AUTOMATIC CALIBRATION AND CO-REGISTRATION FOR A STEREO CAMERA SYSTEM AND A THERMAL IMAGING SENSOR USING A CHESSBOARD

A. Choinowski[a], D. Dahlke[a]*, I. Ernst[a], S. Pless[a], I. Rettig[a]

[a] Institute of Optical Sensor Systems, German Aerospace Center, 12489 Berlin, Germany -
(Andre.Choinowski, Dennis.Dahlke, Ines.Ernst, Sebastian.Pless, Immanuel.Rettig)@dlr.de

Commission I WG 9

KEY WORDS: Thermal, VIS, Chessboard, Co-Registration, Realtime, System calibration

ABSTRACT:

This paper presents a system calibration method for a trifocal sensor, which is sensitive to different spectral bands. The trifocal camera system consists of a stereo camera, operating in the visual (VIS) spectrum and a thermal imaging camera, operating in the Long-Wave-Infrared (LWIR) spectrum. Intrinsic parameters and spatial alignment are determined simultaneously. As calibration target a passive aluminium chessboard is used. Corner detection and subsequent bundle adjustment is done on all synchronized image triplets. The remaining reprojection errors are in the sub-pixel range and enable the system to generate metric point clouds, colored with thermal intensities in real-time.

1. INTRODUCTION

In many industrial inspection applications, thermal images are used to detect temperature anomalies on surfaces. For mapping and documentary reasons it is beneficial to locate such anomalous objects in 3D space. Basically this can be done by stereo matching and subsequent projection of thermal intensities into the point cloud. Beforehand an accurate geometric calibration of the trifocal sensor is crucial. The calibration step becomes even more challenging when the sensors resolve in different electromagnetic spectral ranges, like visual and thermal in our case.

Benefits of fusing thermal imaging data with VIS/RGB data are among others an extended range of operation, increased reliability and compact representation of information. This paper proposes a method to integrate thermal imaging capabilities into the existing IPS (Integrated Positioning System).

IPS (see fig. 1) is a real-time hardware and software solution to determine self-location and orientation without any kind of infrastructure like global navigation satellite systems (GPS, Galileo, etc.) or Wi-Fi (Börner et al., 2018). The system is used to perform different types of inspection tasks in difficult environments like mines, shafts and other tunnel systems, ships and buildings. The IPS navigation-relevant sensors consist of a monochromatic stereo-camera system and an IMU (inertial measurement unit).

The camera images of the sensor head can also be used to generate an accurate 3D point cloud in real-time. Additional sensors like barometer, air mass or wall thickness meter can be included to deliver more relevant information. For instance a RGB camera can be used to generate a coloured point cloud.

This paper describes the integration of a thermal imaging sensor. Practically an extension in the spectral range of LWIR leads to many more use cases in the aforementioned areas of application. Common IPS industrial inspection tasks like monitoring ship hulls or detecting water intrusion in mining shafts can greatly benefit from this additional spectral information.

---
*Corresponding author

Figure 1. IPS Camera System with Optris PI 450 attached on top. The system is mounted on a rotary plate for calibration purposes.

A precursor for such a sensor integration is an accurate geometric camera system calibration. Calibration in general can be defined as a "process of quantitatively defining the system responses to known, controlled signal inputs" (Morain and Zanoni, 2004). Geometric calibration in particular can be used to spatially undistort and rectify imagery, as well as determine the relative poses of the cameras. Basically a thermal imaging sensor can be geometrically calibrated like a conventional camera. Suitable for this task are test fields with control points emitting in the thermal range (e.g. active bulbs) or by reflecting ambient radiation (Luhmann et al., 2013). However, calibration in the mid to long wave infrared poses several challenges. One is the usually low number of image pixels which demands a good fit of target size and camera field of view. Another challenge is the generation of good contrast in-between the elements of the target. Additionally visual cameras have to recognize the target as well for the alignment of both sensor types.

