

# Platform for evaluating sensors and human detection in autonomous mowing operations

P. Christiansen<sup>1</sup>  · M. Kragh<sup>1</sup> · K. A. Steen<sup>1</sup> ·  
H. Karstoft<sup>1</sup> · R. N. Jørgensen<sup>1</sup>

Published online: 13 January 2017  
© Springer Science+Business Media New York 2017

**Abstract** The concept of autonomous farming concerns automatic agricultural machines operating safely and efficiently without human intervention. In order to ensure safe autonomous operation, real-time risk detection and avoidance must be undertaken. This paper presents a flexible vehicle-mounted sensor system for recording positional and imaging data with a total of six sensors, and a full procedure for calibrating and registering all sensors. Authentic data were recorded for a case study on grass-harvesting and human safety. The paper incorporates parts of ISO 18497 (an emerging standard for safety of highly automated machinery in agriculture) related to human detection and safety. The case study investigates four different sensing technologies and is intended as a dataset to validate human safety or a human detection system in grass-harvesting. The study presents common algorithms that are able to detect humans, but struggle to handle lying or occluded humans in high grass.

**Keywords** Safe farming · Sensor platform · Object detection · Computer vision · ISO 18497 · Autonomous farming

## Introduction

Current technology is capable of automatically navigating and operating agricultural machinery, such as tractors and harvesters, efficiently and more precisely compared to manual operation. However, a crucial deficiency in this technology concerns the safety aspects. In order for an autonomous vehicle to operate safely and be certified for

---

✉ P. Christiansen  
pech@eng.au.dk

<sup>1</sup> Department of Engineering-Signal Processing, Faculty of Science and Technology, Aarhus University, Finlandsgade 22, 8200 Aarhus N, Denmark

unsupervised operation, it must perform automatic real-time risk detection and avoidance of humans in the field with high reliability (ISO 18497 2015).

Robust risk detection imposes a number of challenges for the sensor system. Varying weather and lighting conditions influence the image quality of sensing technologies in different ways, and thus no sensor is single-handedly capable of detecting objects reliably under all conditions. Active sensors such as LiDAR, and passive sensors such as RGB camera, stereo camera and thermal camera have different strengths and weaknesses concerning weather, lighting, range and resolution, and therefore a variety of these sensors are needed to cover all scenarios (Rasshofer and Gresser 2005). In addition, attitude estimation sensors such as accelerometers, gyroscopes and GPS are needed for estimating the vehicle position, velocity and orientation and for synchronizing and registering subsequent frames acquired from the imaging sensors.

Today, driver assistance systems are available for a large number of modern passenger cars, and completely autonomous vehicles operating in urban and sub-urban environments are emerging for experimental usage (Paden et al. 2016).

In the agricultural sector, a variety of machines have been operating autonomously for a decade using either precise GPS co-ordinates and/or cameras detecting structures in the field (CLAAS Steering Systems 2011; Pilarski et al. 2002). Efforts have been made to fully automate the process in a driverless solution, but safety aspects currently prevent authorization for this. In Freitas et al. (2012), Yang and Noguchi (2012) and Wei et al. (2005), human detection was performed using only a single sensor (laser scanner or stereo camera). However, multiple sensor modalities should be investigated to evaluate their ability to detect humans. For instance, the QUAD-AV project has investigated microwave radar, stereo vision, LiDAR and thermography for detecting obstacles in an agricultural context (Rouveure et al. 2012). Within the project, a detailed study of stereo vision has shown promising results on ground/non-ground classification (Reina and Milella 2012).

In urban environments, autonomous vehicles can exploit obstacles protruding from the surface. In farming operations, obstacles are commonly placed below or just above an uneven surface of crops introducing specific challenges for autonomous vehicles in agriculture. The likelihood of a human being one of these obstacles is small. However, a child or a fallen, injured or unconscious human provides a risk as these non-protruding objects have reduced mobility. To investigate these challenges, data from agricultural fields and algorithms are needed.

Human safety is addressed in ISO 18497 (an emerging standard for safety of highly automated machinery in agriculture) by defining a minimum obstacle that must be detected with an accuracy of 99.99% (ISO 18497 2015). The minimum obstacle is specified as an olive green barrel shaped object that resembles a small or seated human in green clothing (in this paper defined as an ISO-barrel).

This paper describes a flexible vehicle-mounted sensor platform targeting agricultural fields. The sensor platform records imaging data and vehicle position for a moving vehicle using three passive imaging sensors, one active sensor and two attitude/position estimation sensors. The sensor platform is designed to record simultaneous data from all sensors, thus preparing for subsequent offline processing. Offline processing and visualization of sensor data is presented to investigate the object detection potential for the different sensors. The current paper is an extended version of Christiansen et al. (2015) providing more authentic data in grass-harvesting operations and addressing human safety in more detail. An ISO-barrel was produced under the specification defined in ISO 18497. The ISO-barrel as well as humans and mannequins were placed in standing and lying positions in front of the setup to create recordings that could be used in an actual validation of a human detection system

during grass-harvesting. The extended edition also presents a full procedure for calibrating and registering all sensors using a single calibration thermal checkerboard.

## Materials and methods

### Sensors

An overview of the strengths and weaknesses of the selected imaging and active sensors are presented in Table 1. The qualities are evaluated individually and under various conditions. A weakness is marked with ‘–’ and a strength is marked with ‘+’.

Sensor modalities refer to the information a sensor measures. In this paper, a sensor modality is either visual light, depth or heat radiation.

