



HAL
open science

Robust low resolution thermal stereo camera calibration

Yannick Zoetgnande, Alain-Jérôme Fougères, Geoffroy Cormier, Jean-Louis Dillenseger

► **To cite this version:**

Yannick Zoetgnande, Alain-Jérôme Fougères, Geoffroy Cormier, Jean-Louis Dillenseger. Robust low resolution thermal stereo camera calibration. 11th International Conference on Machine Vision (ICMV18), Nov 2018, Munich, Germany. pp.110411D, <10.1117/12.2523440>. <hal-01919897>

HAL Id: hal-01919897

<https://hal.science/hal-01919897v1>

Submitted on 12 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Robust low resolution thermal stereo camera calibration

Yannick Wend Kuni Zoetgnande^a, Alain-Jérôme Fougères^b, Geoffroy Cormier^c, and Jean Louis Dillenseger^a

^aUniv Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France

^bECAM Rennes - Louis de Broglie, 35170 Bruz, France

^c Neotec Vision, 35740 Pacé, France

ABSTRACT

In this paper, we address the particularly challenging problem of calibrating a stereo pair of low resolution (80×60) thermal cameras. We propose a new calibration method for such setup, based on sub-pixel image analysis of an adequate calibration pattern and bootstrap methods. The experiments show that the method achieves robust calibration with a quarter-pixel re-projection error for an optimal set of 35 input stereo pairs of the calibration pattern, which namely outperforms the standard OpenCV stereo calibration procedure.

Keywords: Stereo calibration; Thermal; Stereo Vision; Low resolution

1. INTRODUCTION

Thermal infrared cameras are sometimes used in surveillance context [1, 2, 3, 4]. Very often in these cases, thermal cameras are preferred to visible images because they can be used in daily and nightly situations and moreover they preserve the anonymity of the people which is a crucial issue for surveillance.

One of the goals of the surveillance is the localization of people and objects in the 3D space. In this context, some authors proposed to associate the thermal camera with another device such as visible camera [5] or a depth sensor [6]. However, it seems also possible to use stereoscopy with a pair of thermal cameras. In this latter case, the stereo camera calibration is the crucial issue for 3D reconstruction.

One of the classical ways to perform this calibration is to image a geometrically well-defined calibration pattern. Usually, a pinhole camera model is assumed for both cameras. The pair of images is then used to estimate the extrinsic (the relative position of the cameras to each other) and intrinsic (focal lengths, principal points) parameters of the stereo camera model.

However, thermal cameras with good resolution are relative expensive [7]. Recently, some manufacturers proposed some very low price thermal cameras, e.g. the FLIR Lepton 2 *, but with very low resolution (80×60 pixels). In this paper, we will propose a solution for the stereo calibration of low-resolution thermal cameras. For this, we had to solve several sub-problems. 1) Given that thermal cameras are only sensitive to heat, we had first to choose and manufacture a calibration grid adapted to our low-resolution context. 2) The cameras gave only low-resolution and extremely noisy images so it is very challenging to exploit these images in 3D vision, especially for the calibration featured detection in the images. We choose to overcome this problem by performing calibration with a high number of image pairs. The accumulation of image pairs allowed to lower the impact of the noise and gain in precision for the estimation of the parameters. However, in our opinion, it was important to know the optimal number of image pairs in order to get a robust calibration. That consisted to find the good trade-off between the ease to perform and the robustness of the calibration.

This paper is structured as follows: Section 2 recalls some background about stereo vision. Section 3 presents some related work about stereo thermal cameras calibration and more specifically about calibration grids and the point detection process in thermal images. Section 4 details our method. Section 5 presents some results and our main contributions. Finally, section 6 concludes the paper and presents some perspectives.

Further author information: (Send correspondence to Yannick Zoetgnande.)

Yannick Zoetgnande.: E-mail: yannick.zoetgnande@univ-rennes1.fr, Telephone: +33 6 19 37 48 32

*<https://www.flir.com/globalassets/imported-assets/document/lepton-2.5-family-datasheet.pdf>

2. BACKGROUND

In the pinhole camera model context, stereo calibration consists to estimate the extrinsic and intrinsic parameters of the pair of camera models. The extrinsic parameters are related to the 3D pose of the cameras. They express the transformation of a point from the 3D world to the camera 2D coordinates. The pose can be view as a rotation (R) with a translation (T). In the stereo calibration, these values express the position of one camera (i.e the right) relative to the other one (i.e. the left). Intrinsic parameters refer to focal length and the principal point of both cameras. They can be estimated through the homography calibration method. Let P_l, P_r two homogeneous points respectively viewed by the left and right cameras. We have the relation:

$$P_r = RP_l + T \quad (1)$$

Given two points P_l and P_r which represent the same point in the 3D space, the idea is to determine R and T .