2. RELATED WORK

There are two main strategies to build a target usable for thermal infrared camera calibration: Active targets which require partial heating or cooling for the generation of contrast have been employed for such calibrations before (Rankin et al., 2011), (Vidas et al., 2012), (Ursine et al., 2012), (Yu et al., 2013), (Zoetgnande et al., 2018), (Chen et al., 2019). Reflective targets using different emissivity surfaces are also possible, but require different temperatures for the target itself and an illumination source (Shibata et al. (2017) and Aguilera et al. (2018)). An extensive evaluation of best suitable material and coating for such a passive board can be found in St-Laurent et al. (2017). However in both cases, visually perceptive contrast between the elements used for the target composition is needed for simultaneous acquisition of VIS/RGB images.

Active targets require thermal insulation between the elements of the target but still may lead to heat transfer to surrounding objects. Also air turbulences may occur and influence the higher resolution imaging of the VIS camera. Similar problems may arise from the use of a thermal illumination source. Furthermore, taking the costs, required equipment and preparation associated with these more complex targets into account, we were motivated to use an existing aluminium chessboard. It has proven to work very well in the VIS spectrum. Its applicability for LWIR sensors and inherent drawbacks will be outlined in Section 3. The key contribution is the precise spatial alignment of a stereo camera system and a thermal imaging sensor with a well-known and inexpensive calibration pattern, finally enabling real-time thermal point cloud generation.

When referring to image sensor calibration, the process of radiometric calibration has to be mentioned as well. Although this task is of special importance in the LWIR spectrum, it has no significant influence on the presented approach and is therefore not part of this paper. A method for performing radiometric calibration of thermal imaging sensors can be found in Kattnig et al. (2015). By using blackbody radiation with known temperatures, the relationship between radiometric intensity level and LWIR radiance could be determined. Since the LWIR radiance is compounded of surface temperature, surface emissivity and reflectance, the material properties have to be taken into account.

3. SENSOR AND CHESSBOARD DESCRIPTION

Cameras , which are used for the presented approach, are briefly described in the following. Example images for both camera types (thermal imaging sensor and panchromatic sensors) can be seen in Figure 2. The different properties in terms of field of view or feature content become clearly visible. While the tripod and the trifocal sensor are mirrored in the thermal image (see fig. 2(b)), they are not visible for the panchromatic sensor (see fig. 2(a)). An overview of the key specifications of both sensor types can be found in Table 1.

The IPS stereo-camera consists of robust industrial-grade cameras with a global shutter mode. The picture release and the synchronisation to the IMU and other sensors are handled by a FPGA. The data grabbing and if necessary navigation and point cloud calculation is done by a PC and the IPS application.

The thermal camera was tightened on the IPS sensor head (see fig. 1) and has no possibility for triggering a picture release.
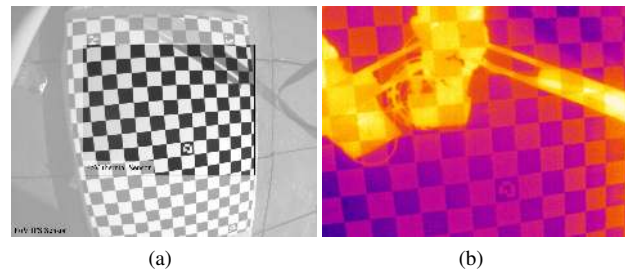

(a)　　　　　　　(b)

Figure 2. Comparison of feature content and field of view (FoV) between VIS and LWIR spectral range: (a) visual image with highlighted area, showing corresponding LWIR FoV (b) colour-coded LWIR image (tripod with camera setup is mirrored on the chessboard surface)

Imaging is done in free-run. In addition the capturing is done in a rolling shutter mode and a radiometric offset calibration is done frequently during the recording session, which in turn makes the camera blind for a short moment. In case of the geometric calibration and some test recordings the Optris PIX Connect software was used. The camera was configured to store a thermal picture triggered by an external push-button release. The signal was also connected to the IPS FPGA, which realizes a rough synchronisation. Due to the rolling shutter, the thermal picture storing was only triggered when the sensor constellation stands fixed on a tripod.