An *RGB camera* captures the modality of visible light. The sensor is useful for identifying the perceived objects as it provides visual characteristics such as texture, color and shape in high resolution at low cost. It is invariant to protrusion, meaning that non-protruding objects such as small animals, a fallen human or humans/animals in high crops are still visible. However, visual characteristics are affected by occlusion from crops, weather conditions (rain, fog and snow) and illumination such as dim light (night) or direct light (causing shadows). An RGB camera is not able to exploit depth information to emphasize protruded objects and the lack of depth makes the positioning of objects in 3D space difficult.

A stereo camera enables 3D imaging data (depth and color information). Depth and color information are registered and the sensor is thus able to exploit the advantages of both modalities. Depth information can be used to see protruded objects and visually camouflaged animals easily while determining the position of an object relative to the vehicle. In this way, depth-aware algorithms can abstract from the very different visual characteristics of objects (shape, color and texture) creating simple detection algorithms. Like the RGB camera, the stereo camera is sensitive to illumination and weather conditions, although the depth information is in some cases still retrievable.

**Table 1** Strengths and weaknesses of sensors (Christiansen et al. 2015)

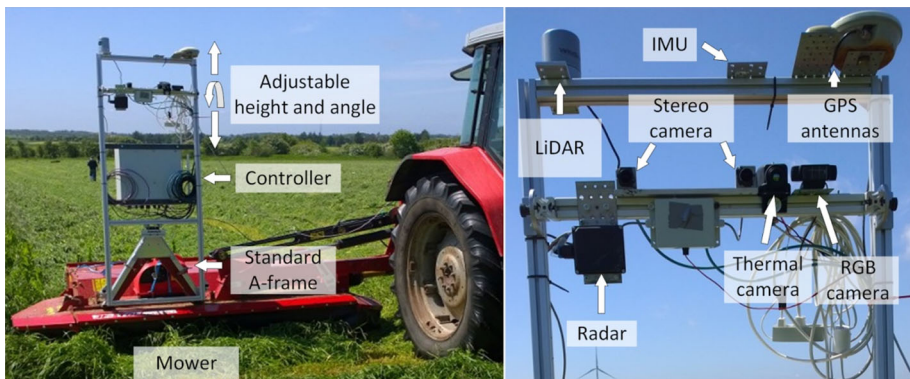
Names	RGB	RGB stereo	Thermal	LiDAR
Specification				
Range	Medium	Medium	Medium	Long
Resolution	+	+	–	–
Depth information	–	+	–	+
Heat information	–	–	+	–
Color information	+	+	–	–
Cost	Low	Medium	Medium	High
Robustness				
Light changes	–	–	+	+
Weather changes	–	–	–	+
Camouflaged objects	–	+	+	+
Protruding objects	–	+	–	+
Non-protruding objects	+	–	+	–

A *thermal camera* is an imaging sensor that captures heat radiation represented by intensities (temperatures) to form a monochromatic image. A thermal camera perceives objects of distinct temperatures, making it ideal for detecting living objects in temperate and colder climates, and even in foggy weather (Serrano-Cuerda et al. 2014). A key ability is that the sensed data are unaffected by non-protruded or visually camouflaged animals and that the distinctness of living objects becomes more apparent at night. However, these capabilities are much affected by the ambient temperature as living objects become indistinct when the temperature difference between the objects and background becomes small (Serrano-Cuerda et al. 2014). The cost of a well-performing and high resolution thermal camera is very high, but low cost cameras are emerging. Object recognition capabilities are low due to a limited resolution and limited visual characteristics.

A *LiDAR* measures range data to a set of surrounding points and generates a point cloud where each point is represented by a 3D position and reflection intensity. The LiDAR is a high cost sensor, but has dropped significantly in price in recent years. Compared to a stereo camera, the LiDAR provides very exact depth information at greater range and some models can capture in 360° horizontally. It is invariant to illumination, temperature and camouflage. The lack of visual and thermal information makes recognition of objects difficult and non-protruding objects are almost or fully undetectable.

### Physical design

The sensor platform consisted of seven sensors and a controller mounted on a common rack of 2 m by 0.8 m in size. The left side of Fig. 1 shows the rack mounted on a tractor and the right side shows the physical placement of sensors. A standard A-frame was mounted at the bottom of the rack to enable easy mounting on tractors. The category 1 A-frame was mounted with dampers for absorbing internal engine vibrations from the vehicle to reduce the amount of mechanical noise acting on the sensors. The horizontal profile in the middle was adjustable in height and angle such that the imaging sensors could be oriented in a downward angle depending on the vehicle height. The LiDAR was placed above the sensor frame to minimize view obstructions for the sensor. The rack allowed sensors to be placed roughly 2 m above ground to provide a more downward view into the crop to better detect hidden obstacles. Placing sensors on top of the tractor would provide a similar downward view. However, the tall rack and the A-frame allowed the sensors to be



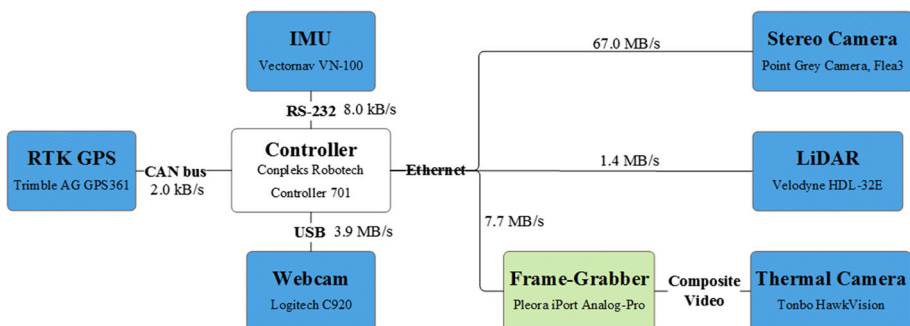
**Fig. 1** *Left* sensor frame including controller. *Right* sensors on the sensor platform

easily swapped to another tractor, an all-terrain vehicle or directly on a ground socket while keeping the downward view under data acquisition.

