

Visual Odometry and Mapping under poor Visibility Conditions using a Stereo Infrared Thermal Imaging System

Raimund Edlinger, Gabriel Himmelbauer, Gerald Zauner; University of Applied Sciences Upper Austria, Wels, Austria
Andreas Nüchter; Informatics XVII - Robotics at the Julius-Maximilians University Würzburg, Germany

Abstract

In the past decades, developments in the field of computer vision have made both the software and hardware more capable and more easily accessible. This has enabled otherwise complex vision systems to be used in other fields, such as autonomous robotics. Although image processing systems in the visible light spectrum are commonplace in robotics today, the IR spectrum is still rarely used, although it offers certain advantages. A thermal camera can sense the temperature of objects, is independent of illumination and can actually see through heavy smoke and fog. This makes it a useful tool in particular in the field of rescue robotics, where poor vision conditions are to be expected. In this paper, the feasibility of using two thermal cameras in a stereo vision setup to map indoor scenes is to be examined. It is meant to allow an autonomous robot to perceive its indoor surroundings as a 3D space, even in poor vision conditions. The biggest challenges are the calibration of thermal cameras and the proper filtering of the raw image and the resulting disparity map. Simple and easily implemented solutions are proposed for each of these issues.

Introduction

In the past two to three decades, a lot of progress has been made in the field of computer vision. Driven by the increased availability and affordability of hardware, in combination with an increase in computational power, a lot of developments have been made in the field. One of these developments, which is relevant to this paper, is the ability of autonomous vehicles to analyse their environment using such computer vision systems and to then draw conclusions and to act accordingly. While a lot of research has been conducted with vision systems that sense visible light, precious little has been done to explore the possibilities of using exclusively infrared vision systems for this task.

The aim of this work is to investigate the application of stereo camera systems operating in the long wave spectral range (LWIR, long wave infrared, 8 μm - 12 μm). In this way, such systems use the thermal radiation emitted by objects instead of reflected light. The main advantage is that autonomous robotic vehicles could now be used in heavily smoky environments. Conventional systems operating in the visible range (VIS, 0.4 μm - 1 μm) no longer provide reliable image data in such situations due to absorption and scattering effects of light. The desire for a system such as the one proposed in this paper arose from concrete prob-



Figure 1: DRZ challenge scenario at the RoboCup German Open 2022: An accident in a chemical laboratory with explosive substances has been reported and limitations of visual and LiDAR sensor in dense smoke [5]

lems in the application area of robotic "Search and Rescue" missions. The question was raised how a robot can find its way through an environment with poor visibility, due to either heavy smoke or complete lack of illumination, and still perform certain tasks, such as localizing humans or mapping the area. This paper presents first experiments with a stereo vision system that enables a mobile robot to analyze indoor environments in situations with poor visibility due to e.g., heavy smoke, thick gas or a lack of light sources in the visible light spectrum.

Related Work

This chapter briefly summarizes the major work and advances in the field of stereo image processing showing that there is only very little literature on infrared based stereo systems. In addition, the main differences in the operation of such systems will be discussed. Finally, the important process of camera calibration in this context is discussed.

Visible spectrum stereo vision systems (VIS)

The central problem in stereo applications is finding corresponding pixels in the two respective camera images. This issue was solved by Hirschmüller in [9], whose semiglobal matching algorithm has become the de facto standard in stereo vision. Most research thereafter has been focused on applying this algorithm in different situations, like in [3], [7], [22], or developing novel algorithms for very specific applications, like in the work [6].

The technology for VIS-stereo-systems has now reached a level of maturity that allows such systems to be purchased commercially. Well known companies, like "FLIR" [26] and "Intel" [11], and lesser known companies that specialise in stereo vision, like "Stereolabs" [23] and "nerian Vision Technologies" [16], offer such stereo vision systems.

