

Depth Resolution of 3D Imaging Techniques for Target Detection in Mobile Work Machines

Olli J Suominen

Faculty of Information Technology and
Communication Sciences
Tampere University
Tampere, Finland
olli.j.suominen@tuni.fi
<https://orcid.org/0000-0001-9766-4249>

Laura Goncalves Ribeiro

Faculty of Information Technology and
Communication Sciences
Tampere University
Tampere, Finland
laura.goncalvesribeiro@tuni.fi

Atanas Gotchev

Faculty of Information Technology and
Communication Sciences
Tampere University
Tampere, Finland
atanas.gotchev@tuni.fi

Abstract—3D imaging is important for enabling autonomous operation of smart mobile work machines. Different sensing techniques have different characteristics, which affect the choice of sensors for each application. We study the depth resolution of three different active imaging technologies and stereoscopic depth estimation with multiple different baselines. We test the effect of depth measurement abilities in measuring average depth, angles between planar structures and diameter from logs. Active sensors show their robustness, while stereoscopic depth estimation follows the error behavior expected from the theory. However, in practice we find that the effect of poor depth resolution can be significantly reduced by averaging over multiple measurements from different points and using more sophisticated stereo processing.

Keywords—3D imaging, stereo, Time-of-Flight, LIDAR, depth resolution

I. INTRODUCTION

Heavy mobile work machines, such as excavators, tractors and forest harvesters, are becoming increasingly intelligent. While still far from full autonomy when compared to autonomous cars, different semi-intelligent features are starting to appear even in commercially available machines. In order for the machines to be able to assist the user or to complete tasks, they must be able to detect their work targets. In the broad meaning of the term, this is a machine vision task. However, current commercial machine vision solutions for determining work targets for grabbing and handling are mainly based on 2D cameras, optimized lighting conditions and image matching, such as having a top down view of a conveyor belt with constant shaped objects at a known distance. In contrast, mobile work machines operate in environments, which are not and cannot be designed for optimal machine vision – the terrain and lighting are constantly changing and the camera has to be placed on the machine itself, limiting the available viewpoints. Using 3D imaging can alleviate these concerns. While the solutions for the field are not as cost sensitive as in the automotive sector, the cost of sensors is still an important consideration.

In this work, we experiment with commonly available sensors and evaluate their suitability for measuring different metrics from the observed targets. Special attention is paid to stereo cameras as the most affordable depth measurement device for dense depth sensing. The distance measurement resolution of stereo pairs is highly dependent on the distance between the cameras, i.e. the baseline. The effect of this on the ability to measure relevant information is quantified.

II. PRIOR WORK

A. Measurement of work targets

The topic of 3D sensing in the context of mobile work machines has been studied before, but the lack of a common umbrella term makes it hard to identify comprehensive surveys. Papers tend to focus on a specific, narrow application area, such as agriculture or palletizing, or on the more generic tasks of obstacle detection and navigation in driving. Out of the common fields, precision agriculture [1] is one of the active and most mature fields with works related to 3D imaging ranging from obstacle avoidance [2] to mapping [3] and from plant detection [4] to assisting in tool attachment [5]. Another relevant application is automatic palletizing of goods for storage and transportation, which employs computer vision in two stages. First, the placing of cargo (e.g. boxes) onto the pallet [6], and second, finding and lifting the loaded pallets to be transported by forklifts or other similar. These tasks do not demand such flexible solutions, since the environment is typically an indoor setting such as a warehouse.

Another largely studied area is related to industrial robotics, but there the typical working distances are much shorter (up to ~1 meter) and the environments are limited to factories and assembly lines, so we consider these works to be not of interest in this application. Conversely, the reconstruction of forests, mines and agricultural environments has been studied through photogrammetry from UAVs [5], where the scale is much larger (upwards ~100 meters). The range of interest for most work machine applications is in between – approximately 1 to 10 meters.

B. Stereo resolution

A specific interest for this work is to study the effect of stereo baseline for the task specific measures. Generic work on stereo depth resolution has been done [7] and the theory is well known (as is described in section III.C Metrics). Our interest is to test the practical implications of the reduced depth resolution and to quantify the effect.