A Logitech HD Pro C920 from Logitech (Silicon Valley, USA) webcam providing  $1\,920 \times 1\,080$  pixels at 30 fps was used as the RGB camera. The stereo camera was composed of two hardware synchronized Flea3/FL3-GE-28S4C-C cameras from Point Grey (Richmond, Canada) with global shutter and  $1\,928 \times 1\,448$  pixels at 15 fps. The thermal camera was a shutterless HawkVision analog thermal camera from Tonbo Imaging (Bangalore, India) providing  $640 \times 480$  pixels at 25 fps (interlaced). The HDL-32E LiDAR from Velodyne (Morgan Hill, USA) was a 32-beam laser scanner providing 70 000 points at 10 Hz with 1–100 m range. Figure 1 shows an automotive Delphi ESR 64-target radar from Delphi (Washington, DC, USA) not addressed in the current paper as it was intended for detecting pieces of metal and not humans. The GPS was an AG GPS361 real time kinematic (RTK) GPS from Trimble (Sunnyvale, USA) enhancing the precision of GPS up to centimeter-level accuracy. The IMU was a VN-100 from Vectornav (Dallas, USA) providing synchronized three-axis accelerometers, gyros, magnetometers and a barometric pressure sensor. The data-collecting controller was an Innovation Robotech Controller 701 from Conpleks (Struer, Denmark). It is an embedded computer with external interfaces for all sensors that using ROS-middleware (robot operating system) to easily integrate them into a common framework.

## System architecture

Figure 2 further illustrates the connections between the sensors and the controller. In ROS, each sensor was given its own node (an executable file) that was responsible for publishing one or more topics. For instance, the IMU had its own node including hardware-specific drivers, and it published different topics related to the readings of the accelerometer, the gyroscopes and the magnetometers. For each topic, the node could send messages containing sensor data whenever a new sensor-reading was available. Each node was connected to the ROS Master which handled interactions between nodes and supplied all messages with exact timestamps. Using the rosbag package, a recording of all desired topics (and all associated messages) to a single rosbag data-file could be obtained. A JavaScript browser interface was developed to easily monitor and record specific sensors, and enabled the platform to be controlled through Wi-Fi using a mobile phone, tablet or computer.



**Fig. 2** System overview illustrating bandwidths and interfaces for sensors

## Data

Data were collected on a grass field of roughly 7.5 ha near Lem in Denmark (latitude 56.059679° N, longitude 8.368701° E) in the beginning of June 2015. To get authentic data, sensors were mounted to a tractor working in a normal grass-harvesting operation. In operation, obstacles were placed in the trajectory of the tractor to simulate collision hazards. For each obstacle, the tractor approached the object and stopped just before collision. To enable some form of reproducibility and to ensure safety, standing/lying adult and child mannequins were used instead of real humans in the field. To incorporate safety standards, the ISO-barrel was also used. Finally, the mower was turned off and two recordings with real humans were captured. Obstacles from the data are presented in Fig. 3.

In Fig. 4, obstacle positions and the tractor route (divided into laps) are presented, where lap 17 and 18 contained real human obstacles.



**Fig. 3** Two real humans, three mannequins and the ISO-barrel



**Fig. 4** Tractor route (*lines*), barrel (*circles*), kid mannequin (*diamonds*), adult mannequin (*squares*), well (*stars*) and lap starting point (*small dots*)



## Registration of sensors

Registration or sensor fusion is essential for a multi-sensor system to merge and exploit information from all sensors. Registration in multiple modalities is non-trivial and can be handled in different ways (Bahnsen 2013; Zhao and Cheung 2014; Krotosky and Trivedi 2007). In particular, Bahnsen (2013) provided a coherent description of registration methods and the complications for registering different modalities, when objects are not positioned at the same distance.

In this work, common camera and sensor view geometry combined with depth information from the stereo camera were used to project points between sensor frames (Johnson and Bajcsy 2008). Such projections require the *intrinsic* parameters to calibrate cameras individually and *extrinsic* parameters—describing the inter-displacement of sensors—to finalize registration. The inter-displacement between LiDAR and stereo camera was found by matching the two point clouds using the iterative closest point algorithm (Zhang 1994).

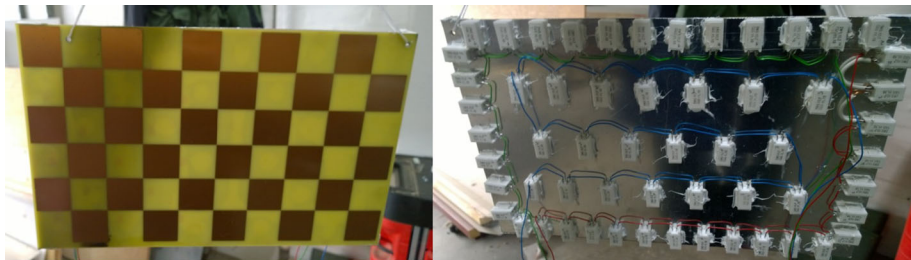
The stereo camera and the webcam was calibrated individually using a normal checkerboard and MATLABs computer vision: calibration tool (2015). The calibration tool was able to detect checkerboards, calibrate cameras, map checkerboard to 3D position automatically and, for the stereo camera, find the inter-displacement between the left and right camera. For the webcam, the extrinsic parameters was determined by finding the transformation that matched corresponding 3D checkerboards to, in this setup, the left stereo camera. However, to calibrate and find inter-displacement between thermal and RGB cameras using a traditional and automated calibration tool, the checkerboard must be visible in both modalities. Therefore, a custom-made visual–thermal checkerboard is proposed.

