train but was slower in runtime. The more classes trained, the slower it became, but fewer training images were needed.

Results showed that the significant features were Height and Temperature. Obstacles were detected well with chromaticity weights zeroed and equal weight on height and temperature. The temperature was an important feature for detecting living things and vehicles, but also stones and poles. It made it easy to detect people hiding 5 meters inside tall maize, and crouching in tall grass. The height made it possible to see fences, half meter tall concrete barriers versus valid ramps. The depth resolution was limited at a distance, so small fences and hedges were not detected until within 15 meters.

Lakes and rain pools were tricky to interpret. The water mirror would trick the stereo system into thinking it was farther away. The temperature signal was often colder then surroundings, but it may also reflect the temperature of a building. In cases where a lake was deep compared to the ground around it, the system saw it as an obstacle. Fusion with radar would improve the interpretation of these.

CONCLUSION

Fusion of stereo and thermography was shown to be a powerful obstacle detector in agricultural settings, where the traversable ground cannot be assumed to be flat. Training can be done by driving the vehicle through traversable paths and crops. However, a vehicle can be pre-initialized with certain knowledge about certain obstacles. Certain classes can also be assigned higher priorities, such as humans.

The system can be used for auto-guided vehicles as well as alert unattentive drivers. The approach can be expanded with other spectral bands (NIR, UV) and waves such as ultrasound, and act complementary to or fused with LIDAR and RADAR systems.

ACKNOWLEDGEMENTS

This project was funded under Era-Net ICT-AGRI framework. Conference participance funded by Robocluster.

REFERENCES

Debevec, P.E. and Malik, J. 1997. Recovering High Dynamic Range Radiance Maps from Photographs. In SIGGRAPH 97, August.

Fusiello, A., Roberto, V., and Trucco, E. 2000 Symmetric stereo with multiple windowing. International Journal of Pattern Recognition and Artificial Intelligence, Vol. 14.

Kim, J.C, Lee, K.M., Choi, B.T., and Lee, S.U. 2005 A dense stereo matching using two-pass dynamic programming with generalized ground control points.

IEEE Computer Society Conference on Computer Vision and Pattern Recognition,. CVPR, Volume 2.
Kuthirummal, S., Das, A., Samarasekera, S. (2011). A graph traversal based algorithm for obstacle detection using lidar or stereo. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1579-1584.
Nielsen, M., Slaughter, D.C., Gliever, C. 2012. Vision-based 3D peach tree reconstruction for automated blossom thinning. IEEE Transactions on Industrial Informatics, 8 (1) pp. 188-196.
Reina, G., Milella, A., Halft, W.,Worst, R. 2013 LIDAR and stereo imagery integration for safe navigation in outdoor settings. IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)
Rouveure, R., Nielsen, M. Petersen, A., Reina, G. Foglia, M.M., Worst, R., Seyed-sadri, S., Blas, M.R., Faure, P., and Lykkegård, K. 2012. The QUAD-AV Project: Multi-Sensory Approach For Obstacle Detection In Agricultural Autonoumous Robotics. In Proceedings International Conference on Agricultural Engineering, Valencia.