In [8], the adverse effect of the baseline on the ability to reconstruct fronto-parallel surfaces was studied. The conclusion was that larger baselines distort the 3D projection and make it more difficult to find matching points.



Fig 1. The test scene consisting of a pallet loaded with a small log, a container with a 90° degree angle and a smooth plane. The pallet is set on a wheeled platform to allow changing the distance from the cameras

These tests hint there is also motivation against arbitrarily increasing the stereo baseline in search for better depth resolution. Though it is possible, that the limitations reported could be circumvented by using stereo algorithms, which, unlike the semi-global matching employed in the study, do not suffer from getting slanted views of planar structures [9].

III. TEST SETUP

For measuring the properties of the sensors, we use a test scene with several objects, each providing a different variable to be measured. The scene is placed at varying distances of an array of sensors, and the relevant metrics are extracted. Some manual assistance in selecting and segmenting the objects of interest is used, since creating automatic systems is not the objective of this work.

A. Test scene

The test scene consists of three objects of interest (Fig 1), each with a distinct purpose relevant to the context – a plane, a log and a box. The fronto-parallel plane is 58 x 58 cm and has a natural wooden texture, which is highly effective for finding stereo correspondences. The log is a short segment from a full size log cut by a forest harvester with a radius of 11.8 cm. The box is 42 x 58 x 46 cm with mock shipping labels taped to the sides for texture to assist the stereo algorithms.

The regions of interest from the plane and the box are marked with red tape to make sure the measurement areas are selected consistently throughout different distances. The clamp across the length of the log achieves the same for the log.



Fig 2. The sensor array used for testing. On the top, 20 identical calibrated Basler cameras. Bottom row: Velodyne VLP-16, Basler ToF, Faro Focus

Table 1. Models and the manufacturer supplied specifications of the 3D measurement devices used in the experiments

	Faro Focus x130	Velodyne VLP-16	Basler ToF640-20gm	Stereo Basler acA1920-50gm
Working range	0.6 – 130 m	100 m	13 m	~Inf
Depth accuracy	2 mm	3 cm	1 cm	1 / distance
FoV	360°x 300°	360° x 30°	57° x 42°	82° x 51°
Resolution (angular)	0.009°	0.1° / 2.0°	0.09°	0.04°
Resolution	49 600 px	3600 x 16 px	640 x 480 px	1920 x 1200 px
Framerate	5-45 min / scan	20 Hz	20 Hz	50 Hz
Approx. price	40 000 €	4 000 €	2000 €	2 000 €

B. Used sensors

We employ four different types of 3D sensing devices for the experiment. The exact models and their specifications are listed in Table 1 and the sensor configuration is shown in Fig 2. We consider a high-end survey laser scanner Faro Focus X130 as the best-case sensor. However, it's operating principle of a single high accuracy laser beam scanning with two rotating axes makes it too slow, fragile and expensive for being deployed in operational use, and therefore the captured data is used only as a ground truth.

For stereoscopic depth estimation, we use a linear multi-camera setting originally constructed for horizontal-parallax-only light field applications. It consists of 20 identical Full HD resolution industrial cameras and low distortion lenses with 6 mm nominal focal lengths. The cameras are jointly calibrated for their intrinsic and extrinsic parameters [10]. For extracting the depth, the first camera from the right and the Nth camera are consecutively chosen as a stereo pair and rectified. Two different stereo algorithms are applied on the data: the simple but fast block matching approach without any attempts at sub-pixel resolution matching [11] and semi-global matching [12].

As a flash LIDAR (also sometimes referred to as a time-of-flight camera), we use ToF640-20gm from Basler. This device suffers from sensitivity to ambient light, making it difficult to utilize it in outdoor applications. Other devices with similar characteristics, but with more ambient light tolerance exists, though.

As a scanning LIDAR, we use a 16-channel laser scanner from Velodyne. It is one of the most commonly used LIDARs in the automotive sector. The single rotating component makes it prone to vibrations in a demanding environment, but there are expectations that similarly capable devices with more robust solid-state designs will emerge in the near future.

C. Metrics

In order to quantify the performance of the different stereo baselines and sensing devices, we adapt the following metrics.