### Visual–thermal registration

A normal checkerboard exposed to sunlight can be used to perform thermal–visual registration as black absorbs more energy than white areas. However, the quality of the thermal calibration is dependent on weather conditions, and heat/energy is transferred in the material between black and white areas making square transitions indistinct.

A registration/calibration board was therefore developed using a circuit board with copper squares as shown in Fig. 5 (left).

The circuit board was heated by attaching an aluminum plate mounted with impact resistors on the backside of the board as in Fig. 5 (right). The 60 resistors delivered 216 W of heat using a 12 V car battery. Copper has a low emissivity coefficient, which effectively made the material work as a reflector. Thus, the non-copper squares emitted heat radiation



**Fig. 5** Front and back side of registration board

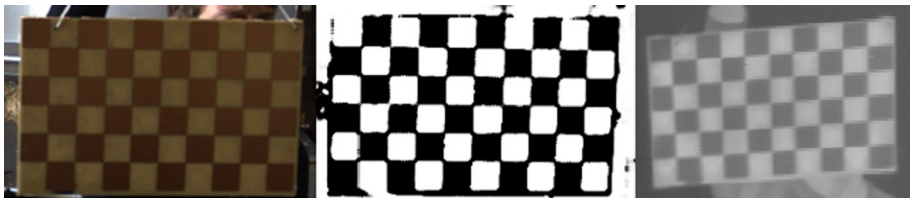
from the heated circuit board, and the copper areas reflected heat of the surroundings, giving a distinct transition between copper and non-copper squares.

The thermal checkerboard would, in a normalized thermal image, resemble a traditional black and white checkerboard as presented in Fig. 6 (right). The thermal camera was then calibrated using traditional and automated calibration tools.

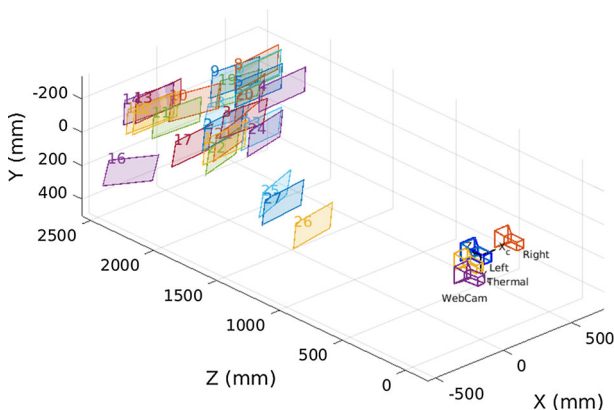
The thermal checkerboard did not, for RGB images, resemble a traditional black and white checkerboard as depicted in Fig. 6 (left). Thus, calibrations tools could not be applied directly. To use RGB images, a MATLAB script was developed to enable a user to mark an area inside the checkerboard. This area was then cropped and converted to the LAB color space. Automatically, the A and B channels were modeled into two clusters using a Gaussian mixture model (McLachlan and Basford 1988). Copper and non-copper areas were separated into two individual clusters. The posterior probability of each pixel belonging to a specific cluster generated a gray-scaled image that made the registration board resemble a traditional black and white checkerboard, see Fig. 6 (mid).

Converting RGB images, enabled all camera sensors to be calibrated and registered using only the proposed registration board. However, the procedure required the user to place a rectangular area inside the checkerboard for each image. In Fig. 7, the detected boards and the inter-displacement of sensors are visualized.

In Fig. 8 (middle left), two humans are annotated in the left stereo camera and projected to the stereo point cloud in Fig. 8 (top). The distance to objects inside the annotation was determined using the median distance of pixels inside the bounding box. The bounding box was then defined as four points in the stereo point cloud that could be projected to other



**Fig. 6** The registration board (*left*) is transformed into a “classic” checkerboard (*mid*) using a Gaussian mixture model. Thermal image of the registration board (*right*)



**Fig. 7** Registration board placements (numbered 1–25) and inter-displacement of sensors





**Fig. 8** Annotations in the *left image* are projected onto the stereo point cloud (*top*). These annotations are then projected to the right and left stereo camera (*middle left and right*), the webcam (*bottom left*) and the thermal camera (*bottom right*)

sensor frames as in Fig. 8. To make a more exact registration of sensors, the registration board should be placed at a broader range of distances from the cameras.

A more quantitative evaluation of the visual–thermal registration is presented in “[Appendix: Thermal–visual registration and evaluation](#)” section.

## Signal processing

To provide an initial qualitative validation of detection performance of the different sensors in an agricultural environment, preliminary tests using different object detection algorithms have been carried out on the sensors.

Using only an RGB camera for detecting all possible obstacles in the field is complex and difficult and requires a very large dataset with many representations of each object. Constraining detection to only humans provided a more realistic case in this preliminary study. The RGB camera images were therefore processed using a state-of-the-art pedestrian detection algorithm (Dollar et al. 2010).

After stereo camera calibration (Zhang 2000), a point cloud could be generated for each stereo image pair. For both stereo and LiDAR, the same algorithm was used to better compare sensors. A ground plane was estimated on the acquired point cloud using the RANSAC algorithm (Fischler and Bolles 1981). Protruding objects were visualized by determining the height of points relative to the estimated ground plane.