Manufacturer Model	AVT GC-1380H	Optris Pi 450
Camera Description	IPS visual	LWIR
Spectral Sensivity (μm)	0.4-0.9	7.5-13
SNR or NETD	n/a	40 mK
Resolution (px)	1360 × 1024	382 × 288
Dynamic Range	12bit	12bit
Pixel Pitch (μm)	6.45	25
Focal Length (mm)	4.8	10.5
Field of View	85 deg	53 deg
Framerate	10Hz	27/80Hz

Table 1. Key specifications for the trifocal sensor systems

Chessboard calibration is an extensively used method in computer vision. They are relatively easy to produce and a variety of automated processing approaches exist. In order to achieve accurate calibration results, the used chessboard is made of two aluminium plates (120 cm × 84 cm each) attached to an aluminium honeycomb core. On the one hand the material is relatively lightweight (7 kg) and on the other hand a honeycomb core ensures a very good stiffness and rigidity. The flatness of such panels is less than 0.5 mm over a span of 4 m. A chessboard with a raster size of 5 cm × 5 cm is printed on a custom static cling and gets attached to one side of the aluminium board. Due to imperfections from the printing and adhesive process, the real pattern size has to be checked with measurements along each dimension of the printed foil. The resulting scale factors are added to the board definition and are taken into account during the calibration process. Due to the open patches along each border, the 19 × 13 raster yields in 20 × 14 possible corners. Moreover the board is equipped with five markers. These tags assure an automatic corner assignment even if just parts of the board are visible in an image.

4. METHODOLGY

In order to remove distortion effects from optically captured imagery, it is necessary to derive calibration parameters. They are gathered under the term "intrinsic calibration" and consist of focal length, principle point and radially symmetric distortion. If several optical sensors are combined to a stereo or trifocal system, an additional prerequisite is the knowledge about the spatial alignment between all sensors. These parameters are described by a rotation (ω, ϕ, κ) and translation (t_x, t_y, t_z). They are gathered under the term "extrinsic calibration". Capturing several images of a chessboard target is a common method in computer vision to derive intrinsic and extrinsic parameters at once. An automatic detection of the chessboard is implemented in different available approaches like Bradski (2010) or Bouguet (2015). Practically those approaches either "require a great deal of manual processing" (Rankin et al., 2011) or rely on a fully seen chessboard. Both drawbacks are inconvenient, especially because of the low resolution of the LWIR sensor. Low geometric and radiometric resolution at the chessboard edges likely lead to incompletely detected corners and therefore to inconsistencies between thermal and visual imagery. An approach that overcomes these drawbacks and allows the calibration of a trifocal system as well, is presented in Wohlfeil et al. (2019).

Image Acquisition of a calibration target in the visual spectrum is a simple matter. It becomes more difficult when the calibration imagery is also used for calibrating thermal infrared sensors. Features of the chessboard target have to be visible in all synchronously captured calibration images. Without an additional external thermal source, the chessboard cannot be reliably located in the LWIR image. Since the thermal radiation differences in-between black and white patches are too faint, all borders will appear with low contrast in the image. The contrast can be enhanced through heating the black patterns, either with an artificial thermal source or by exposing the chessboard to direct sunlight. Both options have its disadvantages. Either the thermal source lacks the capability to homogeneously heat the chessboard surface or the method relies on optimal weather conditions. In the presented approach we make use of the reflection properties of both, the different inks used for black and white patterns and the underlying aluminium surface. When positioned facing the sky, the thermal gradient between sky and ambient temperature combined with different emissivity of the ink types yields in sufficient contrast of the chessboard pattern (see fig. 2(b)). Due to the aluminium surface the chessboard is comparable to a mirror, reflecting strongly in the LWIR range.

A setup for calibration and spatial alignment of multi camera systems with planar reference targets can be found in Luhmann et al. (2013). Following this setup the chessboard pattern is captured with several observations and rotations under clear sky. Figure 3 shows the prescribed set of 36 positions used for calibration. Note that the different poses present a balanced set, with different rotations and distances to decorrelate intrinsic and extrinsic calibration parameters as good as possible.