Let K_l and K_r respectively the intrinsic parameters of the left and right cameras, q and p the 2D points in resp. the left and the right cameras. We have:

$$P_r^t K_r^t R [T]_{\times} K_l^{-1} P_l = 0 \quad (2)$$

where $F = K_r^t R [T]_{\times} K_l^{-1}$ is the fundamental matrix. It encodes the information of the intrinsic and extrinsic parameters. $K_{i \in \{r, l\}}$ is the left (right) camera matrix. It is expressed as follows:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where f_x and f_y represent the focal length, c_x and c_y represent the principal point.

When the two cameras are uncalibrated, the matching between 8 points allows getting F . However, when the two cameras are calibrated (we already have the values of K_l and K_r), so the goal is to determine the essential matrix E which is:

$$E = R [T]_{\times} \quad (4)$$

where $[\]_{\times}$ is the cross product matrix. E encodes the information about the extrinsic parameters.

When we calibrate the cameras, the idea is to use corresponding points to determine R, T (which is also the baseline-the distance between both cameras), F and E . The quality of the calibration can be assessed by the RMS re-projection error (usually given by the calibration algorithm) and the average epipolar error. The average epipolar error refers to the error given by:

$$P_r^t \times F \times P_l = \epsilon \quad (5)$$

3. RELATED WORKS

3.1 Calibration grid

They are many works about calibration in stereo vision [8]. Most of these works concern visible images. A few works have been done for thermal stereo calibration. The main difference is that the points we want to detect must be hot. Generally, there are two ways to make the calibration grid visible by the thermal cameras: passive and active heating.

Passive heating consists to heat some black/white calibration grid. Because of the color, the black and white parts will not get the same temperature. This allows getting a gradient of temperature that is sufficient to distinguish features [9, 10]. Once the grid is heated, it can be used for imaging. Another solution is to use

sunlight to heat an asymmetric calibration pattern painted on a Dibond board [11]. In the best cases, they were able to get an RMS re-projection error of 0.348 on a 640×512 pixels camera. Shibata et al. [5] use the same calibration grid to make a joint geometric calibration of visible and far-infrared cameras. Each of these cameras is prior calibrated separately and then they estimate the extrinsic parameters. They get good results compared to the state of the art. The resolution of their visible and thermal images was respectively 640×480 and 160×128 . St Laurent et al. [12] try to find the best calibration pattern in order to calibrate stereo thermal infrared cameras. They present a passive calibration grid that is able to produce optimal image contrast in order to get a robust calibration. They are able to get an RMS re-projection error of around 0.1. They use thermal cameras with a resolution of 640×480 pixels.

Active heating consists to use a self-heating grid. The self-heating features can be small light bulbs [13, 14, 15] or resistors [16]. Several grids have been proposed: 81 light bulbs arranged in a 9×9 matrix [14], black/white chess with 25 bulbs arranged in a 5×5 matrix to be used to calibrate simultaneously thermal and visible cameras [15],...

It is important to notice the fact that the lowest resolution used in these works is 160×128 [5]. But they used a hybrid stereo calibration with a good resolution visible camera (640×128) and a low-resolution thermal camera. None of these works doesn't concern low-resolution cameras such as ours.

3.2 Point detection

The problem is twice, first the features must be segmented and then they must be located on the images. Generally, a simple threshold is sufficient to segment the regions that represent features (e.g. the bulbs). However, depending on the resolution of the thermal camera, it is more or less important to locate the features in sub-pixel precision.

Yang et al. [15] propose a semi-automatic calibration method. They select first the four extremities of the matrix containing the bulbs. Then the features are estimated in sub-pixel accuracy using a quadratic approximation. The resolution of their thermal and visible cameras were respectively 240×320 and 768×576 .

Ellmauthaler et al. [14] propose an algorithm to get a robust calibration. The algorithm has three steps. First for each calibration image each point in the grid is estimated in sub-pixel accuracy by computing the centre of mass of each point after a grey-scale thresholding. In the second step, they estimate the camera parameters. They used then the estimated homography matrix H to refine the position of the centre of mass. This process is iteratively performed until convergence. It is important to note that this calibration, done in monocular vision, can be applied to stereo vision. But, the calibration is performed on a camera with a resolution of 320×244 .