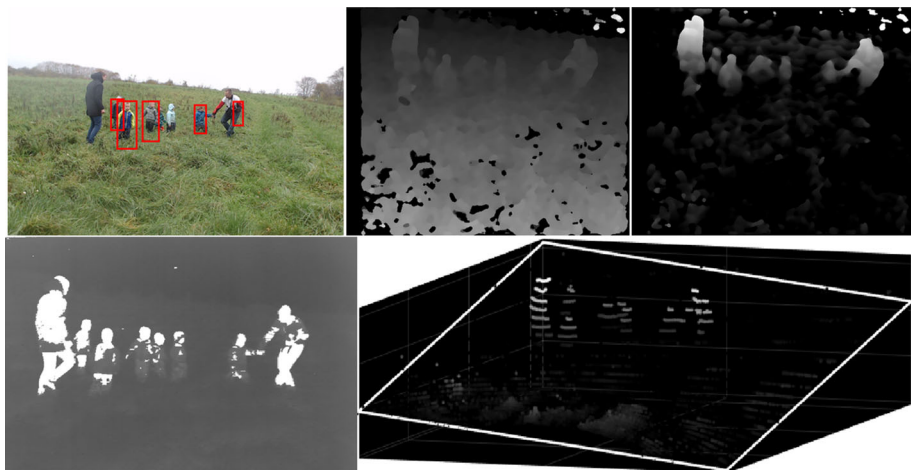
The thermal camera images were processed by thresholding the (temperature-related) intensities by a constant value above the median intensity of the image (Christiansen et al. 2014). Subsequent connected components analysis was used for extracting only components that exceeded a certain area.

## Results and discussion

An initial validation of detection algorithms is presented in four scenarios. The first scenario is humans of different sizes, appearances and postures similar to Christiansen et al. (2015) in low grass. Scenarios 2–4 are, respectively, a barrel, a lying child mannequin and a sitting human in high-grass taken from the above described data.

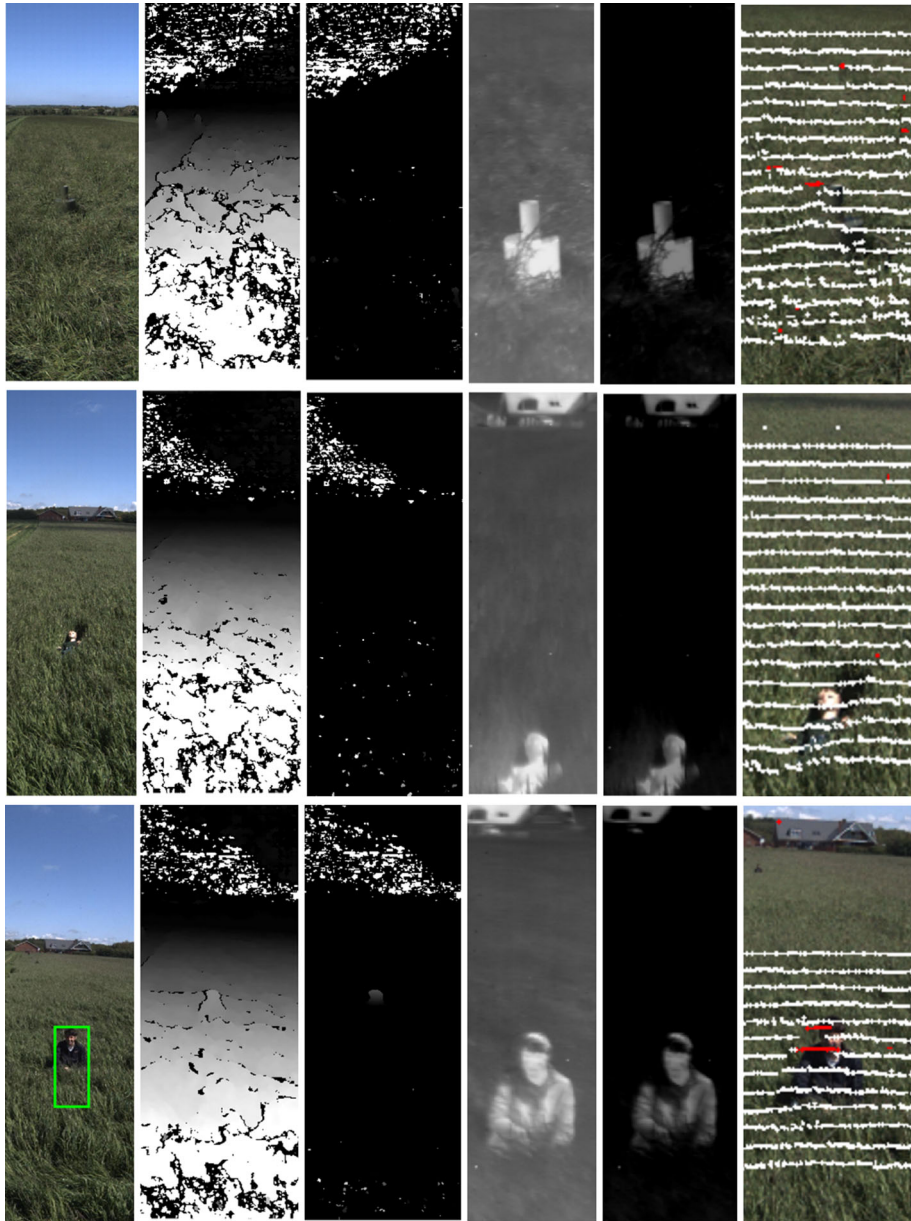
Figure 9 depicts the human detection performance evaluated at single, synchronized frames for the RGB camera, the stereo camera and the LiDAR. At the top left, the RGB camera is shown with bounding boxes indicating results of the pedestrian detection algorithm. In the top middle, the disparity map of the stereo camera is shown and, at the top right, a protrusion map indicating objects that protrude from the ground plane is visualized. At the bottom left, the thermal camera is shown with overlaid thresholded components and, at the bottom right, the LiDAR data are visualized with a ground plane and protruding points.

