Practical Infrared Visual Odometry

Paulo V K Borges, *Senior Member, IEEE* and Stephen Vidas, *Member, IEEE*

***Abstract*—The use of cameras as a sensor for odometry estimation is an active research topic that has seen significant growth in recent years. Most methods, however, are only suitable for standard cameras that rely on reasonable lighting. An alternative to overcome low light conditions is the use of thermal or long-wave infrared imaging. Although visible spectrum and thermal imaging share many characteristics, it is not straightforward to apply standard visual odometry algorithms to thermal imaging data. In this paper we propose a practical visual odometry system based on a monocular thermal camera. As monocular odometry suffers from an unknown scale factor, the system performs efficient ground plane detection for targeted feature extraction, such that the scale factor can be reliably estimated if the camera height and pitch are known. We also address the problem of periodic non-uniformity correction, a necessary characteristic of thermal cameras which freezes the output potentially for over a second and can severely affect motion estimation. In this sense, we automatically determine appropriate times to perform non-uniformity correction based on the current and predicted camera rotations. Experiments illustrate the applicability of the system and compare it with other state estimation approaches.**

***Index Terms*—Visual odometry, autonomous vehicles, thermal imaging, long-wave infrared.**

I. Introduction

Visual odometry is a well known functionality of computer vision systems, in which the egomotion of a camera mounted on a vehicle or robot is estimated. The estimation process considers that only the visual input from one or more cameras is used. With the term initially proposed by Nistr et al. [1], visual odometry finds application in intelligent vehicles, robotics and augmented reality, among others. In vehicles in particular, it can provide important prior information in Simultaneous Localization And Mapping (SLAM) or obstacle detection, serving as an alternative or supplement to wheel odometry, Global Positioning System (GPS) or Inertial Measurement Units (IMUs), for example. Compared to those localization modalities, visual odometry offers different advantages. In relation to wheel speed sensors, visual odometry can present higher accuracy, particularly in slippery terrain [2]. Regarding GPS and IMU-based systems, they are generally low accuracy or present very high sensor costs. GPS-denied environments are arguably one of the most significant examples in which visual odometry has great relevance, such as in underwater or underground navigation. Visual odometry is made possible through the analysis of changes in position of particular points in the image from frame to frame, generally induced by the camera motion. As with many computer vision tasks, sufficient illumination and enough texture (such that features can be extracted) are necessary for motion estimation to work satisfactorily. In practice, however, challenging or low light conditions are often part of real application scenarios.

Tackling this problem, in this paper we consider the "visual" odometry task using a thermal camera. The term "visual" is quoted as in theory the information used lies in the non-visible (i.e. thermal) spectrum, although it can be represented in a visible format for human interpretation, and can be subject to image processing in a similar way to conventional imagery.

Far infrared imaging has been previously combined with other sensors for ego-motion estimation [3], [4]. These works indicate that statistics obtained from long-wave infrared (LWIR) images can be fused with other sensors to satisfactorily improve the results achievable using visible cameras. In this paper we do not address the information fusion problem, but focus on practical and theoretical challenges that arise when using thermal imaging only. The proposed solutions move towards the limits of this sensing modality for visual odometry, applied to vehicle egomotion estimation in outdoors scenarios.

The contributions presented are motivated by the fact that night-time operations of autonomous vehicles find applications in many fields. Aside from the generic goal of dependable autonomous navigation, in industry and agriculture, for example, some systems must operate 24 hours a day, motivating the use of thermal cameras as a sensor for localization. Other modalities such as Short-Wave Infrared (SWIR) also offer complementary statistical properties to visual images. However, LWIR was preferred for the experiments because it represents a more distinct imaging modality. While regular visible-spectrum images capture radiation (light) reflected off an object, thermal infrared images primarily capture radiation emitted from objects, even at room temperature, and are therefore independent of the effects of lighting.

A. Contributions of this Work

Although thermal and visible spectrum images share several statistical characteristics, some fundamental aspects make direct application of standard visual odometry algorithms to thermal imaging data hard or unfeasible. A significant problem in thermal images is a generally reduced number of features, in particular on road surfaces, due to a lower signal-noise ratio than visible images under daylight conditions. This is a relevant issue, as a reliable estimation of road feature points is essential (when other types of speed estimation are not available) to estimate the scale factor in monocular odometry [5]. For this reason, the proposed system performs efficient ground plane detection for targeted feature extraction. This is achieved by performing intelligent segmentation of the ground, combining information from motion and appearance

P. Borges is with the Autonomous Systems Laboratory, CSIRO Computational Informatics. Address: 1, Technology Court, Pullenvale, QLD, 4069, Australia. paulo.borges@csiro.au

S. Vidas is with Digital Applications International, Aberdeen, Scotland. stephen.vidas@gmail.com

to segment the road. If the road is satisfactorily segmented, the feature point detection sensitivity can be increased for points lying on the road. Hence, reliable road (ground) points in the odometry computation allow for better scale estimation.

Another inherent aspect of many thermal cameras is the Non-Uniformity Correction (NUC) [6], whose aim is to eliminate fixed pattern noise that accumulates with time in thermal digital imaging sensors. The NUC freezes the camera for a small amount of time (typically between 0.3 and 2 seconds) and can severely affect motion estimation. We mitigate this problem by determining suitable times for NUCs based on the current and predicted camera rotations. We show that this proposed 'NUC Trigger Manager' reduces the negative impact of NUCs on visual odometry, improving position estimation. The experiments performed compare the performance of the standard algorithm to the proposed method. For the tests, the thermal camera was mounted on a ground vehicle and driven in structured and unstructured environments. The results indicate key points of failure of the standard algorithm and illustrate improved visual odometry using the system described.