Combined approaches (VIS + IR)

In contrast to VIS-stereo-vision with disparity map algorithms [8], vision systems using multiple spectral ranges at the same time are not commonly used and are currently a popular research topic [2], [4]. The aim of such systems is basically to use or compensate the advantages and disadvantages of the respective spectral ranges. Different camera combinations (with different wavelength ranges) are also possible, e.g.:

- 2 cameras with different spectral ranges each
- 2 cameras with the same spectral ranges
- 2 complete stereo systems each with different spectral ranges (i.e. 2+2 cameras)

The general goal of these works is to map thermal LWIR data on to VIS images. Hung et al. have used two stereo setups, one utilising the visible light spectrum, the other utilising the thermal spectrum, to extract a spatial region of interest from the thermal stereo data, in order to filter visible light stereo data [10]. Stojcsics et al. have proposed a novel method to improve the resolution of cheap thermal cameras using a single visible light camera, to improve the mapping of thermal data to spatial data obtained using photogrammetry for outdoor scenes [24]. Jung et al. have also used a single thermal and a single visible light camera to map thermal data to the data of the visible light camera, although in this case only in 2D [12]. Schramm et al. have proposed a setup utilising 3 spectra: visible light, short wave infrared (SWIR) and LWIR. It consists of a visible light stereo setup, a SWIR structured light depth sensor and a single thermal camera. This approach was chosen to create a robust thermography system, capable of mapping thermal data to 3D objects and spaces [21].

Another application was presented recently by Mouats et al. They propose a stereo setup using a single visible light and a single thermal camera, to allow for egomotion estimation of ground vehicles [15]. The results of this work seem promising (especially with respect to mobile robotics applications), but clear drawbacks are also apparent - in particular, the thermal camera generally extracts far fewer points for stereo matching than its visible light counterpart

(due to less textured image structures). Conversely, the VIS camera appears to be more susceptible to glare.

Infrared spectrum stereo vision system (IR)

In contrast to visible spectrum vision systems and combined approaches, not much work seems to have been done in stereo vision using solely thermal cameras. An attempt to put this into practice was made by Kim et al. in [13], and while they managed to implement a simple calibration method and the detection of fire and humans, they did not manage to extract any depth information using a thermal stereo setup. Mouats et al. in [14] have, however, managed to implement such a system for Unmanned Aerial Vehicles (UAVs). They have proven that visual odometry, 3D reconstruction and egomotion estimation using only thermal cameras is possible, with an accuracy comparable to a setup using only the visible light spectrum. The scenarios they tested their methods were limited to urban outdoor areas, in which their setup worked well even under adverse weather conditions and at night.

Calibration

One thing all stereo vision systems have in common, regardless of their used spectra, is that the cameras have to be calibrated. This calibration process corrects lens errors and is used to geometrically determine the relative camera poses to enable correct depth measurements. Practically, a calibration plate with an imprinted pattern consisting of regularly arranged structures (such as a checkerboard pattern or circles) is used for this purpose. While this process is relatively simple in the visible range, visualizing such patterns in infrared light is somewhat more difficult. A few works simply tried printing out a paper calibration pattern and heating this up with a heat lamp. In theory, the black parts of the pattern should radiate more heat than the white spaces in between, allowing a standard calibration procedure. They have concluded, however, that this method is too inaccurate or too unreliable to be used [27], [14]. The work of Saponaro et al. has proven, however, that this approach is possible with the proper processing of the images [20]. In summary, it takes considerable effort to implement such a method, and it seems easier to create calibration plates specifically adapted for this use case, which can subsequently be used in a more practical way.

Most works have utilised one of two approaches to design a special calibration target for thermal cameras. The first one involves putting a pattern on top of a reflective metal surface. This approach has the advantage that it can be used for all visible light and infrared spectra, but it is rather difficult to create an object like this with the necessary accuracy [14], [15]. The second approach generally involves recreating a common pattern with active heating elements. This can be done either by inserting active heating elements in an existing target [13], which does not promise great accuracy, or cutting out a plate with the calibration pattern and mounting a large, active heating or cooling element on the back [14], [10], [12]. In our opinion, this seems to be the most promising approach, as it is the easiest to put into practice while providing very accurate

results.