Using only single frames, pedestrian detection applied to the RGB camera failed to detect all humans in the image. Problems concerning occlusion and humans seen from the side or from behind have been observed. However, utilizing a sequence of frames would greatly improve detection performance, as the algorithm most often failed for just a single frame and not for an entire sequence of frames. The stereo camera performed well for detecting humans that protruded from the ground plane. However, the algorithm assumed a certain level of protrusion in order to detect an object. The thermal camera detected all humans when their faces were visible. However, potential problems concern well-insulated clothes that cover an entire body and warm weather where temperature differences are much smaller than in the present recording. The LiDAR detected most humans robustly when they protruded significantly from the ground.



**Fig. 9** Human detection. RGB (*top left*), stereo camera disparity map (*top middle*), stereo camera protrusion map (*top right*), thermal camera (*bottom left*), LiDAR (*bottom right*; Christiansen et al. 2015)

Figure 10 depicts three cropped scenarios in high grass with respectively a barrel, a lying child mannequin and a sitting human. The pedestrian detector was able to detect the sitting human as the face and torso were upright and visible. To detect the lying



**Fig. 10** The three rows show respectively a barrel, a lying child mannequin and a sitting human. The columns show respectively pedestrian detections, a disparity map from stereo imaging, an object height map based on this, the thermal signature, thermal signature after subtracting the median temperature of the bottom half of the image, and the LiDAR projected onto the left stereo camera, where points protruding from the surface by more than 0.25 m are visualized

mannequin, the detector needs to be trained on new data showing humans in similar scenarios. However, the given detector had limited capacity in terms of detecting objects with huge inter-class variation. In the high-grass case, there was a limited reliability of the stereo point cloud which impacted detection performance such that only the sitting human and not the barrel were visible. Exploiting also visual information from the stereo camera should be utilized to improve performance. The LiDAR was more reliable and was able to visualize that both the sitting human and the barrel protruded. The thermal camera achieved robust and reliable detection performance. In scenarios 2–4, all sensors apart from the thermal camera had problems with high grass/crop, presenting a specific challenge that should be addressed in agriculture. The thermal camera will undoubtedly be significantly worse on a warm and sunny day as experienced by Steen et al. (2012) and Serrano-Cuerda et al. (2014). A single sensor is therefore insufficient for detecting all objects reliably, invariant of temperature and lighting changes.

## Conclusions

A flexible vehicle-mounted sensor platform was developed for capturing time-stamped data in the agricultural domain using imaging sensors (RGB, thermal and stereo camera), an active sensor (LiDAR) and attitude estimation sensors (RTK GPS and IMU). A registration board was proposed to provide a simple tool for calibrating and registering all sensors in the setup using a single registration board. Authentic data in an actual high grass harvesting operation with a specific focus on human detection were recorded, and an initial evaluation of the potential of different sensor modalities for detecting standing and lying humans including an ISO-barrel was given. Using a common pedestrian detection algorithm, an RGB camera was able to detect upright humans, but degraded rapidly in performance for more complex scenarios. The depth aware sensors (LiDAR and stereo camera) were efficient for detecting objects that protruded significantly above the ground. The LiDAR was invariant towards changing weather and lighting conditions, whereas the stereo camera had the highest resolution making it useful for classifying objects. The thermal camera showed great capabilities in the captured dataset as it was able to detect objects of distinct temperature using a simple procedure that worked well for humans regardless of posture. However, the detection would be much more complicated in environments of higher temperature, where the heat signatures of living objects become indistinct.

The authenticity of the data enabled an initial validation of a human detection system using multiple sensors in a high grass harvesting operation. However, the above arguments and the case study concludes that the use of multiple modalities, more complicated procedures and a fusion of the different modalities is required to achieve robust human detection in high grass harvesting.

**Acknowledgements** This research is sponsored by the Innovation Fund Denmark as part of the Project “SAFE - Safer Autonomous Farming Equipment” (Project No. 16-2014-0) and “Multi-sensor system for ensuring ethical and efficient crop production” (Project No. 155-2013-6).