It is important to note that one of the goals of this contribution is to propose strategies that require little adaptation of standard visual odometry algorithms (enabling them to work successfully with thermal-infrared images), rather than proposing a fundamentally different visual odometry approach. As far as our literature review could tell, there are no other visual odometry methods that have been adapted to or designed explicitly for the thermal-infrared modality.

This paper is organized as follows. In Section II we review background concepts that are relevant for the proposed method. In Section III we describe the proposed method and the NUC Trigger Manager. In Section IV we present the method for enhanced road estimation, followed by experiments in Section V. We finish the paper with relevant conclusions in Section VI.

II. Background

In this section we review background concepts related to monocular visual odometry as well as to thermal imaging. A description of these concepts allows for better understanding of the contributions presented in this paper.

A. Monocular Visual Odometry

When performed with a single camera, motion estimation suffers from scale ambiguity. To overcome this limitation without integrating IMUs or GPS into the system [7], [8], a common strategy is to make assumptions about the camera mounting and the environment, such as the road [5], [9] and the structure of surrounding buildings [10]. For ground vehicles, assuming a locally planar road model in the vicinity of the vehicle is reasonable, as illustrated in Figure 3. Although roads usually present some degree of curvature, a linear model is a commonly used approximation [5], [11] that is very effective in structured environments. In this case, if the intrinsic camera calibration parameters K are known and fixed, as well as the height and pitch of the camera, scale can be retrieved [5]. Given a sequence of input images, the odometer must estimate the pose of the camera at each frame using only image information. To achieve this, salient points x_j^k are tracked over a number of frames. These salient points can be corners or blobs [12], for example. In the salient points notation, j corresponds to the index of the feature and k indicates the frame index. Tracking these points from frame to frame, it is possible to estimate the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$ between the current and previous frames. Therefore, the rigid body transformation relating the poses at instants k and $k-1$ is given by

$$\mathbf{T}_{k,k-1} = \begin{bmatrix} \mathbf{R}_{k,k-1} & \mathbf{t}_{k,k-1} \\ 0 & 1 \end{bmatrix} \quad (1)$$

Transformations can be integrated over each frame for the full trajectory, so that the poses of the moving camera with respect to a predefined world reference frame are determined.

B. Infrared Visual Odometry

Video-based localization using only thermal-infrared video has been found to be highly challenging (particularly in 3D), but possible in highly constrained environments [13]. For applications such as loop closure detection, the thermal-infrared modality has been demonstrated to be complimentary in nature to the visible spectrum [14]. In principle, thermal images could be used directly as input for the framework described in Section II-A using standard methods that work in the visible domain. However, two important aspects make the application of standard algorithms in thermal-infrared challenging: (i) more difficult salient point selection and (ii) the occurrence of NUCs. These problems, and their proposed solutions, are detailed in the following two sections.

1) Ground Point Detection for Scale Estimation: In the visible domain, image statistics are generally richer, particularly in terms of texture and high frequency spatial information, enabling better detection and tracking of salient points. This is especially important in the case of monocular visual odometry methods which rely on ground points being tracked. The importance of detecting features on the ground plane is that, with a known camera height and angle, it is possible to retrieve an approximate scale information from a monocular camera. A low number of detected points on the road severely reduces the algorithm's performance. In the case of thermal images, unfortunately, roads generally present low contrast and the number of features able to be tracked reliably is very low, limiting the effectiveness of the road estimation procedure. For monocular visual odometry, reliable tracking of enough ground points is essential for scale estimation. To mitigate this problem, in Section IV we perform an intelligent salient point selection, increasing the detectability of points that lie on the ground. In this scenario, we employ a road detection algorithm applied to thermal images that indicates potential road regions.

2) Non-Uniformity Correction: The raw digital output of thermal-infrared cameras is greatly affected by the presence of spatial non-uniformities, producing a phenomenon known as Fixed-Pattern Noise [15]. If this noise is allowed to accumulate, it can substantially reduce the reliability of feature

Fig. 1: Effect of the accumulation of 5 minutes of non-uniformity noise (image on the right) on the quality of thermal-infrared video. Note that non-uniformity noise is always present in thermal-infrared images. However, the right image contains much more of this noise, as it was captured much longer after the most recent NUC had been performed. The left image, in contrast, was captured only a few seconds after the NUC.

extraction and tracking algorithms [16]. Figure 1 shows the effect of non-uniformity noise.

As a result, many thermal cameras have a built-in mechanical shutter that is used to re-calibrate the sensor during operation and reset the sensor noise [17], performing an operation known as a Non-Uniformity Correction, or NUC. This procedure results in a delay of approximately 1 second during which no new data can be obtained. Such a disruption can be fatal for conventional feature-tracking methods which assume a smooth motion of features from frame to frame.

Most thermal cameras perform NUCs periodically, but the NUC frequency depends on the sensor quality and application. Typical values range between 1 and 3 NUCs per minute. During the NUC, the input stream is frozen for a period of time. Although short, this image freezing is sufficient to damage severely the quality of pose estimation (in particular because odometry errors accumulate over time), depending on the current camera state. In the following section we describe a NUC management approach that reduces the impact of NUCs on motion estimation.

III. NUC Management

Given the challenges discussed in Section II, in this section we present practical solutions that improve camera motion estimation significantly in the thermal domain. In particular, we introduce a module called 'NUC Manager' that triggers the NUC at suitable times.