Corner Detection and Bundle Adjustment is done on all image triplets at once. For the automated detection of chessboard corners, as well as the following bundle adjustment, a solution presented in Wohlfeil et al. (2019) is used. Basically it works similar to commonly available approaches, but has the aforementioned advantages. Corner points are found in two steps. Firstly, candidates are searched with particular attention to the accuracy of the edges' position. Secondly, these candidates are analysed
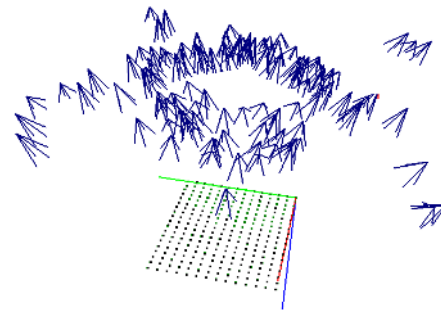

Figure 3. 3D plot of 36 poses used for the calibration and the chessboard corners (each pose is described by 3 cameras)

to be an actual corner point of the chessboard in order to reduce the number of false positives to a minimum. Each corner point is defined relative to the tags with intervals of the corner points (i_x and i_y) and their number (n_x and y_y) in x and y direction (i_x and i_y). Given the print specific scale factors, mentioned in the previous section, the intervals of the corner points are assumed to be identical. After all corner points in all images taken for the calibration have been found and assigned automatically (see fig. 5(a)), the calibration can be performed using a bundle adjustment.

The intrinsic and the extrinsic camera parameters are estimated by minimizing the re-projection error between the metric chessboard reference frame and each corner point coordinate, given in pixel. Since the pinhole model assumes zero distortion in the image, a lens distortion model has to be considered before the pinhole model could be applied. An extension of the model including radially symmetric components is given in the Brown-Conrady model (Brown, 1971).

While each additional calibration pose contributes to the estimation of the camera model, it also adds 6 parameters to the optimization problem. For a trifocal camera system, two more cameras are rigidly mounted w.r.t. the first camera. This is modeled by two additional relative orientations $(\mathbf{R}, \mathbf{t})_{c1}^{c2}$ and $(\mathbf{R}, \mathbf{t})_{c1}^{c3}$. The two step approach firstly estimates intrinsic and extrinsic parameters with a linear optimization (Zhang, 2000). Given these initial values, the non-linear least squares problem of the distortion model is now solved with the Gauss-Newton algorithm.

5. RESULTS

21 783 corner points have been detected on 36 image triplets. Each triplet and pose is estimated with an average of 605 corner points. This in turn means on each image roughly 200 points have been detected. An investigation on the robustness of the derived calibration parameters is done by repetitively using random samples from the complete set of 36 image triplets. 10, 20 and 30 triplets have been randomly chosen 1000 times. Each calibration parameter is hereby estimated 1000 times for the 3 sets. Figure 4(a) and fig. 4(b) shows the outcome exemplary for the focal length of the thermal imaging sensor. The maximum of a Gauss fitting function is used as value for the calibration parameter. While this maximum value is shifting by 0.5 px between the two sets, the corresponding standard deviation is significantly decreasing with 3 times more images. In figures this means from 3.66 px (for sets with 10 images) to 0.8 px (for sets with 30 images).
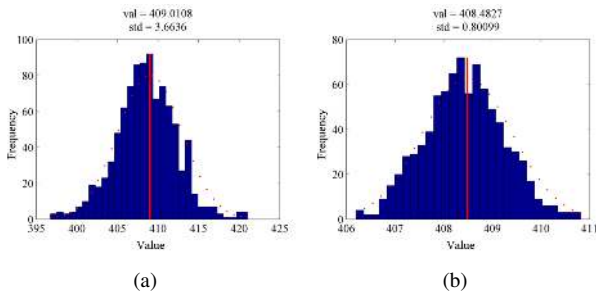
(a)   (b)

Figure 4. Fitting a gaussian function to 1000 adjustment results of the focal length (thermal imaging sensor). Maximum of the gaussian function is used as value. (a) Value derived from 10 random triplets. (b) Value with lower standard deviation derived from 30 random triplets.