4. OUR METHOD

The cameras, we were using are more challenging given their resolution and their noise. So, it is necessary to adopt a different way to calibrate and evaluate the results. As expressed in the introduction, we will use multiple pairs of images to reduce the impact of the noise and low resolution. The calibration grid must be relatively easy to handle and moreover, the thermal property of the features must be constant over time.

4.1 Calibration grid

First, we discarded passive calibration grids even if they are easier to handle but the external thermal condition influences the quality of the image [12]. Given that we wanted to collect many images, the grid can get cool more and less quickly changing the measured information in the images over time.

For the design of our active grid, we had to take into account the low resolution and the reduced field of view of our cameras. We have manufactured a calibration grid (Fig. 1) which is composed of automobile bulbs placed on a wooden board. The temperature of the bulbs can higher than 37 Celsius degrees, which correspond to human body temperature. The wooden board allows us to make sure that the bulbs are well isolated and so they will be distinguished by the cameras. We placed 36 bulbs as a 6×6 matrix. Each bulb was separate to its row-wise and col-wise nearest neighbours by 160 mm. The size of the grid is 1 meter \times 1 meter. For the calibration purpose, we should make sure that all lights have almost the same brightness.

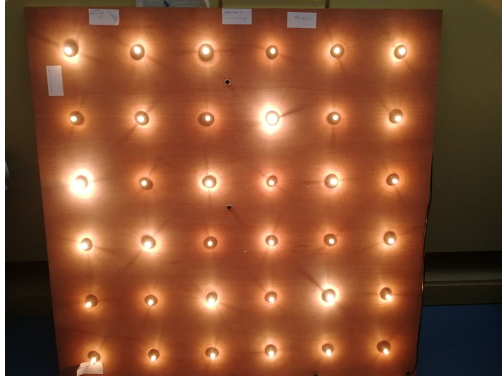


Figure 1: Calibration grid

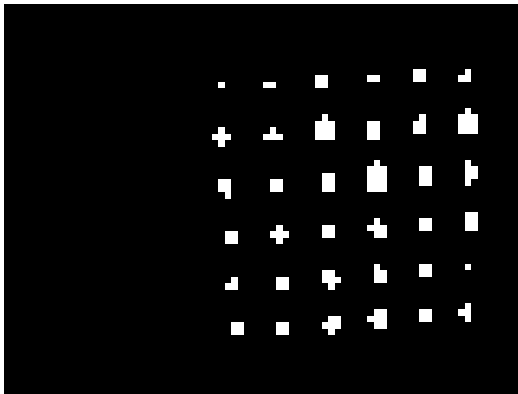
4.2 Stereo calibration

We used two fixed cameras separated by a distance of 140 mm.

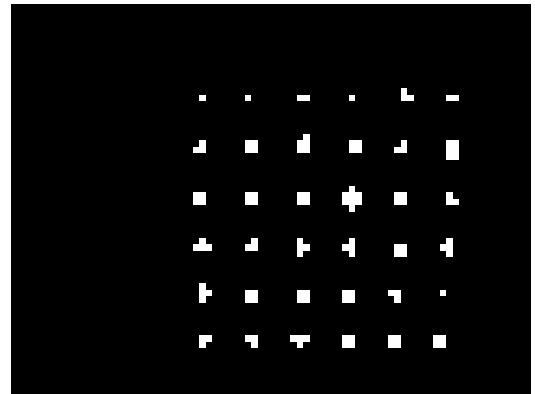
To perform the stereo calibration it is important to accurately detect the points. In the OpenCV calibration module [17], points are detected in pixel mode accuracy. This accuracy is sufficient for most of the cameras which have an acceptable resolution but not in our case of low resolution and noisy images. Inspired by the work of Ellmauthaler et al. [14], we developed our own sub-pixel point detection tools. Because we manufactured a wooden support for the calibration grid which is less susceptible to drive the heat, a simple grey scale threshold is sufficient to separate the bulbs information from the background.