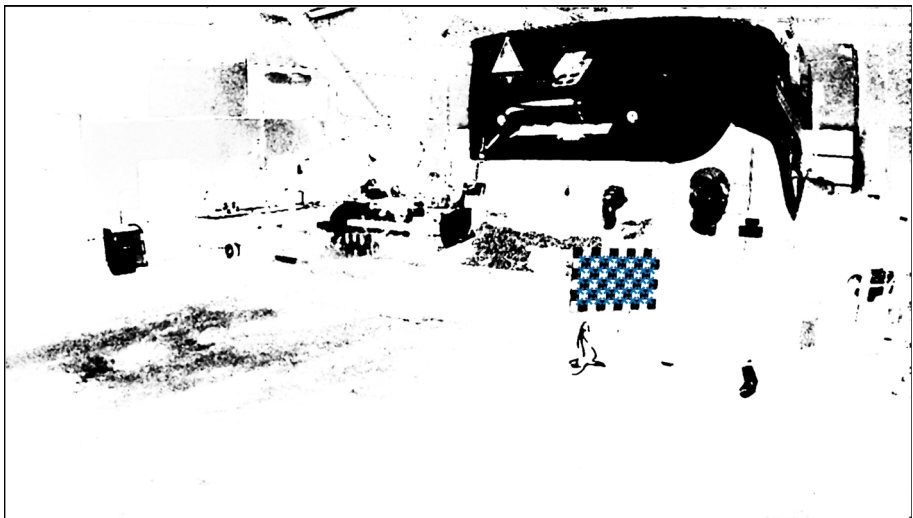


## Appendix: Thermal–visual registration and evaluation

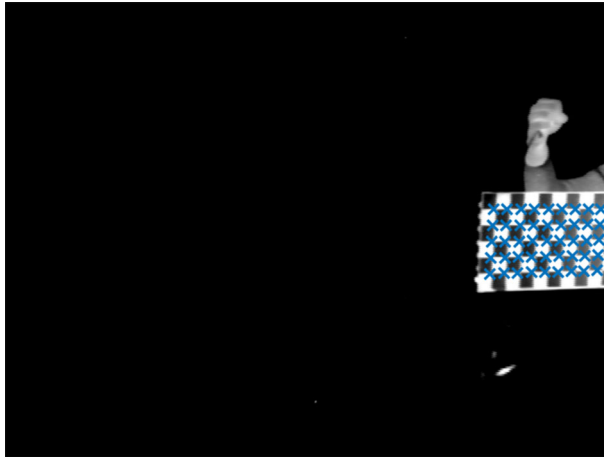
First a total of 47 thermal and stereo synchronized images were selected from a single calibration recording. For each image, a rectangle area inside the checkerboard was marked manually to specify an image cropping, see Fig. 11. For RGB images, the cropped image was converted to the LAB color space and a Gaussian mixture model separated the pixels into two clusters (copper and non-copper areas). The posterior probability of belonging to one of the Gaussian clusters was determined for all pixels in the original image, see Fig. 12. For thermal images, the cropped image was normalized—transforming pixel



**Fig. 11** Image example and a manually marked *rectangle*



**Fig. 12** Posterior probability of belonging to one of the Gaussian clusters for all pixels in the image example. Checkerboard detection is marked with *blue crosses* (Color figure online)

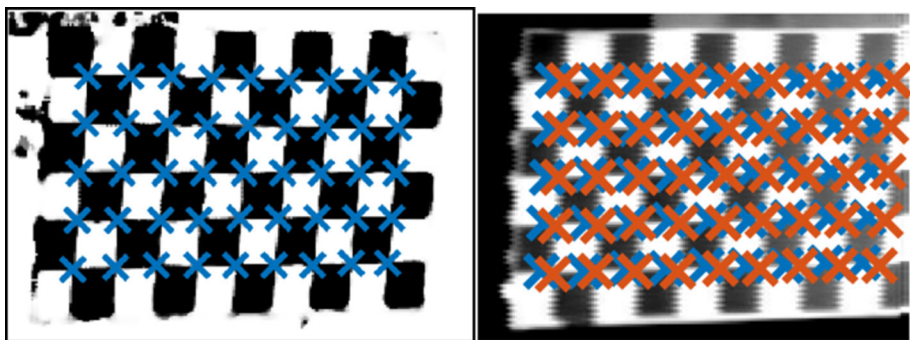


**Fig. 13** Thermal image is normalized relative to the checkerboard. Checkerboard detection is marked with *blue crosses* (Color figure online)

values in the range  $[0\ 1]$  by shifting and scaling. The same normalization was applied to the whole thermal image, see Fig. 13. The MATLAB calibration toolbox was able to automatically detect checkerboards of the transformed RGB and thermal images. The calibration toolbox was able to detect the checkerboard in 27 and 43 out of the 45 images for respectively stereo and thermal images. The 27 stereo images were used for calibrating the intrinsic and extrinsic parameters of the stereo camera. The 43 thermal images were used for determining the intrinsic parameters of the thermal camera.

In 25 out of 47 synchronized images, the checkerboard was successfully detected by the MATLAB calibration toolbox for both RGB and thermal images. The toolbox estimated the 3D position of the checkerboard in all 25 images for each camera. The extrinsic parameters of the thermal camera were determined as the least square rigid transformation that mapped the estimated checkerboards from the left RGB camera to the thermal camera (in 3D).

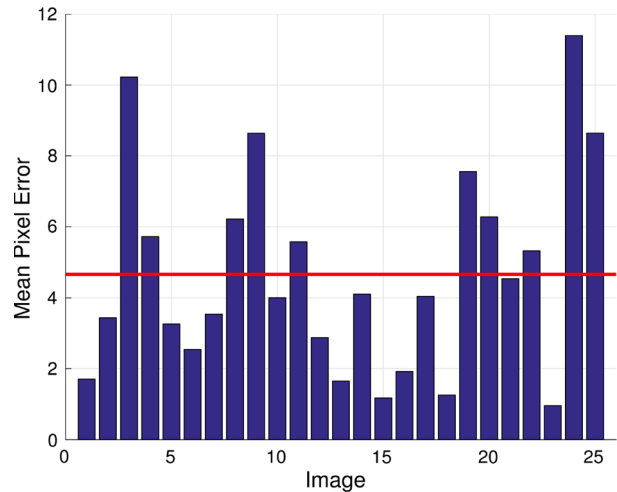
The registration was evaluated on the 25 images to provide a quantitative evaluation of the thermal–visual registration. The camera calibration for the left stereo camera



**Fig. 14** Zoomed images. *Blue crosses* mark *corners* detected by the MATLAB calibration toolbox for both an RGB image (*left*) and a thermal image (*right*). The *red crosses* (*left*) show how 3D points are projected to the thermal camera (Color figure online)



**Fig. 15** The mean pixel error for 25 images (blue bars) and the mean pixel error across all images (red line) (Color figure online)



estimated—as already described—the checkerboard positions in 3D. These positions were then projected to the thermal image using the estimated extrinsic and intrinsic parameters of the thermal camera, see Fig. 4 (right).