For a camera viewing relatively large areas (like in our application scenario, where the camera is mounted on an outdoor ground vehicle), the error due to NUC freezing is most prominent when the camera is undergoing significant rotation. This is because feature displacement in the 2D image is significantly greater during rotation than regular forward-motion. Therefore, an efficient way to perform NUC management is to analyze the current camera rotation state. For this purpose, we evaluate the rotation (in our framework we represent rotations via quaternions with respect to an original reference) to determine whether or not NUCs are appropriate at a given time, with respect to the rotation variable. As the NUC lasts for a significant amount of time, not only the current rotation is relevant, but also the amount of rotation predicted for the near future. This prediction can be done by analyzing the structure of the environment, in particular of the road direction on which the vehicle navigates. A third factor that is also considered in deciding when to perform a NUC is the elapsed time since the last NUC, as this indicates how urgent it is. Finally, the temperature change experienced by the camera since the last NUC is also important - with a greater temperature change increasing the urgency.

Summarizing the considerations above, Figure 2 shows all variables considered by the NUC Trigger Manager for determining whether or not the NUC is allowed at a given instant. In that figure, $d_{\mathbf{r}}$ represents the angle of the current rotation quaternion, given by $2\arccos(w)$, where w is the real element of the quaternion. Equivalently, $p_{\mathbf{r}}$ is the angle of the predicted rotation quaternion. t_e and Δ_{t} correspond to the time elapsed and the temperature change since the last NUC, respectively. The output $T_{\mathrm{NUC}} \in 0, 1$ is a flag indicating whether the NUC should be triggered.

Let functions $f(d_{\mathbf{r}}, p_{\mathbf{r}}, t_e)$, $g(t_e)$, and $h(\Delta_{\mathrm{t}})$ be defined as

$$f(d_{\mathbf{r}}, p_{\mathbf{r}}, t_e) = \alpha_1 t_e + \alpha_2 d_{\mathbf{r}} + \alpha_3 p_{\mathbf{r}} \tag{2}$$

and

$$g(t_e) = \begin{cases} 0, & \text{if } t_e < t_{\max} \\ 1, & \text{if } t_e \geq t_{\max} \end{cases} \tag{3}$$

and

$$h(\Delta_{\mathrm{t}}) = \begin{cases} 0, & \text{if } \Delta_{\mathrm{t}} < \Delta_{\mathrm{t}_{\max}} \\ 1, & \text{if } \Delta_{\mathrm{t}} \geq \Delta_{\mathrm{t}_{\max}} \end{cases} \tag{4}$$

where α_1, α_2 and α_3 are normalization parameters. $t_{\max}$ is the maximum allowed time without NUC and $\Delta_{\mathrm{t}_{\max}}$ is the maximum temperature change allowed since the previous NUC. The NUC Trigger Manager rule is defined as

$$T_{\mathrm{NUC}} = \begin{cases} 0, & \text{if } f(d_{\mathbf{r}}, p_{\mathbf{r}}, t_e) + g(t_e) + h(\Delta_{\mathrm{t}}) < 1 \\ 1, & \text{if } f(d_{\mathbf{r}}, p_{\mathbf{r}}, t_e) + g(t_e) + h(\Delta_{\mathrm{t}}) \geq 1 \end{cases} \tag{5}$$

The output of this equation is evaluated on a frame by frame basis, preventing or allowing NUC. Respecting the output from (5), NUC can be executed at any time except when the camera has not yet completed the previously triggered NUC.

$d_{\mathbf{r}}$ is obtained directly from the current output of the visual odometry algorithm (before integration with previous measurements for trajectory generation). $p_{\mathbf{r}}$ depends on the geometry of the road, since future rotations depend on upcoming road curvature. Therefore, Section IV discusses a method for estimating the road surface and direction to generate $p_{\mathbf{r}}$.

The results in Section V illustrate the benefits of using the NUC manager, comparing the position estimation performance with and without this functionality.

IV. Road Estimation

The road estimation has two goals in this work: (i) to generate $p_{\mathbf{r}}$ used in the NUC manager and (ii) to highlight the ground as an area of interest for feature point extraction in visual odometry. A number of works have been proposed

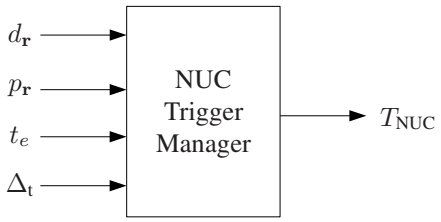


Fig. 2: Block diagram illustrating the inputs and output of the NUC Trigger Manager. $d_{\mathbf{r}}$ represents the normalized derivative of the current rotation quaternion and $p_{\mathbf{r}}$ is the predicted rotation. t_e and Δ_t corresponds to the time elapsed and the temperature change since the last NUC, respectively. The output $T_{\text{NUC}} \in 0, 1$ is a flag indicating whether the NUC should be triggered.