4.2.1 Calibration features localization

Given N views, the goal is to extract and locate the points through sub-pixel accuracy. The images furnished by the FLIR sensor are in 16 bits, so first we normalized the image pixels in the range $[0,255]$. The image is then binarized using a threshold estimated by the Otsu method [18] (Fig. 2). We extracted then all the connected components, estimated their centroid and give them a label. We get the list of the points located in pixel accuracy but, as shown in Fig. 2, the points are not always well located. The subpixel accuracy was obtained by computing the centre of mass on the grey-level information around each centroid. This process is performed on all the original thermal images.



(a) Left image



(b) Right image

Figure 2: Stereo pair binary images

4.2.2 Parameters estimation

The parameters estimation was performed once we have detected all the 36 points in all the image pairs. We used these points as inputs for calibration methods available in the OpenCV calibration module [17]. As outputs we

got intrinsic parameters such as cameras matrices (M_l and M_r) and distortion vectors, and extrinsic parameters such as fundamental matrix F , essential matrix E , rotation matrix R and translation matrix T . Our calibration method took into account radial and tangential distortion.

4.3 Robustness of the calibration

The main particularity of our calibration is the low resolution and the noise of our images. It can be easily shown that if the number of input image pairs used for calibration is too low, the noise can have a direct impact on the parameters estimation i.e. if we perform the calibration on 2 different set of input images, it is highly probable that to get different values of stereo calibration outputs. This is mainly due to the difficulty to detect accurately the positions of the points in the images. If we increase the number of input image pairs, the calibration becomes more robust to noise in the sense that if we perform the calibration on 2 different set of input images we will get similar estimated parameters. But in this case, we also increase the difficulty and the handling time of the process. A tradeoff must be found between accuracy or robustness and handling facility.

In order to estimate the optimal number of images to get a robust calibration, we used the following bootstrap method.

- First we create a set of 100 image pairs of the calibration grid. We verified that in these images all the 36 points were visible.
- The robustness of the calibration using N pairs of input images was estimated by:
 1. We randomly selected N pairs of images from the 100 and performed calibration. The calibration outputs are stored.
 2. We repeated this process 1000 times in order to get 1000 calibration outputs.
 3. The dispersion of the 1000 calibration outputs was a good indicator of the robustness of the calibration.
- This process was repeated for several N with $N \in \{5, 10, 15, \dots, 40\}$.

From all the different calibration outputs, we choose to observe more particularly four specific outputs which are:

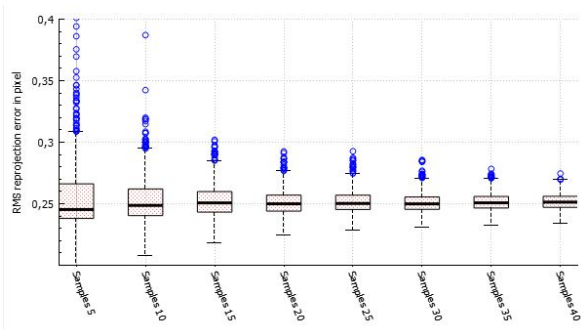
- the estimated focal length
- the RMS re-projection error
- the average epipolar error (see Eq. 5)
- the estimated baseline: this value can be easily compared to the ground truth which is 140 mm.

5. RESULTS

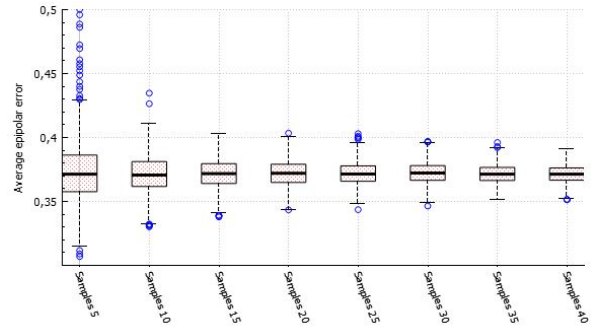
Through the bootstrap method, we can determine how the values converge to a certain value. We use box plots to observe the dispersion and to detect outliers.

In Figs. 3a-3d we can see that the median distribution of the calibration results is relatively similar regardless x and close of the ground truth for the baseline (140 mm). However, we can see that if $N \leq 30$ image pairs, the dispersion of the calibration results is very high with a lot of outliers. From 35 image pairs, we can see that the dispersion is stabilized and relatively small. We can thereby consider that from 35 image pairs, the calibration will almost estimate the same parameters regardless of the input images, and so our calibration can be considered robust.