One advantage of the used bundle adjustment is to estimate intrinsic and extrinsic calibration for all sensors at once (aio). Nevertheless it is also possible to perform the adjustment separately (sep). Hereby the intrinsic parameters are estimated for the 3 cameras. Subsequently and with fixed intrinsic parameters the extrinsic calibration is estimated. Both approaches are compared against each other by using the aforementioned random sample method. The resulting values and standard deviations for focal length and principle point are compared in table 2.

	IPS left					
	c_k		x_0		y_0	
	value	stdv	value	stdv	value	stdv
aio_{10}	776.7	0.96	711.3	0.78	546.4	0.79
aio_{20}	776.6	0.55	711.2	0.44	546.3	0.44
aio_{30}	776.6	0.28	711.2	0.24	546.3	0.24
sep_{10}	776.7	0.46	711.0	0.68	546.1	0.69
sep_{20}	776.7	0.27	710.9	0.39	546.0	0.38
sep_{30}	776.7	0.15	710.9	0.21	546.0	0.20

	IPS right					
	c_k		x_0		y_0	
	value	stdv	value	stdv	value	stdv
aio_{10}	773.7	1.06	681.6	0.80	540.8	0.79
aio_{20}	773.5	0.62	681.5	0.45	540.7	0.44
aio_{30}	773.5	0.32	681.5	0.23	540.7	0.24
sep_{10}	774.0	0.57	681.8	0.72	540.6	0.65
sep_{20}	773.8	0.31	681.7	0.40	540.5	0.36
sep_{30}	773.8	0.18	681.7	0.22	540.5	0.20

	Optris Pi 450					
	c_k		x_0		y_0	
	value	stdv	value	stdv	value	stdv
aio_{10}	409.0	3.66	188.4	4.19	146.6	3.19
aio_{20}	408.5	1.64	188.1	2.16	146.3	1.67
aio_{30}	408.5	0.80	188.1	1.11	146.2	0.85
sep_{10}	409.9	10.44	193.3	4.32	144.8	6.11
sep_{20}	409.6	5.55	193.4	2.32	145.3	3.14
sep_{30}	409.9	2.77	193.3	1.17	144.8	1.66

Table 2. Comparison for an all-in-one (aio) or seperate (sep) bundle adjustment with 10, 20 and 30 random samples. On top and in the middle are intrinsic parameters of the visual IPS stereo cameras and at the bottom the values for the thermal imaging sensor in [px].

The parameters for the LWIR sensor benefit from the aoi solution. Especially the standard deviations for the focal length suffer from sep bundle adjustment. This could mean that the pose estimation becomes much more accurate when all sensors are calibrated at once. For the thermal sensor this means, the position of
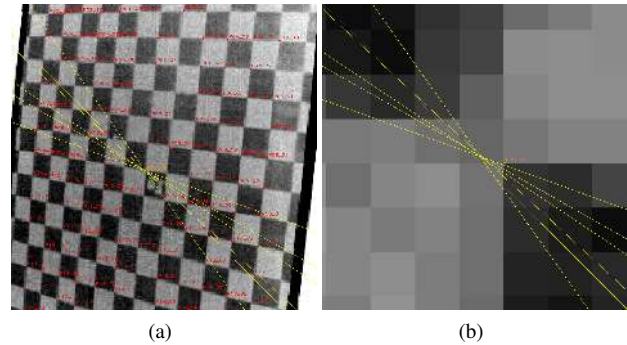

(a)   (b)

Figure 5. Epipolar geometry for one corner point seen in 7 images (a) LWIR image with detected chessboard corners (b) magnification for one point with sub-pixel accurate epipolar rays

the chessboard is much more reliable and therefore the intrinsic parameters are estimated with more accuracy. On the other hand it can be seen that the intrinsic parameters for the visual stereo system suffers from the aio solution. Which makes sense, since the low resolution data of the thermal imaging sensor is used as well for the pose estimation. However the values for the visual cameras are almost identical between the aio and sep solution.