Basically, the solutions presented are based on proven concepts of camera calibration, especially since available calibration software (e.g. OpenCV or Matlab) was used. However, there are exceptions, especially for specific applications. Yu et al. have proposed a novel calibration target in their work [27]. It consists of a large cross-shaped structure with active heating elements mounted in defined positions. This approach is well suited for outdoor applications at large distances, e.g. for UAVs or autonomous ground vehicles, as it is easier to build such large calibration object than a traditional target of the same size [27].

Overview of the processing steps applied

Generally, a stereo algorithm takes two raw images and calculates a 3D map of the scene. In order to extract real world measurements from the images, the stereo camera setup must be calibrated first. This determines the intrinsic and extrinsic camera parameters, which in turn are needed for the rectification of the images and the stereo matching algorithm. The output of this system is expected to be a type of 3D map with a minimum accuracy of 5 cm at an operating distance of 1.5 m that can then be used by other algorithms, e.g. SLAM or navigation algorithms. A practical restriction in terms of hardware is the limited space available on the robot which limits the size of the camera to 20 x 10 x 10 cm. Another aspect is a necessary video real-time capability with a minimum frequency of 0.5 Hz and an easy integration into existing ROS environments (Robot Operating System).

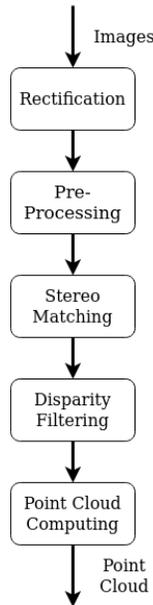


Figure 2: Image Pipeline.

Rectification

The images are rectified and undistorted using the parameters found during the calibration procedure. The camera matrix is one of the results of the camera calibration

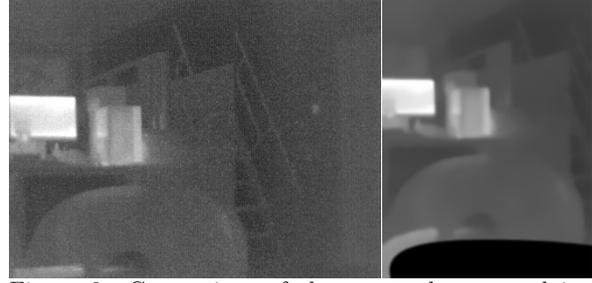


Figure 3: Comparison of the raw and processed image: (left) the raw, 8bit image and (right) the rectified and processed image:

and it is defined in (1), where (f_x, f_y) is the focal length and (c_x, c_y) is the optical centre.

$$\text{camera matrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

The values $(k_1, k_2, p_1, p_2, k_3, k_4, k_5, k_6, s_1, s_2, s_3, s_4, \tau_x, \tau_y)$ describe the distortion coefficients, which are also derived from the camera calibration. The matrix R describes a rectification transformation in the object space and the vector (u, v) describes the pixel position in the corrected and rectified image.

$$\begin{aligned} x &\leftarrow \frac{(u-c'_x)}{f'} \\ y &\leftarrow \frac{(v-c'_y)}{f'} \\ [X \ Y \ W]^T &\leftarrow R^{-1} * [x \ y \ 1]^T \\ x' &\leftarrow X/W \\ y' &\leftarrow Y/W \\ r^2 &\leftarrow x'^2 + y'^2 \\ x'' &\leftarrow x' \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6} + 2p_1x'y' + p_2(r^2 + 2x'^2) + s_1r^2 + s_2r^4 \\ y'' &\leftarrow y' \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6} + p_1(r^2 + 2y'^2) + 2p_2x'y' + s_3r^2 + s_4r^4 \\ s \begin{bmatrix} x''' \\ y''' \\ 1 \end{bmatrix} &= \begin{bmatrix} R_{33}(\tau_x, \tau_y) & 0 & -R_{13}(\tau_x, \tau_y) \\ 0 & R_{33}(\tau_x, \tau_y) & -R_{23}(\tau_x, \tau_y) \\ 0 & 0 & 1 \end{bmatrix} R(\tau_x, \tau_y) \begin{bmatrix} x'' \\ y'' \\ 1 \end{bmatrix} \\ \text{map}_x(u, v) &\leftarrow x''' f_x + c_x \\ \text{map}_y(u, v) &\leftarrow y''' f_y + c_y \end{aligned} \quad (2)$$