1) Theoretical depth resolution

The theoretical upper bound of a stereo pair can be derived from the geometrical relations between the two cameras and can be found stated in many sources, such as [13]. Since the derivation offers some insight into the cause of the resolution problem, we present it here briefly. The fundamental relation between the different stereo parameters is

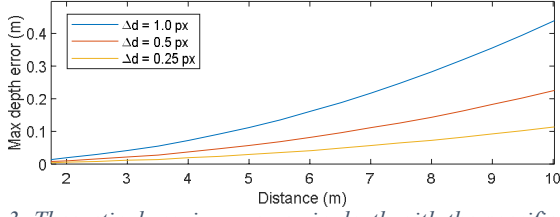


Fig 3. Theoretical maximum error in depth with the specifications of two Basler acA1920-50gm with 6mm lenses given different assumptions on the resolvability of disparity (Δd)

$$z = \frac{bf}{d}, \quad (1)$$

where b is the baseline between cameras, f the focal length of the cameras, d the disparity between the cameras for a point, and z the metric depth. The disparity can be detected only up to a certain resolution, which in turn depends on the algorithm and assumptions made. A straightforward viewpoint is that since the scene is sampled by its pixels, the location of a point can be only determined within one pixel. A more optimistic view is that one can assume certain characteristics about the smoothness of the sampled colours. Thus, fitting functions of varying degrees as a part of the matching process can help in locating a point at resolutions higher than one pixel [14]. Therefore, varying opinions on the achievable disparity estimation resolution exist.

The maximum depth error we are looking for is the difference between the true value z_t and the measured value z_m ,

$$\Delta z = z_t - z_m = \frac{bf}{d} - \frac{bf}{d + \Delta d}. \quad (2)$$

To discover the depth error as a function of depth, we insert Eq. (1) back into the relation, and get

$$\Delta z = \frac{z^2 \Delta d}{bf + z \Delta d}. \quad (3)$$

For example, on a stereo pair of two Full HD resolution cameras with 6 mm nominal focal length optics separated by a 20 cm baseline, the error rate follows the curves presented in Fig 3.

2) Metric depth and planar smoothness

The scene contains a large, smooth planar structure roughly orthogonal to the surface. We segment out the plane as indicated by the marked ROI and orient it to align with the XY-plane using Principal Component Analysis on the high quality laser scan from Faro. From the selected points, we compute the values of interest: mean and standard deviation.

3) Angle between container sides

From the box, we select the area marked by tape. Planes are fit on the sides of the box using M-estimator Sample Consensus (M-SAC) as implemented in Matlab, and the angle between them is measured. The expectation is for the corner to be 90 degrees, and any variations from that are attributed to sensing errors.

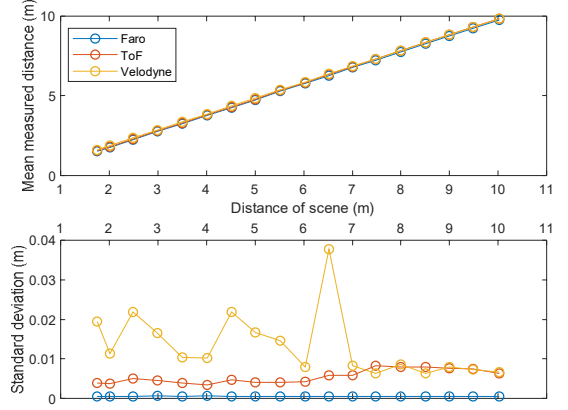


Fig 4. Mean distance and standard deviation using the active sensors

4) Log diameter

From the log, we select the area limited by the clamp and the end of the log. The aim is to measure the circumference of the log based on the reconstructed points on the visible arc of the cylindrical shape. First, a simple background removal is done using distance based clustering. Non-linear optimization is applied to orient the length of the log with the y-axis, i.e. finding a pose for the log that minimizes surface normals with vertical components. The points are projected to the XZ-plane and a circle is fitted on them using Direct least-squares fitting [15] with M-SAC. To minimize the effect of the inherent M-SAC randomness, the circle is detected in 100 iterations and averaged to find the final estimate.

IV. ANALYSIS OF RESULTS

In the experiments, we captured the tests scene described in section III using the four different sensors. The cart was gradually moved further away from the sensors at 0.5 m increments from 2 to 10 meters and captured at each position. The distance increments were verified using a laser measurement tool pointed at the large planar structure, reported as the ‘scene distance’ axis in the figures. The following details the extracted measurements as described in section III.C.