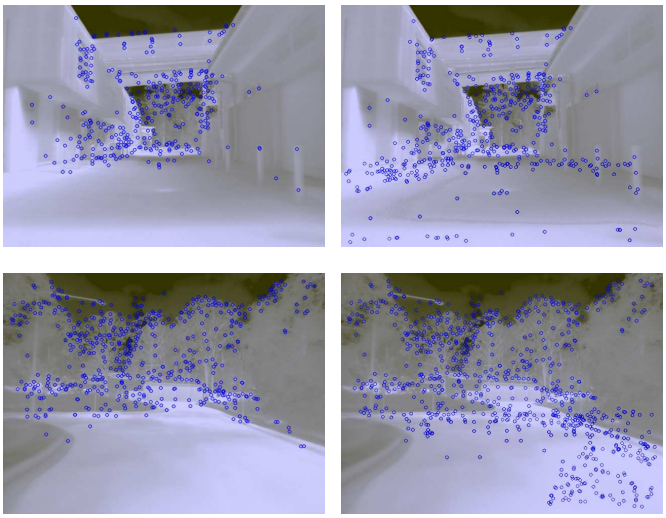


Fig. 3: Example thermal-infrared images of roads in industrial environments, where traditional road detection methods often perform poorly. The images on the left illustrate feature detection without the sensitivity parameter adapted for the road, where as the images on the right show additional points detected due to increased sensitivity.

for road surface estimation, based on various features such as color, texture and shape of the road [18], [19], [20]. Most methods, however, focus on roads with either well-defined geometry or with data from visible-spectrum cameras. Many industrial application scenarios, for example, require road detection on non-urban traffic roads without specific lane markings and curbs, aside from the data being from thermal cameras (see the examples in Figure 3). In this context, an alternative is to combine motion and appearance for the generic segmentation of objects in the scene [21]. Based on semantics (e.g. the expected position of a road in front of a vehicle), the road surface and direction can be determined.

The main pipeline for the road segmentation method is shown in Figure 4. The key aspect is the use optical flow vectors as supplementary information in image segmentation. In this paper, we employ a gradient-based watershed method [22] for segmentation. Therefore, we extend the common watershed implementation to include not only brightness information,

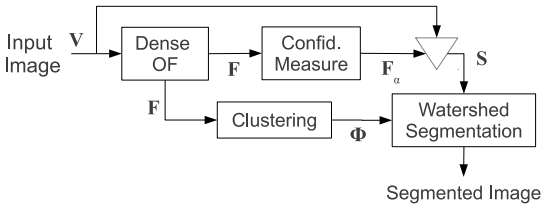


Fig. 4: Block diagram illustrating the basic pipeline of the proposed road detection method. The segmentation stage considers a multi-channel input, consisting of the raw input image and the dense optical flow representation in the corresponding frame. It also considers seed locations provided by clustering of the optical flow vectors.

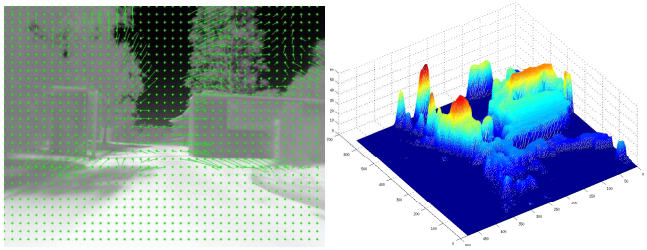
but also topological optical flow information, obtained from the difference between two consecutive frames. Optical flow data is applied in two ways: (i) by merging (via weighted sum) the optical flow "image" $\mathbf{F}$ and the original image $\mathbf{V}$ into another image $\mathbf{S}$, which is then segmented, and (ii) by clustering optical flow vectors and using the center of the clusters as seeds (represented by the set $\mathbf{\Phi}$) in the segmentation process. Figure 5(b) shows an example of the topological 3-D structure of the flow representation corresponding to the visible spectrum image in Figure 5(a). This figure illustrates that the optical flow topology is applicable to a watershed segmentation, with well-defined "hills" and "valleys", therefore being less prone to over-segmentation, as addressed in Section IV-B.

A. Semi-Dense Optical Flow

In our implementation, we use semi-dense optical flow (i.e. optical flow in a sub-sampled version of the original image) as it provides a satisfactory indication of the flow over the whole frame, which is not the case when relying only on feature points. Among dense optical flow techniques, one family of methods that has gained popularity (and we use in this paper) are three dimensional (3D) structure tensor methods [23], [24], mainly due to their low systematic error and robustness to noise. In order to work adequately, 3D structure tensor methods assume that the brightness change constraint is valid, such that a given feature will have constant brightness between two consecutive frames, and that it will only be translated. In order to regularize the results, it builds the tensor for each pixel element within its surrounding pixels, where local optical flow is considered invariant. The optical flow estimation is then converted to an eigenvalue analysis problem. Because the locally constant optical flow assumption is often not met in real applications, the 3D structure tensor technique can be extended to use the affine motion model. In this situation the tensor is determined by projecting the image into a second degree polynomial and integrating the affine model into the tensor, where the affine parameters are solved using linear systems [24].

B. Joint Segmentation

The joint segmentation merges the input image and its corresponding optical flow image into a single matrix, as indicated



(a) Thermal-infrared image flow. (b) 3D representation of flow magnitude, illustrating well defined "hills" and "valleys".

Fig. 5: Illustration of a thermal image and its flow representation.

by the block $\triangledown$ operation shown in Figure 4. Let $\mathbf{V}$ be the input image and $\mathbf{F}$ represent the flow image corresponding to $\mathbf{V}$. The two images are combined based on a weighted sum given in (6), which depends on a confidence level α of each flow vector. This confidence measure estimates the correctness of each displacement vector such that vectors with a lower α have less influence in the segmentation. α is determined following an intuitive technique based on linear subspace projections [25], obtaining "eigenflows" from the original optical flow signal. In summary, the confidence measure is determined based on the assumption that if displacement vectors can be well reconstructed, they are more reliable in their surrounding neighborhood. Hence, the reconstruction error of the flow vector is used to quantify the confidence measure.

The final segmentation using watershed is performed on "image" $\mathbf{S}$, whose i-th element is described as

$$s_i = \frac{v_i + \alpha_i |f_i|}{1 + \alpha_i} \tag{6}$$

where α_i is the confidence measure for the i-th optical flow vector in $\mathbf{F}$, and v_i and f_i are the i-th element of $\mathbf{V}$ and $\mathbf{F}$, respectively. The operator $|\cdot|$ represents the magnitude of the flow vector.