Equations (2) describe the rectification process. The resulting $\text{map}_x(u, v)$ and $\text{map}_y(u, v)$ are then used as transformation maps to correct and rectify the image using an image transformation with linear interpolation [19].

Pre-processing

The rectified images are then filtered using a median filter to remove impulse noise, and enhanced by using the non-local-means denoising algorithm proposed in [1] and finally filtered again using a bilateral filter to further improve the image quality. The result of these processing steps is shown as an example in Figure 3.

Stereo Matching

The left/right image pair is then used to calculate a so called disparity map using the global matching algorithm proposed in [9].

Disparity Filtering

The disparity map is then filtered using a weighted least squared filter, in the form of a global fast smoother to further improve performance, as described in [18].

Point Cloud Computing

The resulting disparity map is then transformed into a point cloud, which is the standard data representation in robotics (e.g., for visualization, localization, pose estimation, mapping, etc.). OpenCVs implementation is used for this purpose, based on the formula given in (3). The vector $[X, Y, Z, W]$ describes the point cloud, while the vector $[x, y, disparity(x, y), z]$ describes the disparity image, and Q is a 4x4 matrix describing the disparity-to-depth mapping matrix, which is found during the calibration process [17].

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = Q \begin{bmatrix} x \\ y \\ disparity(x, y) \\ z \end{bmatrix} \quad (3)$$

Experimental Results

Flat Field Calibration

Uncooled infrared cameras typically require corrections due to heating of the camera itself to maintain constant image quality over time. In this "Flat Field Calibration" (FFC), a shutter of uniform temperature is applied to the detector, but this results in a regular interruption of the image acquisition. [14]. This was shown to be an issue for real-time algorithms for localisation and pose estimation [14]. For the application described here, however, this effect can be compensated for because each image can be time-stamped (as long as the system is not moving too fast) and the shutter speed was fast enough to be neglected.

Monocular and Stereo Calibration

When using a camera in the visible light spectrum, a black calibration pattern printed on a white sheet of paper is usually sufficient as a calibration target. In the LWIR spectrum, this is not sufficient, as several authors have shown [14], [27], at least without extensive processing as e.g., in [20]. For the application of open source calibration algorithms, a standard calibration pattern must also be used. Therefore, the method proposed in [14] was chosen instead of the very specific setup described in [27]. Here, the calibration pattern is cut into a plate and this plate is placed over an active heat source, e.g., a heating mat. In this way, the LWIR radiation passes through the holes of the calibration pattern and can be clearly seen by the cameras.

Figure 4 shows the comparison between the heated calibration pattern as viewed through a thermal camera (left) and the result of the calibration process visualised as

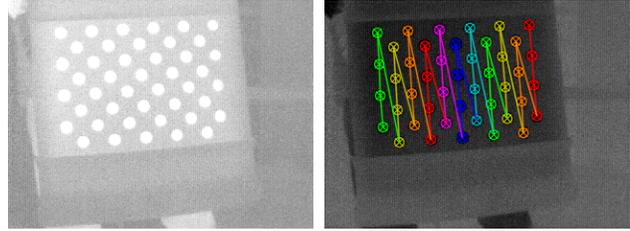


Figure 4: Comparison of the calibration pattern and the result of the calibration process: (left) Image of the heated calibration pattern and (right) image depicting the result of the calibration process.

overlay on the input image (right). The pattern, in this case an asymmetrical circle grid, can easily be extracted from images like this using simple, existing algorithms, for both monocular and stereo camera setups. The detected pattern and the real pattern align almost perfectly, indicating that the calculated intrinsic and extrinsic camera parameters are accurate. The average RMS re-projection error lies between 0.15 and 0.20 for monocular calibration and between 0.40 and 0.50 for the stereo calibration. Thus, both results are within a satisfactory range of values.