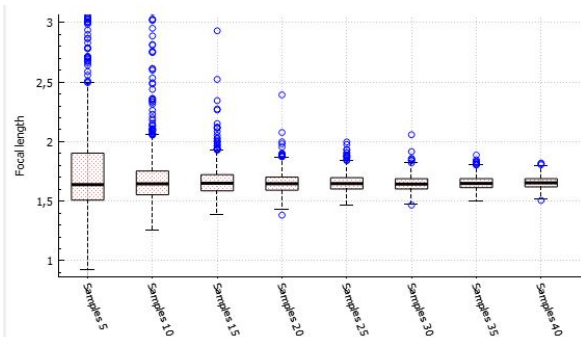
Concerning the accuracy, with 35 image pairs, the focal length and the baseline were estimated relatively precisely. The RMS seems to be stabilized at 0.25 pixels which correspond to an RMS of 1 pixel in 320×240 resolution. This could be considered as relatively important but in our case, we are faced with very noisy images which is not the case in high-resolution images. Moreover, if we would apply our method directly on



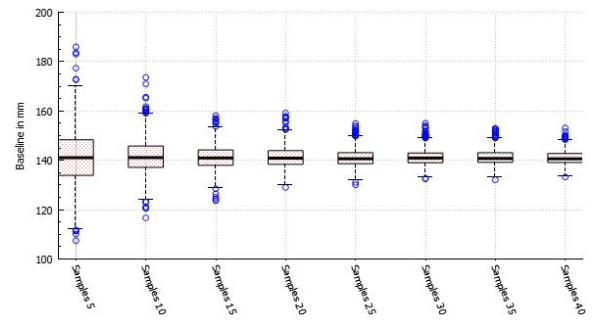
(a) Box plot of the RMS re-projection error



(b) Box plot of the average epipolar error



(c) Box plot of the focal length



(d) Box plot of the baseline

Figure 3: Results

Method	Camera	Resolution	MRE $_{pixels}$
Gschwandtner et al. [16]	PathfindIR	360×288	0.4918
Yang et al. [15]	GUIDE IR112	320×240	1.2214
Vidas et al. [19]	Miricle 307K	640×480	0.3031
Proposed method (35 images pair)	FLIR Lepton 2	80×60	0.20

Table 1: Comparison of our method with others

high-resolution images, we would probably have better accuracy than the classical methods because of our noise robustness and sub-pixel feature localization.

In order to compare our method with previous ones, we have compute the mean reprojected error (MRE) because the previous methods use this metric for comparison. The results are shown in the table 1. It can be noticed that despite the fact that we have a very low resolution we are able to compete with other methods that have higher resolutions.

6. CONCLUSION

In many applications, thermal images are used to characterize hot objects. In such a situation, it is important to locate an object in 3D space. This is can be done under stereoscopic vision. However, this later needs an accurate stereo camera calibration. This calibration is made more difficult when the images come from low-resolution cameras (80×60 pixels in our case).

In order to overcome this difficulty, first, we proposed a calibration grid composed of a matrix of 36 bulbs placed on a wooden board. In a second step, we proposed a method to locate features on the thermal images with sub-pixel accuracy. Finally, we estimate the optimal number of image pairs which has to be used for a robust calibration.

We have shown that in low 80×60 pixels resolution context we can get a robust calibration. We can retrieve the calibration parameters and also some physical values such as the focal length of the cameras or the baseline between the cameras.

Our other main contribution concerns the estimation of the number of images needed to get a robust calibration. In the state of the art, none have made a study to estimate the optimal number of images that give almost always a robust calibration. Calibration is generally performed only one time without any warranty of the relevance of the obtained calibration parameters. We have shown that in our context, with 35 images it is always possible to get a robust calibration.

Generally, once the camera is calibrated it is possible to perform the 3D reconstruction. This is done after a matching process to associate each point in the left image to a point in the right image. Due to the low resolution of our images, some techniques applied to high resolution visible and thermal images will not work well. So we must find the best way to do it.

ACKNOWLEDGMENTS

This work was part of the PRuDENCE project (ANR-16-CE19-0015) which has been supported by the French National Research Agency (ANR).