A. Metric depth

The metric depth measurements (Fig 4) using the active sensors are highly robust. The average distances as measured from the wooden plane are consistent between all devices and the incremental motion of the cart. When looking at the spread of values through standard deviation, Faro remains consistent throughout the measuring scale. The ToF camera starts diverging slightly at the farther end of the scale. The Velodyne appears to suffer from the relatively small amount of points hitting each objects, making it prone to outliers.

In Fig 5, the same measurement is done using Block Matching stereo on the varying baselines of the multi-camera rig. Though the expectation is for the depth resolution to degrade as the distance increases, the averaging over the whole plane keeps the mean value constrained and it starts diverging only at the far end with the smallest baselines. The standard deviation tells that the individual measurement points do increase, and the spread (error) is bigger for smaller baselines. The same can be seen in Fig 6 for all baselines. Furthermore, the spread starts to increase for close distances with larger baselines, confirming the findings from [8].

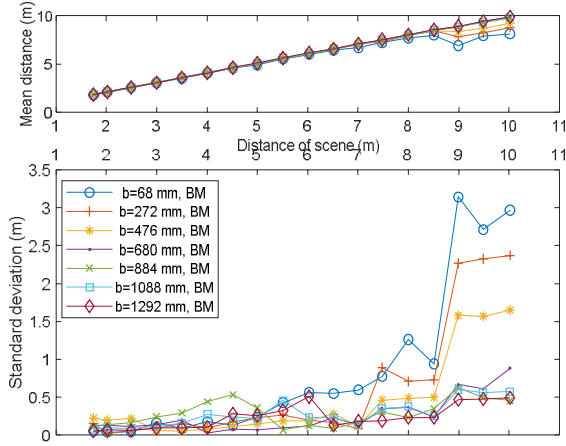


Fig 5. Distance measured by Block Matching stereo

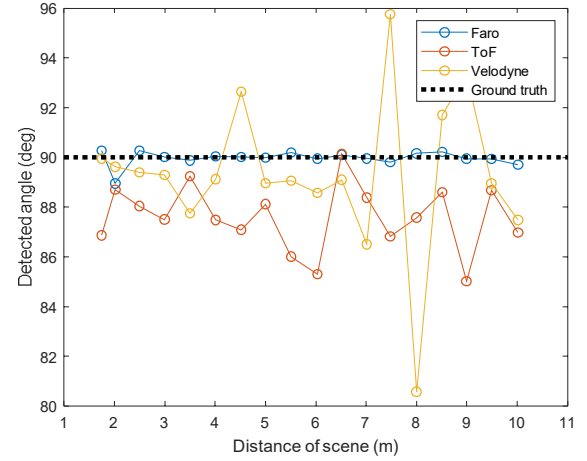


Fig 8. Measured angle between container sides using active sensors

B. Angle between container sides

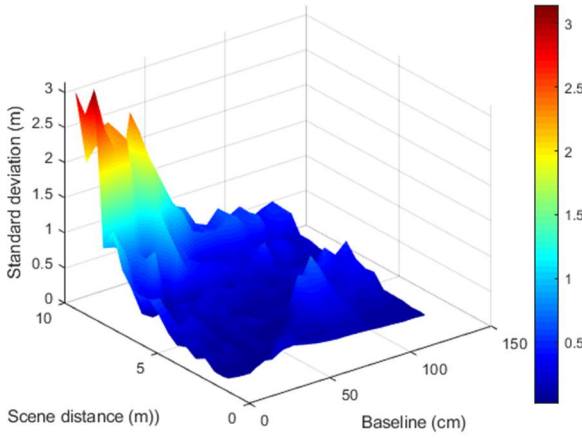


Fig 6. Standard deviation of the distance measurements from all evaluated stereo baselines for Block Matching stereo