Hardware Solution

Somewhat similar systems already exist and are in use e.g., by firefighters in buildings that are filled with thick smoke. These handheld devices are called "Thermal Imaging Cameras" (TICs), that show an image often composed of a number of different spectral ranges (sometimes referred to as multi-spectral devices). A LWIR sensor seems to always be present, since this spectral range penetrates smoke well [25]. LWIR sensors have become reasonably cheap in recent years, with a considerable amount of systems being developed for smartphones and personal use. Therefore, the proposed hardware setup consists of two commercially available "CompactPRO" LWIR cameras by "Seek thermal", set up as a stereo camera pair with a baseline of about 10 cm, see Figure 5. This results in a measurable depth difference between 0.34cm and 3.02cm, at operating distances between 0.5m and 1.5m, respectively (according to (4)). In the following equation ΔH describes the measurable depth difference, H describes the operating dis-

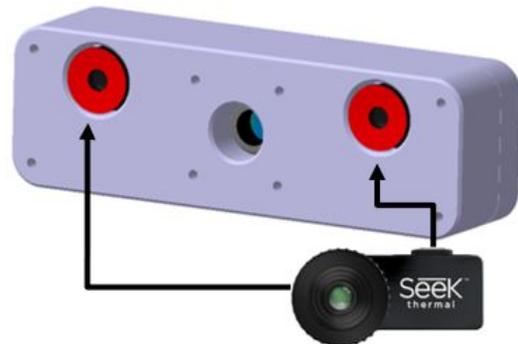


Figure 5: CAD model of the stereo thermal camera with an additional RGB camera in the center.

tance, P_x describes the pixel side length, f describes the focal length and b describes the baseline.

$$\Delta H = \frac{P_x * H^2}{f * b} \quad (4)$$

The sensors are connected via USB to a "Jetson Nano" by "NVIDIA" on which the image processing algorithms are executed, and which in turn is connected to the robot via Ethernet.

Data Gathering and Visualization

For test purposes, a program periodically took frames out of a pre-recorded video and fed them to the algorithm. At first, experiments were conducted on a desktop computer (to show the general effectiveness of the approach). The results were then saved to the hard disk. Finally, the raw images, the processed images, the raw disparity, the filtered disparity and the point cloud derived from the filtered disparity images could be visualized.

Test Scenarios

To further test the effectiveness of the algorithm, this chapter considers in detail three common scenarios that an autonomous rescue robot is likely to encounter. The first scenario shows an environment with no active heat sources, where only the floor, walls, and objects to be avoided are visible. The second scenario additionally shows active heat sources and the third scenario furthermore shows a working person.

Scenario #1 - Scenes with no heat sources

The scene shows a robot lying on the floor with four fins and an outstretched arm with a hot module at its base. Above the robot, the lower part of a door and a sink can be seen. Figure 6 Fig. (x) shows the raw images from the left and right cameras, which are still noisy, and the two processed images, which now look much better. However, it seems that this scene is even less suitable for the stereo algorithm, since the processed disparity map shows nothing but the hot spot of the active module.

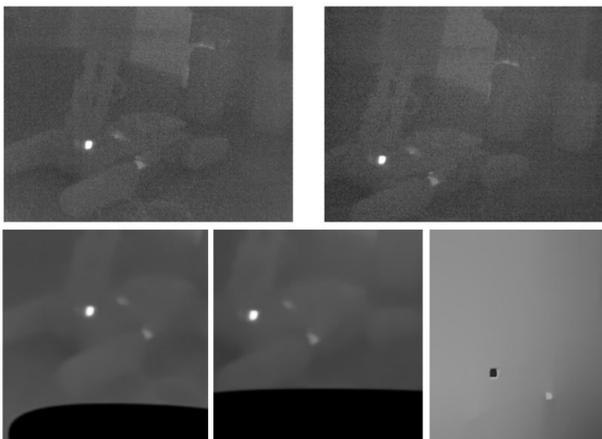


Figure 6: Comparison of the raw, processed data and the disparity map of the first scenario.