References

- [1] Anindya Maiti and S Sivanesan, “Cloud controlled intrusion detection and burglary prevention stratagems in home automation systems,” in *2nd Baltic Congress on Future Internet Communications (BCFIC), 2012*. IEEE, 2012, pp. 182–186.
- [2] Massimo Bertozzi, Emanuele Binelli, Alberto Broggi, and MD Rose, “Stereo vision-based approaches for pedestrian detection,” in *IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR’05)*, 2005, pp. 16–16.
- [3] Frédéric Suard, Alain Rakotomamonjy, Abdelaziz Bensrhair, and Alberto Broggi, “Pedestrian detection using infrared images and histograms of oriented gradients,” in *IEEE Intelligent Vehicles Symposium, 2006*, 2006, pp. 206–212.
- [4] Massimo Bertozzi, Alberto Broggi, Claudio Caraffi, Mike Del Rose, Mirko Felisa, and Guido Vezzoni, “Pedestrian detection by means of far-infrared stereo vision,” *Computer vision and image understanding*, vol. 106, no. 2-3, pp. 194–204, 2007.
- [5] Takashi Shibata, Masayuki Tanaka, and Masatoshi Okutomi, “Accurate joint geometric camera calibration of visible and far-infrared cameras,” *Electronic Imaging*, vol. 2017, no. 11, pp. 7–13, 2017.
- [6] Johannes Rangel, Samuel Soldan, and A Kroll, “3d thermal imaging: Fusion of thermography and depth cameras,” in *International Conference on Quantitative InfraRed Thermography*, 2014.
- [7] CNET, “Heat seeker: Meet the thermal-imaging camera you can afford,” <https://www.cnet.com/news/heat-seeker-thermal-imaging-camera-for-the-masses/>, 2014.
- [8] Christoph Strecha, Wolfgang Von Hansen, Luc Van Gool, Pascal Fua, and Ulrich Thoennessen, “On benchmarking camera calibration and multi-view stereo for high resolution imagery,” in *IEEE Conference on Computer Vision and Pattern Recognition, 2008 (CVPR 2008)*, 2008, pp. 1–8.
- [9] W Ursine, F Calado, G Teixeira, H Diniz, S Silvino, and R De Andrade, “Thermal/visible autonomous stereo visio system calibration methodology for non-controlled environments,” in *11th International Conference on Quantitative Infrared Thermography*, 2012, pp. 1–10.
- [10] Weiwei Kong, Daibing Zhang, Xun Wang, Zhiwen Xian, and Jianwei Zhang, “Autonomous landing of an uav with a ground-based actuated infrared stereo vision system,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2013*, 2013, pp. 2963–2970.

- [11] Josh Harguess and Shawn Strange, “Infrared stereo calibration for unmanned ground vehicle navigation,” in *Unmanned Systems Technology XVI*, 2014, vol. 9084, p. 90840S.
- [12] L St-Laurent, M Mikhnevich, A Bubel, and D Prévost, “Passive calibration board for alignment of vis-nir, swir and lwir images,” *Quantitative InfraRed Thermography Journal*, vol. 14, no. 2, pp. 193–205, 2017.
- [13] Arturo Rankin, Andres Huertas, Larry Matthies, Max Bajracharya, Christopher Assad, Shane Brennan, Paolo Bellutta, and Gary W Sherwin, “Unmanned ground vehicle perception using thermal infrared cameras,” in *Unmanned Systems Technology XIII*, 2011, vol. 8045, p. 804503.
- [14] Andreas Ellmauthaler, Eduardo AB da Silva, Carla L Pagliari, Jonathan N Gois, and Sergio R Neves, “A novel iterative calibration approach for thermal infrared cameras,” in *20th IEEE International Conference on Image Processing (ICIP), 2013*, 2013, pp. 2182–2186.
- [15] Rongqian Yang, Wei Yang, Yazhu Chen, and Xiaoming Wu, “Geometric calibration of ir camera using trinocular vision,” *Journal of Lightwave technology*, vol. 29, no. 24, pp. 3797–3803, 2011.
- [16] Michael Gschwandtner, Roland Kwitt, Andreas Uhl, and Wolfgang Pree, “Infrared camera calibration for dense depth map construction,” in *IEEE Intelligent Vehicles Symposium (IV), 2011*, 2011, pp. 857–862.
- [17] Gary Bradski and Adrian Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*, ” O’Reilly Media, Inc.”, 2008.
- [18] Nobuyuki Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [19] Stephen Vidas, Ruan Lakemond, Simon Denman, Clinton Fookes, Sridha Sridharan, and Tim Wark, “A mask-based approach for the geometric calibration of thermal-infrared cameras,” *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 6, pp. 1625–1635, 2012.