C. Outlier Removal and Clustering

In addition to being included as a topological region in the segmentation process, flow vectors can also be used as seeds in the watersheds. Since the cameras are mounted on a vehicle, it is reasonable to assume that the cameras move on a surface plane and the great majority of correct optical flow lines follow the same direction as their neighbors. Consequently, we can remove outliers by simply eliminating optical flow vectors whose directions differ more than a threshold τ_θ from the average optical flow direction. This operation reduces the occurrence of false local minima in the watershed segmentation that is caused by outliers.

In a scene, optical flow vectors are rarely induced from only one object. Generally there are several objects inducing optical flows in the image. Hence, it is possible to cluster optical flow vectors according to their location and characteristics.

In a given frame, each induced optical flow vector is described by its x position, y position, angle θ, and magnitude A. Let $\mathbf{D}_f$ represent the description of the optical flow vectors within the given frame, f. For this frame, the optical flow information is given as

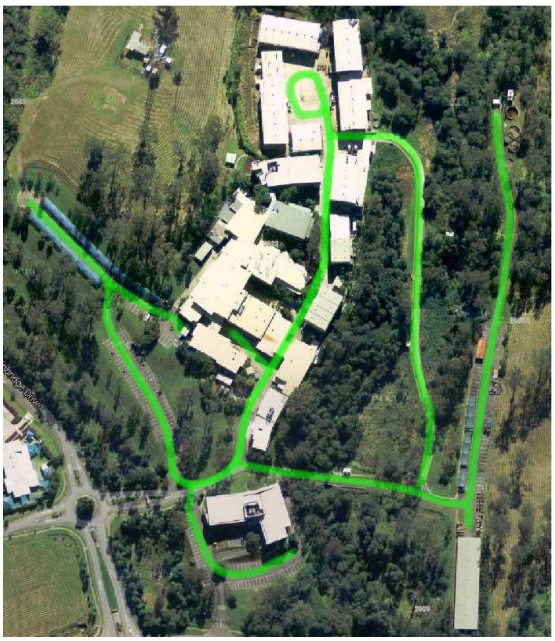
$$\mathbf{D}_f = \begin{bmatrix} x_1 & y_1 & \theta_1 & A_1 \\ x_2 & y_2 & \theta_2 & A_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_K & y_K & \theta_K & A_K \end{bmatrix}, \tag{7}$$

where K is the number of flow vectors in the image. Optical flow vectors are clustered according to their location, magnitude and angle by employing a Gaussian Mixture Model (GMM) in a 4 dimensional space. The mean of each cluster is projected onto the 2-D image space, and it is employed as a seed in the watershed segmentation algorithm. GMMs are probabilistic models that represent the presence of subpopulations within an overall population. Formally, a mixture model describes the mixture distribution that represents the probability distribution of observations in the overall population, and GMMs are particularly effective in clustering within normal distributions. Assuming that the distribution of the magnitudes of induced optical flow vectors is normal [26], GMMs can be employed to cluster the data according to their source. The number of clusters is determined using a standard methodology in which an expectation-maximization algorithm estimates the finite mixture models corresponding to each number of clusters considered and using Bayesian inference criterion to select the number of mixture components, which is then set as the number of clusters [27]. This methodology commonly generates between 10 and 25 clusters in typical outdoors thermal images.

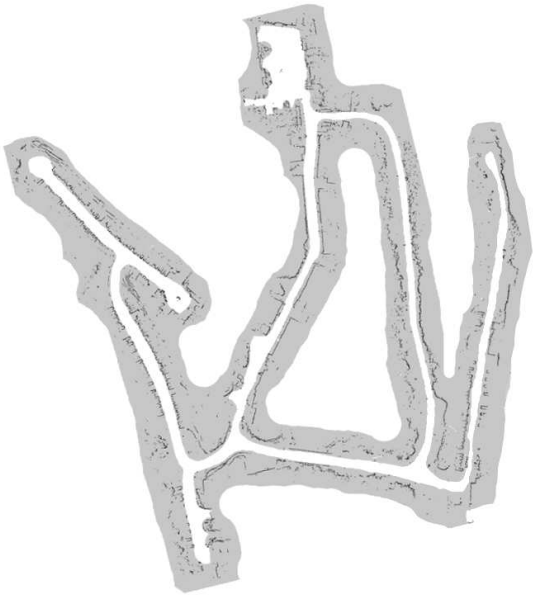
D. Segmenting the road

Once the image is segmented, the immediate cluster in a region R in front of the vehicle is set as the road cluster. If multiple clusters are found in R, the largest cluster is selected. As discussed, this cluster is used for (i) the calculation of the road direction and (ii) to assist in the identification of ground points in the plane estimation required by the odometry module. The main road direction is estimated using a triangular model [28] via least-squares linear fitting from the points on the edges of the segmented area. Although a triangular model is not highly precise, it suffices in providing a general direction of the upcoming road section, which is used as input as the variable $p_{\mathbf{r}}$ in the NUC manager.

The visual odometry algorithm estimates the ground plane to allow for the scale factor, which can be determined given that the height and pitch of the camera are known. With the knowledge of the areas that represent ground, the feature point detection threshold τ_{nms} is reduced, therefore increasing the number of points in that region. In particular, we change the peakiness threshold in the non-maxima-suppression, allowing for more feature points to be considered. Tuning of τ_{nms} must consider the trade-off between an increased number of points of interest and the robustness of these points. To determine a suitable value, an evaluation is performed where τ_{nms} is varied and the odometry estimation error is obtained. As an example, Figure 3 shows feature points with different



(a) Satellite view of the test area (drivable routes are highlighted in green), illustrating the structured and unstructured routes. The scale is 410×360 meters.