Scenario #2 - Scenes with active heat sources

In this scenario only one scene is examined. It shows a desktop PC running and a monitor on a desk with a chair in front of it and stairs to the right. The results of the second scenario as shown in Figure 7 which seem to be more promising than the results of the first scenario. Now, the disparity map shows more information than in the previous example in Figure 6. The shape of the PC, the monitor, the chair and even parts of the stairs can be seen. However, the distances are completely wrong. For example, the monitor shows a clear distance gradient across its screen, although it should be virtually flat. Also, while the dark lines at the bottom and right edges of the monitor reasonably accurately reflect the distance of the background from the monitor, they should extend all the way to the PC and also appear above and to the left of the monitor, since the background around the monitor is the same distance. The PC itself is divided into parts of very different distances, with the top part of the PC being the furthest object according to the disparity map, although it should be the same distance as the rest of the PC and the monitor. The closest object according to the disparity map is the monitor's screen at a distance of, again, about 0.35m. In reality, the monitor is about 2 m away, and the closest object is the chair at the very bottom of the images, which is about 1 m away.

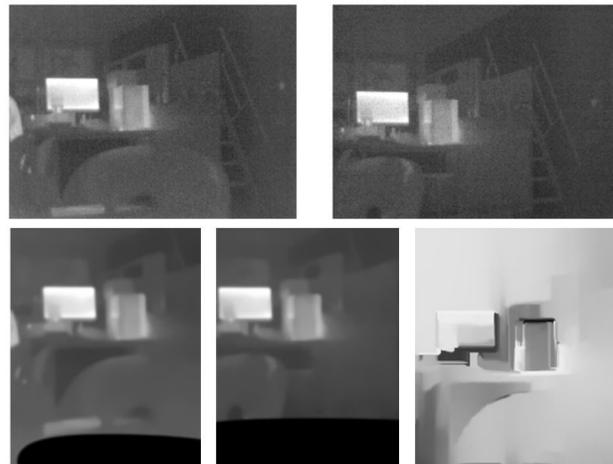


Figure 7: Comparison of the raw, processed data and the disparity map of the second scenario.

Scenario #3 - Scenes with a working person

This scenario, a person is shown working at a table. To the right of this person is the monitor of the scenario shown in Figure 8. In front of the person, a set of screwdrivers is placed on the table and a chair covers part of the person's body. The results are similar to Figure 6 and Figure 7. The structures of the objects and the person in the scene can be seen in the disparity map, even relatively accurately, but individual objects are divided into different distances, all of which are incorrect. For example, the person's head is shown detached from the body and at the same distance as the monitor at about 0.35 m, even though the person is about 1 m away and the monitor is 2 m away.



Figure 8: Comparison of the raw and processed data of the third scenario.

Point Cloud Visualization

The resulting disparity map is then transformed into a point cloud which is the final output of the algorithm. Figure 9 shows the person working from Figure 8 but from a different camera position and the according final point cloud visualized with the visualization tool *rviz* in ROS.

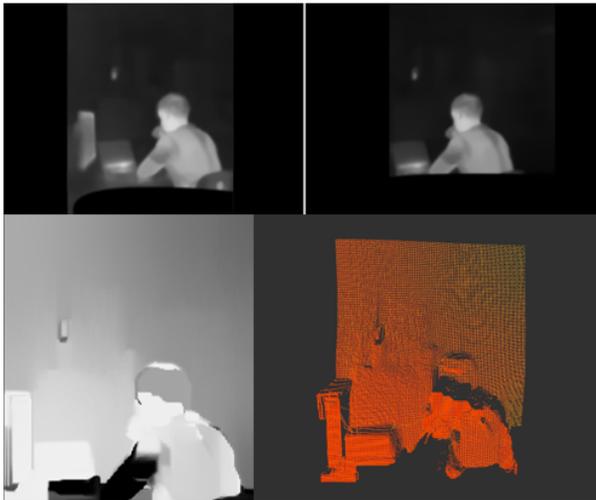


Figure 9: Visualization of 3D point cloud.