The radially symmetric distortion parameters as well as the extrinsic calibration are derived in the same way as the focal length in fig. 4(b). Since the overall influence of the aio solution leads to more accurate calibration results, it is used for the intrinsic and extrinsic calibration (see table 3). Furthermore this guarantees consistency in-between the derived parameters.

	IPS left	IPS right	Optris PI 450
c_k	776.6	773.5	408.5
x_0	711.2	681.5	188.1
y_0	546.3	540.7	146.2
k_1	−0.273	−0.257	−0.187
k_2	0.168	0.118	−0.008
k_3	−0.069	−0.029	0.434
t_x	-	−20.162	−0.314
t_y	-	−0.036	7.968
t_z	-	0.060	3.368
ω	-	0.0087	−0.0183
ϕ	-	0.0077	0.0231
κ	-	0.0048	0.0009

Table 3. Calibration Results for the trifocal sensor (intrinsic calibration is given in px and extrinsic calibration is given in cm or in rad)

Concerning the accuracy, with 30 image triplets, the intrinsic and extrinsic calibration parameters were estimated precisely enough to perform sub-pixel matching between the sensor image data. The accuracy of the calibration model is measured by the standard deviation of the residuals (RMS). For a set with 30 image triplets and roughly $18\,000$ corner points, the RMS seems to be stabilized at $0.293\,\text{px}$. Another visual interpretation for the accuracy of the calibration model is the epipolar geometry, which is a prerequisite for subsequent matching tasks. From fig. 5(b) it can be seen that the epipolar error is in the sub-pixel range. Furthermore the standard deviation of the focal length (for the LWIR sensor) of $0.8\,\text{px}$ is slightly better than Vidas et al. (2012) approach with masked backdrops and far better than their test results with heated chessboards (standard deviation of $14\,\text{px}$).

6. CONCLUSION AND OUTLOOK

We proposed an effective way of estimating the intrinsic and extrinsic calibration parameters for three cameras at once, including a thermal imaging sensor and two panchromatic sensors. In order to solve this problem, we proposed a calibration grid composed of a chessboard raster placed on an aluminium board. In a second step, we performed a corner detection on the visual and thermal image triplets with sub-pixel accuracy. Subsequently all alignment and intrinsic orientation parameters where derived through a bundle adjustment.

Through the random sample method, we can determine how the values converge to a certain value. We use the maximum of a fitted gaussian function to determine the intrinsic and extrinsic calibration values. However, we can see that if N is less than 30 image triplets, the standard deviation of the calibration results become very high. It can be stated, that 30 triplets or more lead to sufficiently accurate and robust results. Furthermore we have shown the influence of the image quality to the calibration results. A better resolution, estimated from the stereo camera system, can be used to improve the calibration parameters of the thermal imaging sensor.

It is planned to equip the Optris Pi 450 sensor with another optic, which has a FoV of $80°$. The advantage is a comparable FoV for all sensors. The drawback is another loss of spatial resolution when capturing the imagery from the proposed distance of roughly $1.5\,\mathrm{m}$. In order to overcome the resolution issue during the calibration, a new custom static cling was printed (see fig. 6). The geometric size of a chessboard pattern was increased to $8\,\mathrm{cm}$. Furthermore the white patches have been left blanc. Hereby the thermal contrast could be enhanced from $1\,\mathrm{K}$ to more than $50\,\mathrm{K}$.
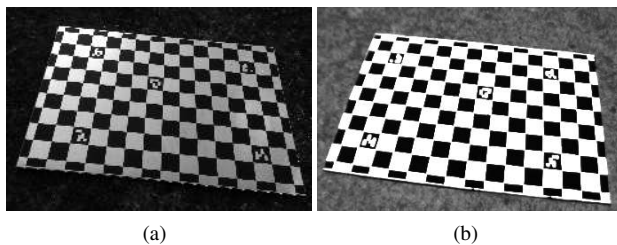

(a)   (b)

Figure 6. Improved chessboard with higher thermal contrast and larger patterns (a) VIS (b) inverted LWIR image (dark tones mean cooler temperatures)

ACKNOWLEDGEMENTS

REFERENCES

Aguilera, C. A., Aguilera, C. and Sappa, A. D., 2018. Melamine faced panels defect classification beyond the visible spectrum. *Sensors*.