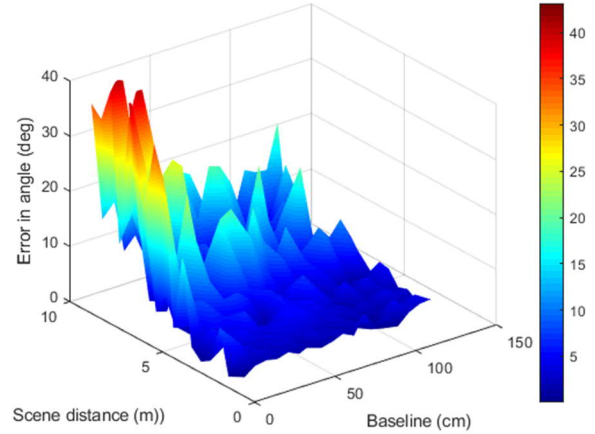


Fig 9. Measured angle between container sides using BM stereo

When measuring the angle between the sides of the container Fig 8, Faro performs consistently, but both Velodyne and the ToF are starting to have some difficulties. Especially the sparse sampling of points for Velodyne makes it difficult to find the planes, often having only a dozen or so points hitting the container.

For BM stereo (Fig 9), the angle is being found reasonably (considering an error below 5 degrees reasonable) for distances up to 5 meters, after which the performance starts to degrade. Larger baselines improve the situation, but good results are not achieved even then. SGM performs slightly better with short baseline and large distance, but there is no significant differences.

C. Log diameter

Faro is again reliable in measuring the log at all distances (Fig 11). The ToF starts to under-estimate the radius as the distance increases. As the log segment is fairly short, Velodyne has again problems in catching enough points from it at longer distances.

The radius measurement using Block Matching stereo in Fig 13 follows the same general trend in the depth error. Until 5 meters, any baseline is able to measure the objects reasonably. When the distance increases, the circular shape is

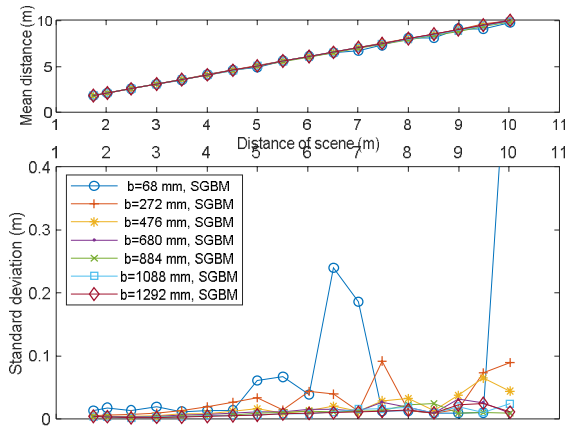


Fig 7. Distance measured by Semiglobal Matching

For the somewhat more advanced version of stereo depth estimation, Semi-Global Matching, there is even less degradation of quality even with small baselines (Fig 7), with all baselines agreeing on the depths. The spread of the values is also almost an order of magnitude less. This is no surprise, since the algorithm intentionally tries to find consistent depth maps by propagating disparity estimates over the image.

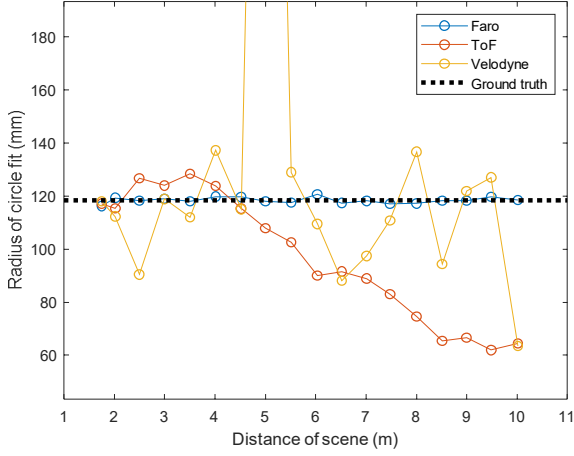


Fig 11. Measured log radius using active sensors

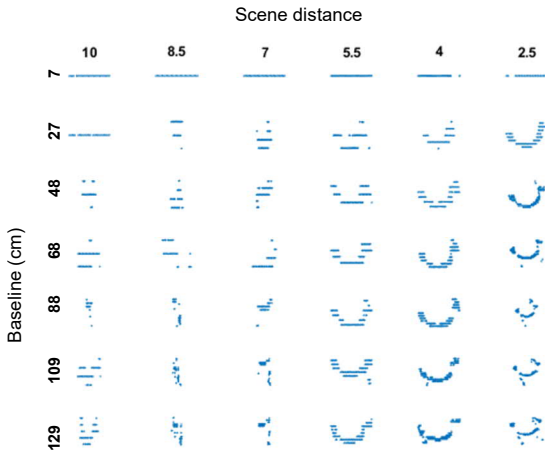


Fig 10. XY-projections of the surface reconstruction of the log at different distances and baselines, demonstrating the challenge in detecting and measuring the circle and its radius