Conclusion and Future Work

The results of the processing pipeline and the first prototype have not yet provided the expected results. The main reason for this is that interior scenes in particular generally do not provide a distinct thermal "texture" for the stereo algorithm to work with. The results show that the system currently has problems with small temperature differences, which as a result do not provide an accurate disparity map and ultimately a 3D point cloud of the environment. Without textures, the block matching of the stereo algorithm cannot work properly. This could be improved by using higher quality thermal cameras, since the low image quality of infrared cameras, currently 320x240

pixels, and subsequent processing further reduce the visible textures and gradients in the image. Although the method proposed in this work does not currently meet the requirements, it still offers the prospect of further developing this initial design and using it in other scenarios.

References

- [1] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. Non-Local Means Denoising. *Image Processing On Line*, 1:208–212, 2011. https://doi.org/10.5201/ipol.2011.bcm_nlm.
- [2] Yong-Jun Chang, Byung-Geun Lee, and Moongu Jeon. Warping-based spectral translation network for unsupervised cross-spectral stereo matching. *Information Sciences*, 588:214–230, 2022.
- [3] K.S. Chidanand Kumar. Stereo-vision based smart tv control. In *2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS)*, pages 67–71, 2015.
- [4] Weichen Dai, Yu Zhang, Donglei Sun, Naira Hovakimyan, and Ping Li. Multi-spectral visual odometry without explicit stereo matching. In *2019 International Conference on 3D Vision (3DV)*, pages 443–452. IEEE, 2019.
- [5] Raimund Edlinger, Roman Franz Froschauer, and Andreas Nüchter. Robuste Algorithmen zur Situationsanalyse von flexiblen und mobilen Roboterassistenten. Available at <http://www.fh-ooe.at/comingsoon> (19/09/2019), September 2019. Coming Soon Conference.
- [6] Xinjian Fan, Xuelin Wang, and Yongfei Xiao. A shape-based stereo matching algorithm for binocular vision. In *Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pages 70–74, 2014.
- [7] Saeid Fazli, Hajar Mohammadi Dehnavi, and Payman Moallem. A robust obstacle detection method in highly textured environments using stereo vision. In *2009 Second International Conference on Machine Vision*, pages 97–100, 2009.
- [8] Rostam Affendi Hamzah and Haidi Ibrahim. Literature survey on stereo vision disparity map algorithms. *Journal of Sensors*, 2016, 2016.
- [9] Heiko Hirschmüller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, 2008.
- [10] P. Hung, K. Chen, H. Hsu, and R. Wang. Integration of autofocus and object tracking in an infrared stereo vision-based video surveillance system with multi-lens module. In *2018 IEEE International Conference on Applied System Invention (ICASI)*, pages 1280–1283, 2018.
- [11] Intel Corporation. Intel® RealSense™ Technology. Available at <https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html> (22/06/2021).
- [12] H. Jung and J. Lyou. Matching of thermal and color images with application to power distribution line fault detection. In *2015 15th International Conference on Control, Automation and Systems (ICCAS)*, pages 1389–1392, 2015.
- [13] S. Kim, S. Jun, and J. Park. Thermal stereo system for

visible range extension of disaster robot. In *2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 1–2, 2018.