Börner, A., Irmisch, P., Ernst, I. and Baumbach, D., 2018. Cameras for navigation and 3d modelling on planetary exploration missions. In: *International Workshop on Planetary Missions*.

Bouguet, J. Y., 2015. Camera Calibration Toolbox for Matlab. `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

Bradski, G., 2010. The OpenCV Library. `https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html`.

Brown, D. C., 1971. Close-range camera calibration. *Photogrammetric Engineering* 37(8), pp. 855–866.

Chen, Y., yu Huang, F., ming Shi, F., qi Liu, B. and Yu, H., 2019. Plane chessboard-based calibration method for a lwir ultra-wide-angle camera. *Appl. Opt.* 58(4), pp. 744–751.

Ernst, I., Zhang, H., Zuev, S., Knoche, M., Dhiman, A., Chien, H.-J. and Klette, R., 2018. Large-scale 3d roadside modelling with road geometry analysis: Digital roads new zealand. In: *15th Int. Symp. Pervasive Systems, Algorithms and Networks, IEEE online, YiChang, China, 2018*.

Kattnig, A., Thetas, S. and Primot, J., 2015. Ensuring long-term stability of infrared camera absolute calibration. *Opt. Express* 23(14), pp. 18381–18390.

Luhmann, T., Robson, S., Kyle, S. and Boehm, J., 2013. *Close-Range Photogrammetry and 3D Imaging*. De Gruyter Textbook, De Gruyter.

Morain, A. and Zanoni, M., 2004. Joint isprs/ceos-wgcv task force on radiometric and geometric calibration. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1(35), pp. 354–360.

Rankin, A., Huertas, A., Matthies, L., Bajracharya, M., Assad, C., Brennan, S., Bellutta, P. and Sherwin, G. W., 2011. Unmanned ground vehicle perception using thermal infrared cameras.

Shibata, T., Tanaka, M. and Okutomi, M., 2017. Accurate joint geometric camera calibration of visible and far-infrared cameras. *Electronic Imaging* 2017(11), pp. 7–13.

St-Laurent, L., Mikhnevich, M., Bubel, A. and Prévost, D., 2017. Passive calibration board for alignment of vis-nir, swir and lwir images. *Quantitative InfraRed Thermography Journal* 14(2), pp. 193–205.

Ursine, W., Calado, F., Teixeira, G., Diniz, H., Silvino, S. and de Andrade, R., 2012. Thermal / visible autonomous stereo visio system calibration methodology for non-controlled environments. In: G. Cardone (ed.), *QIRT E-book Proceedings*.

Vidas, S., Lakemond, R., Denman, S., Fookes, C., Sridharan, S. and Wark, T., 2012. A mask-based approach for the geometric calibration of thermal-infrared cameras. *IEEE Transactions on Instrumentation and Measurement* 61(6), pp. 1625–1635.

Wohlfeil, J., Grießbach, D., Ernst, I., Baumbach, D. and Dahlke, D., 2019. Automatic camera system calibration with a chessboard enabling full image coverage. In: *The International Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. in Print.

Yu, Z., Lincheng, S., Dianle, Z., Daibing, Z. and Chengping, Y., 2013. Camera calibration of thermal-infrared stereo vision system. In: *2013 Fourth International Conference on Intelligent Systems Design and Engineering Applications*, pp. 197–201.

Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11), pp. 1330–1334.

Zoetgnande, Y., Foug'eres, A.-J., Cormier, G. and Dillenseger, J.-L., 2018. Robust low resolution thermal stereo camera calibration. In: *11th International Conference on Machine Vision (ICMV18)*, Munich, Germany.