(b) Map of the test area from 2D laser scans mounted on the vehicle. The position from laser-based localization is used as ground truth for evaluating odometry error.

Fig. 6: Illustration of the test area and corresponding 2-D map.

sensitivity parameter for road regions. The experimental evaluation shown in Section V indicates that, for road regions in thermal data, a reduction of approximately 75% in τ_{nms} in comparison to visual spectrum data works satisfactorily. Hence, in our implementation we use $\tau_{\text{nms}} = 50$ for visual spectrum data and $\tau_{\text{nms}} = 13$ for thermal data.

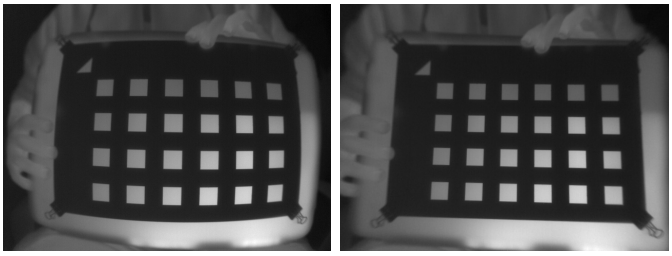


Fig. 7: Correction of distortion using the mask-based approach.

V. Experiments

In this section we run a number of experiments comparing the proposed system with visual odometry using standard cameras.

A. Practical Considerations

1) Hardware: For the experiments, a Thermoteknix Miricle 307K thermal-infrared camera and a Basler scA780 visible spectrum camera were used. The cameras were mounted on a John Deer "Gator" (Figure 8), an electric two-person utility vehicle, and driven in structured and unstructured environments in an industrial park in Australia. The area contains both urban and rural characteristics, with roads, trees, dense buildings and open fields. A satellite view of some of the approximate routes traversed is shown in Figure 6. The thermal-infrared camera consists of a long-wave uncooled microbolometer detector sensitive in the $7-14\mu$m range, with spatial resolution of 640×480. It is able to see objects in the -20 to $150\,^{\circ}$C temperature range, and it has a NEDT (Noise-Equivalent Differential Temperature) of 85mK. Images were geometrically calibrated to remove distortion and determine the intrinsic parameters of the camera using a mask-based approach [29]. The effect of this calibration is demonstrated in Figure 7.

(a) Robotic vehicle.

(b) Setup of the visible and thermal cameras.

Fig. 8: Vehicle and cameras used in the experiments.

2) Software: In our implementation, for the thermal camera data extraction and pre-processing we used the ROS (Robotics Operating System) [30] package developed by Vidas et al [13], which contains a FFMPEG4 based driver for streaming off the thermal camera.[1]. For this paper, we modified this package

[1]This open-source ROS package is available online for download from http://wiki.ros.org/thermalvis

TABLE I: Divided by the forward slash '/', the median positional errors (in meters) and heading errors (in degrees) are presented in the table, for varying traveling distances (e.g. 10 m, 20 m, etc). The blank fields (-) correspond to the night tests with the visual spectrum cameras, where no meaningful results are obtained. The expanded acronyms for each 'Odometry Setup' are given in the beginning of Section V-B.

	Day-time				Night-time			
Odometry Setup	10 m	20 m	50 m	100 m	10 m	20 m	50 m	100 m
VC	0.43 / 1.10	0.89 / 1.86	2.79 / 2.49	6.09 / 3.15	- / -	- / -	- / -	- / -
VC + RE	0.30 / 1.05	0.60 / 1.91	2.60 / 2.40	5.30 / 3.10	- / -	- / -	- / -	- / -
TC	1.02 / 12.00	3.95 / 20.23	12.40 / 25.31	21.05 / 42.75	1.08 / 13.87	4.01 / 18.46	12.61 / 18.14	21.43 / 41.03
TC + RE	0.74 / 10.29	2.98 / 18.70	10.51 / 21.49	18.78 / 32.45	0.80 / 15.33	2.89 / 11.76	10.33 / 17.04	19.80 / 32.50
TC + RE + NM	0.55 / 1.40	1.50 / 1.98	4.01 / 4.75	7.96 / 6.23	0.48 / 1.50	1.47 / 1.91	4.20 / 3.48	7.95 / 6.22
Wheel Odometry	0.49 / 1.70	1.67 / 3.23	6.44 / 7.39	15.47 / 12.45	0.50 / 1.70	1.64 / 3.22	6.32 / 7.44	15.10 / 12.11

adding the NUC management system (discussed in Section III) as a module linking motion estimation with NUC triggering.

Data was logged using ROS and the resulting frame rate was approximately 14 frames per second. For the odometry computation, as discussed in Section II-A, we use a state-of-the-art method for feature extraction, essential matrix computation and motion integration, implemented in the package Libviso2 [31]. In this algorithm, the only difference between the implementations visual and thermal modalities is the sensitivity parameters for applying optical flow in regions of interest, as explained in Section IV-D.