The error was determined as the distance between the detected checkerboard and the projected 3D positions. Figure 15 shows the mean pixel error for each of the 25 images and the mean pixel error across all images on 4.66 pixels. The image example used in Figs. 11, 12, 13, and 14 is image 21 with a mean pixel error close to the mean pixel error across all images.

## References

- Bahnsen, C. (2013). Thermal-visible-depth image registration. Unpublished Master Thesis, Aalborg University, Aalborg, Denmark.
- Christiansen, P., Kragh, M., Steen, K. A., Karstoft, H., & Jørgensen, R. N. (2015). Advanced sensor platform for human detection and protection in autonomous farming. *Precision Agriculture*, 15, 291–298.
- Christiansen, P., Steen, K. A., Jørgensen, R. N., & Karstoft, H. (2014). Automated detection and recognition of wildlife using thermal cameras. *Sensors*, 14(8), 13778–13793.
- CLAAS Steering Systems. (2011). *Tracking control optimisation*. Retrieved 2016, 26 September from <http://claas.via-us.co.uk/booklets/gps-steering-systems/download>.
- Dollar, P., Belongie, S., & Perona, P. (2010). The fastest pedestrian detector in the west. In F. Labrosse, R. Zwigglelaar, Y. Liu & B. Tiddeman (Eds.), *Proceedings of the British machine vision conference 2010* (pp 68.1–68.11). BMVA Press, Durham University, UK.
- Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.
- Freitas, G., Hamner, B., Bergerman, M., & Singh, S. (2012). A practical obstacle detection system for autonomous orchard vehicles. In *2012 IEEE/RSJ international conference on intelligent robots and systems* (pp 3391–3398).
- ISO/DIS 18497:2015: *Agricultural and forestry tractors and self-propelled machinery—Safety of highly automated machinery*. Retrieved 2016, 26 September from <https://drive.google.com/file/d/0B1ilODNTH9nzRUV2N0JzbklubFU/view>.
- Johnson, M. J., & Bajcsy, P. (2008). Integration of thermal and visible imagery for robust foreground detection in tele-immersive spaces. In P. Solbrig (Ed.), *Proceedings of the 11th international conference on information fusion* (pp. 1265–1272). Piscataway, USA: IEEE.

- Krotosky, S. J., & Trivedi, M. M. (2007). Mutual information based registration of multimodal stereo videos for person tracking. *Computer Vision and Image Understanding*, 106(2–3), 270–287.
- McLachlan, G. J., & Basford, K. E. (1988). Mixture models: Inference and applications to clustering. In *Statistics: textbooks and monographs*. New York, USA: Dekker.
- Paden, B., Cáp, M., Yong, Z. S., Yershov, D., & Frazzoli, E. (2016). A survey of motion planning and control techniques for self-driving urban vehicles. *IEEE Transactions on Intelligent Vehicles*, 1(1), 33–55. [arXiv:cs.CV/1604.07446v1](https://arxiv.org/abs/1604.07446v1).
- Pilarski, T., Happold, M., Pangels, H., Ollis, M., Fitzpatrick, K., & Stentz, A. (2002). The Demeter System for automated harvesting. *Autonomous Robots*, 13, 9–20.
- Rasshofer, R. H., & Gresser, K. (2005). Automotive radar and lidar systems for next generation driver assistance functions. *Advances in Radio Science*, 3, 205–209.
- Reina, G., & Milella, A. (2012). Towards autonomous agriculture: Automatic ground detection using trinocular stereovision. *Sensors*, 12(12), 12405–12423.
- Rouveure, R., Nielsen, M., & Petersen, A. (2012). The QUAD-AV Project: Multi-sensory approach for obstacle detection in agricultural autonomous robotics. In *International conference of agricultural engineering*. Valencia, Spain: EurAgEng.
- Serrano-Cuerda, J., Fernández-Caballero, A., & López, M. (2014). Selection of a visible-light vs. thermal infrared sensor in dynamic environments based on confidence measures. *Applied Sciences*, 4(3), 331–350.
- Steen, K. A., Villa-Henriksen, A., Therkildsen, O. R., & Green, O. (2012). Automatic detection of animals in mowing operations using thermal cameras. *Sensors*, 12(6), 7587–7597.
- The MathWorks, Inc. (2015). *MATLAB and computer vision system toolbox*. Natick, MA, USA: The MathWorks, Inc.
- Wei, J., Rovira-Mas, F., Reid, J. F., & Han, S. (2005). Obstacle detection using stereo vision to enhance safety of autonomous machines. *Transactions of the ASAE*, 48(6), 2389–2397. doi:[10.13031/2013.20078](https://doi.org/10.13031/2013.20078).
- Yang, L., & Noguchi, N. (2012). Human detection for a robot tractor using omni-directional stereo vision. *Computers and Electronics in Agriculture*, 89, 116–125.
- Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2), 119–152.
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.
- Zhao, J., & Cheung, S. S. (2014). Human segmentation by geometrically fusing visible-light and thermal imageries. *Multimedia Tools and Applications*, 76(1), 7361–7389.