no longer detected at all (baselines shorter than 50 cm), or the error grows steeply to the point of being unusable. No increase in baseline helps in accurately measuring the distance. The limiting factor is likely the spatial resolution, as the log is reconstructed as a plane instead of a cylindrical shape. This can clearly be seen in Fig 10, where the XY-projection of far-away targets is essentially a line. Interestingly, the supposedly better SGM stereo has worse performance in measuring the radii. This can be attributed to the tendency of SGM to adjust depth estimates based on consistency more so than BM simply aggregating with local windows, losing the needed detail for accurate measurements.

V. CONCLUSIONS

The performed experiments highlight the capabilities of different techniques for different tasks. Despite the underwhelming theoretical depth resolution of stereo, it can actually be used reliably for measurement of distances given that the measured target surface can be identified. The distribution of the depth ranging error appears to be symmetric, allowing it to be averaged out with relative ease. For most tasks, the cut-off distance for reliable operation (with the tested optics and resolution) seems to be around 5 meters for the best performing baseline.

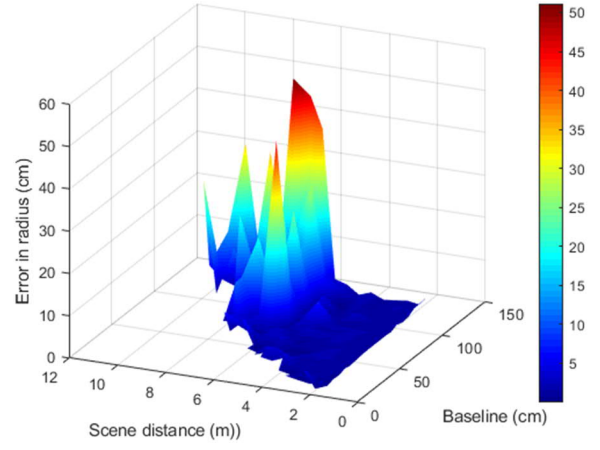


Fig 13. Measured log radius using block matching stereo

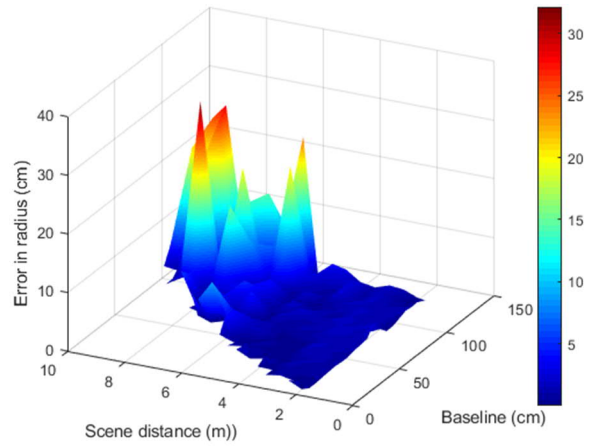


Fig 12. Measured log radius using SGM stereo

For tasks requiring identification from the reconstructed scene (in this paper, measuring the angle of the container corner and log radius), the stereo depth resolution really is an issue. This limits the usability of stereo in so-called RGB-D approaches, where classification or segmentation of performed on the combination of colour and geometry. If such processing is intended, it may be justified to use active sensing to have a more robust reconstruction.

We also confirm the compromise in increasing the stereo baseline. While the increase in baseline increases depth resolution at a distance, there is also a reason not to increase it arbitrarily. When the baseline increases too much, the perspective distortion of nearby objects makes reconstructing them by finding correspondences more difficult.