3) Ground-truth: All types of odometry suffer from unbounded error and incremental drifts. In addition, a typical characteristic is that the systematic error in odometry systems grows with the distance from the starting point, but the error is frequently reduced as the vehicle loops around and returns towards the origin [32]. This indicates that driving in a loop and computing the average Euclidean distance error to the origin can be a weak measure for the performance of odometry systems. A practical alternative is suggested by Johnson et al. [33], where a short segment of the odometry path is initially registered with the ground-truth path using a least squares method. The Euclidean distance error of their odometry system is then measured x meters down the path. The origin is slightly shifted and the procedure is repeated until the origin is x meters from the end of the driven path. Observing the median and the variance of the errors obtained, a more meaningful interpretation of the performance of the odometry module is achieved. In large open scenarios like our test area, an adequate value for x was determined empirically to be 100 meters.

For comparison purposes, the vehicle is equipped with four 2D lasers (one in each corner of the vehicle) which are used as input for a localization algorithm that is used as ground-truth. Figure 6(b) shows the map of the test area generated from 2D laser scans mounted on the vehicle. Although RTK-GPS is often used for ground-truth, in the industrial environment where the experiments are performed this type of localization is not a suitable solution. There are several metal buildings and sheds which cause severe multi-path and form a "semi-urban canyon", indicated by the central trajectory in Figure 6a.

B. Results

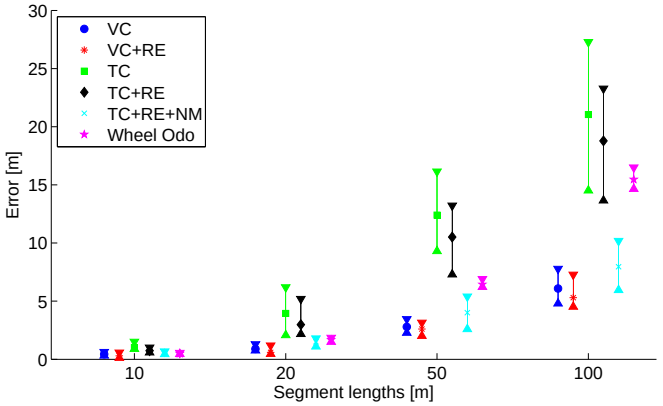
The results presented in the following compare the proposed framework against the direct application of monocular visual odometry to thermal imaging, with the goal of evaluating the improvements suggested. The results are presented in 5 parts, considering the following setups:

1) VC: Visual spectrum camera.
2) VC + RE: Visual spectrum camera with road estimation.
3) TC: Thermal camera (without NUC management or road estimation).
4) TC + RE: Thermal camera with road estimation (without NUC management).
5) TC + RE + NM: Thermal camera with road estimation and NUC management.

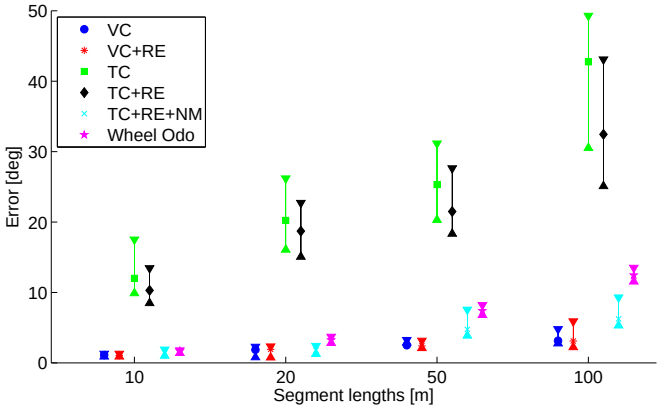
All setups above are tested for both day time and night time situations, where the benefit of using the thermal camera is evident.

Figure 11 shows the trajectories for the cases mentioned above for the day time trials. This figure illustrates that although the visual spectrum camera still performs better at day time, the proposed method improves significantly the performance of the thermal visual odometry. The advantage, of course, lies in the fact that the thermal camera can operate in the dark or night time, where the visible camera does not provide meaningful results. A key result is the fact that the average position error is reduced by more than 70% between simple thermal camera odometry and thermal camera odometry with road estimation and NUC management. Numerical results are summarized in Table I, for different distances (10, 20, 50 and 100 meters), as discussed in Section V-A3. Correspondingly, Figure 9 shows the spreading of results for the different runs, presenting both absolute position and heading errors. The figure illustrates the high variance in the case of thermal data without the NUC manager.

The trajectories for the night time tests present similar characteristics to those in Figure 11, aside from the fact that the visible camera fails as expected. Overall, results show that the system proposed greatly improves the performance of thermal camera odometry (compared to direct 'out-of-the-box' use of VO algorithms with thermal data), as indicated in Table I. It is important to observe that errors for thermal imaging without the NUC compensation vary significantly between different tests, depending on what part of the run (i.e. during a curve or a straight segment) NUC was performed. The results discussed above use $\tau_{\text{nms}} = 50$ for visual spectrum data and $\tau_{\text{nms}} = 13$ for thermal data. This choice is based on the experimental evaluation shown in Figure 10, which present errors (in distance) for TC+RE for varying τ_{nms}, considering



(a) Distance error growth.



(b) Heading error growth.

Fig. 9: Experimental errors for the different methods at different travelling distances from the origin. The center dots represent the median of each method and the triangles correspond to the maximum and minimum values obtained in the experiment. The expanded acronyms for each different modality are given in the beginning of Section V-B.

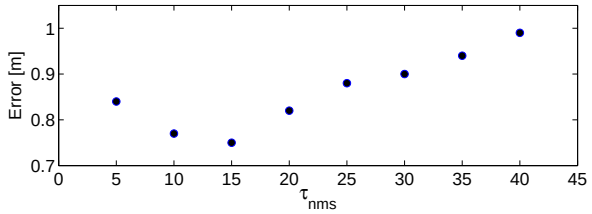


Fig. 10: TC+RE odometry errors for different values of τ_{nms}. In this evaluation, the errors were generated for 10 meters segments (as opposed to also 20, 50, and 100 meters).

the 10 meter segments.