- [14] T. Mouats, N. Aouf, L. Chermak, and M. A. Richardson. Thermal stereo odometry for uavs. *IEEE Sensors Journal*, 15(11):6335–6347, 2015.
- [15] T. Mouats, N. Aouf, A. D. Sappa, C. Aguilera, and R. Toledo. Multispectral stereo odometry. *IEEE Transactions on Intelligent Transportation Systems*, 16(3):1210–1224, 2015.
- [16] Nerian Vision Technology. HKarmin3 3D Stereo Camera. Available at <https://nerian.com/products/karmin3-3d-stereo-camera/> (22/06/2021).
- [17] OpenCV Team. OpenCV: Camera Calibration and 3D Reconstruction. Available at https://docs.opencv.org/3.4/d9/d0c/group_calib3d.html (22/06/2021).
- [18] OpenCV Team. OpenCV: cv::ximgproc::DisparityWLSFilter Class Reference. Available at https://docs.opencv.org/4.x/d9/d51/classcv_1_1ximgproc_1_1DisparityWLSFilter.html (22/06/2021).
- [19] OpenCV Team. OpenCV: Geometric Image Transformations. Available at https://docs.opencv.org/3.4/da/d54/group_imgproc_transform.html (22/06/2021).
- [20] P. Saponaro, S. Sorensen, S. Rhein, and C. Kambhamettu. Improving calibration of thermal stereo cameras using heated calibration board. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4718–4722, 2015.
- [21] S. Schramm, J. Rangel, and A. Kroll. Data fusion for 3d thermal imaging using depth and stereo camera for robust self-localization. In *2018 IEEE Sensors Applications Symposium (SAS)*, pages 1–6, 2018.
- [22] Arvind Sharma, Shraddha Chaudhary, Sumantra Dutta Roy, and Prakash Chand. Three dimensional reconstruction of cylindrical pellet using stereo vision. In *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pages 1–4, 2015.
- [23] Stereolabs Inc. Stereolabs - Capture the World in 3D. Available at <https://www.stereolabs.com/> (22/06/2021).
- [24] D. Stojcsics, I. Lovas, Z. Domozi, and A. Molnar. High resolution 3d thermal imaging using flir duo r sensor. In *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, pages 000311–000316, 2018.
- [25] Teledyne FLIR LLC. Handheld & Aerial Thermal Imaging Cameras for Firefighting . Available at <https://www.flir.eu/instruments/firefighting/> (22/06/2021).
- [26] Teledyne FLIR LLC. Stereo Vision. Available at <https://www.flir.com/iis/machine-vision/stereo-vision/> (22/06/2021).
- [27] Z. Yu, S. Lincheng, Z. Dianle, Z. Daibing, and Y. Chengping. Camera calibration of thermal-infrared stereo vision system. In *2013 Fourth International Conference on Intelligent Systems Design and Engineering Applications*, pages 197–201, 2013.

Author Biography

Raimund Edlinger is an assistant professor at the University of Applied Sciences Upper Austria. He received his

DI(FH) in sensors and microsystems (2007) and his MSc. in Automation Engineering (2013) from the University of Applied Sciences Upper Austria. Since 2007 he has worked in the Research and Technology at the University in Wels/Austria. His work has focused on the development of mobile robots and sensor systems. He is a IEEE Member and on the board of RoboCup Rescue League as technical member and since 2018 Phd student at Graduate School Science and Technology at University of Würzburg.

Gabriel Himmelbauer is a researcher at the University of Applied Sciences Upper Austria, where he received his BSc. (2020) and DI (2022). His work has focused on the application of sensor technologies and the automation of difficult, mundane or dangerous tasks.

Gerald Zauner received his PhD in applied physics from Vienna University of Technology (2005). Since then he has worked at the University of Applied Sciences Upper Austria (School of Engineering) where he is professor for signal processing (2015). His work has focused on industrial machine vision, optical metrology and non-destructive testing.

Andreas Nüchter is professor of computer science (robotics) at University of Würzburg. He holds a doctorate degree (Dr. rer. nat) from University of Bonn. His thesis was shortlisted for the EURON PhD award. Andreas works on robotics and automation, cognitive systems and artificial intelligence. His main research interests include reliable robot control, 3D environment mapping, 3D vision, and laser scanning technologies, resulting in fast 3D scan matching algorithms that enable robots to perceive and map their environment in 3D representing the pose with 6 degrees of freedom. The capabilities of these robotic SLAM approaches were demonstrated at RoboCup Rescue competitions, ELROB and several other events. He is a member of the GI and the IEEE.