The crucial factor for using stereo as a depth measurement tool is in distinguishing between measuring a distance to a known object, and identifying an object based on the measured geometry. For instance, if a cargo container can be identified based on its colour information from a 2D image, its distance can be measured using a stereo camera setup to a precision much higher than implied by the theoretical limits. On the other hand, it may not be the best idea to use stereo to analyse trees for harvesting from afar, since the ability to measure the shape is questionable.

ACKNOWLEDGMENT

The work presented in this paper was supported by FIMA Forum for Intelligent Machines ry and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764951. The research infrastructure of Centre for Immersive Visual Technologies (CIVIT) at Tampere University provided the sensors and the laboratory space for conducting the experiments.

REFERENCES

- [1] M. Vázquez-Arellano *et al.*, "3-D Imaging Systems for Agricultural Applications—A Review," *Sensors*, vol. 16, no. 5, p. 618, Apr. 2016, <https://doi.org/10.3390/s16050618>.
- [2] G. Reina and A. Milella, "Towards Autonomous Agriculture: Automatic Ground Detection Using Trinocular Stereovision," *Sensors*, vol. 12, no. 9, pp. 12405–12423, Sep. 2012, <https://doi.org/10.3390/s120912405>.
- [3] F. Rovira-Más, J. F. Reid, and Q. Zhang, "STEREOVISION DATA PROCESSING WITH 3D DENSITY MAPS FOR AGRICULTURAL VEHICLES," *Trans. ASABE*, vol. 49, no. 4, pp. 1213–1222, Oct. 2013, <https://doi.org/10.13031/2013.21721>.
- [4] U. Weiss, P. B.-R. and autonomous systems, and undefined 2011, "Plant detection and mapping for agricultural robots using a 3D LIDAR sensor," *Elsevier*.
- [5] E. Honkavaara *et al.*, "Processing and Assessment of Spectrometric, Stereoscopic Imagery Collected Using a Lightweight UAV Spectral Camera for Precision Agriculture," *Remote Sens.*, vol. 5, no. 10, pp. 5006–5039, Oct. 2013, <https://doi.org/10.3390/rs5105006>.
- [6] C. Prasse *et al.*, "SSMI 12 - New Approaches for Singularization in Logistic Applications Using Low Cost 3D Sensors," *Smart Sensors, Meas. Instrum.*, vol. 12, 2015, https://doi.org/10.1007/978-3-319-12898-6_10.
- [7] S. D. Blostein and T. S. Huang, "Error Analysis in Stereo Determination of 3-D Point Positions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 6, pp. 752–765, Nov. 1987, <https://doi.org/10.1109/TPAMI.1987.4767982>.
- [8] W. Boonsuk, "Investigating Effects of Stereo Baseline Distance on Accuracy of 3D Projection for Industrial Robotic Applications," in *5th IAJC/ISAM Joint International Conference*, 2016, p. 9.
- [9] D. Gallup, J.-M. Frahm, P. Mordohai, Q. Yang, and M. Pollefeys, "Real-Time Plane-Sweeping Stereo with Multiple Sweeping Directions," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8, <https://doi.org/10.1109/CVPR.2007.383245>.
- [10] L. Goncalves Ribeiro, A. Durmush, O. Suominen, and A. Gotchev, "Photogrammetric Multiple Camera Calibration Using an Industrial Robotic Arm," in *Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis*, 2019.
- [11] K. Konolige, "Small Vision Systems: Hardware and Implementation," in *Robotics Research*, 1998, pp. 203–212, https://doi.org/10.1007/978-1-4471-1580-9_19.
- [12] H. Hirschmuller, "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 807–814, <https://doi.org/10.1109/CVPR.2005.56>.
- [13] D. Gallup, J.-M. Frahm, P. Mordohai, and M. Pollefeys, "Variable baseline/resolution stereo," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8, <https://doi.org/10.1109/CVPR.2008.4587671>.
- [14] M. Shimizu and M. Okutomi, "Precise sub-pixel estimation on area-based matching," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, pp. 90–97, <https://doi.org/10.1109/ICCV.2001.937503>.
- [15] V. Pratt, "Direct least-squares fitting of algebraic surfaces," in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques - SIGGRAPH '87*, 1987, vol. 21, no. 4, pp. 145–152, <https://doi.org/10.1145/37401.37420>.