VI. Conclusions

This paper presented a new method for using thermal imaging as an alternative to conventional imaging for monocular visual odometry. The results show that the proposed method can achieve similar results to those with a visible spectrum

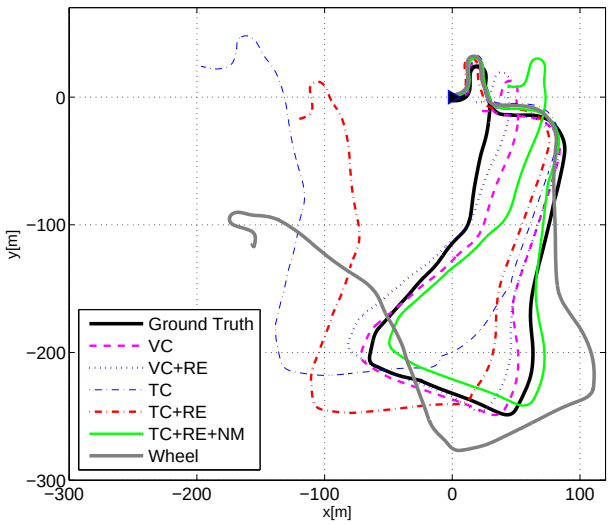


Fig. 11: 2-D illustration comparing different types of odometry. The loop shown in this figure corresponds to the loop shown in Figure 6. The expanded acronyms for the legends are given in the beginning of Section V-B.

camera using traditional algorithms, with the significant advantage of being able to operate effectively at night time. The method combines efficient NUC management with road lane estimation in order to compensate for the inherent challenges of thermal infrared data, namely the unavoidable occurrence of data interruptions and the low signal-to-noise ratio. The proposed method could be implemented as a standalone odometry system, or potentially combined with conventional visual odometry or other approaches to enhance accuracy. One limitation of the proposed system is that it is tailored to applications where the camera is mounted on a vehicle or ground robot. In different scenarios (e.g. aerial vehicles or handheld devices), the NUC manager would have to be modified, although most of the underlying concepts (time since and temperature change since last NUC, current rotation) are still applicable.

Future work includes investigating different image pre-processing methods for feature matching in the thermal domain. In general, to increase saliency in thermal images, pre-processing methods can enhance substantially the image quality and make it easier to extract features, enhance segmentation, etc [34], [35]. However, in the case of temporal data, one issue with pre-processing methods that modify the image on a per-frame basis and that are image-statistics dependent (e.g., anisotropic diffusion, histogram equalization, bilateral filters, etc) is that modifications with different parameters across consecutive frames can be counterproductive for feature matching and description purposes. A small difference in the local statistics of a feature point (caused by different filtering parameters for different images) between consecutive frames can decrease the performance of the algorithm. Our own brief investigation indicated that the use of anisotropic diffusion, bilateral filters or histogram equalization did not

bring improvements to the feature point matching, although there are conflicting reports in the literature discouraging [36] or encouraging [37] the use of pre-processing in the visible spectrum.

REFERENCES

[1] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1. IEEE, 2004, pp. I–652.

[2] D. M. Helmick, Y. Cheng, D. Clouse, L. H. Matthies, and S. I. Roumeliotis, "Path following using visual odometry for a mars rover in high-slip environments," in *Aerospace Conference, 2004. Proceedings. 2004 IEEE*, vol. 2. IEEE, 2004, pp. 772–789.

[3] E. Nilsson, C. Lundquist, T. B. Schön, D. Forslund, and J. Roll, "Vehicle motion estimation using an infrared camera," in *Proceedings of the World Congress of the International Federation of Automatic Control, Milan, Italy*, 2011.

[4] T. B. Schon and J. Roll, "Ego-motion and indirect road geometry estimation using night vision," in *Intelligent Vehicles Symposium, 2009 IEEE*. IEEE, 2009, pp. 30–35.

[5] B. M. Kitt, J. Rehder, A. D. Chambers, M. Schonbein, H. Lategahn, and S. Singh, "Monocular visual odometry using a planar road model to solve scale ambiguity," in *Proceedings of European Conference on Mobile Robots*, September 2011.

[6] O. Riou, S. Berrebi, and P. Bremond, "Non uniformity correction and thermal drift compensation of thermal infrared camera," in *Proc. of SPIE Vol*, vol. 5405, 2004, p. 295.

[7] M. Agrawal and K. Konolige, "Real-time localization in outdoor environments using stereo vision and inexpensive gps," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3. IEEE, 2006, pp. 1063–1068.

[8] C. Dornhege and A. Kleiner, "Visual odometry for tracked vehicles," in *In Proc. of the IEEE Int. Workshop on Safety, Security and Rescue Robotics (SSRR)*, 2006.

[9] H. Wang, K. Yuan, W. Zou, and Q. Zhou, "Visual odometry based on locally planar ground assumption," in *Information Acquisition, 2005 IEEE International Conference on*. IEEE, 2005, pp. 6–pp.

[10] O. Saurer, F. Fraundorfer, and M. Pollefeys, "Homography based visual odometry with known vertical direction and weak manhattan world assumption," *ViCoMoR 2012*, p. 25.

[11] X. Lu and R. Manduchi, "Detection and localization of curbs and stairways using stereo vision." in *ICRA*. IEEE, 2005, pp. 4648–4654.

[12] F. Fraundorfer and D. Scaramuzza, "Visual odometry: Part ii: Matching, robustness, optimization, and applications," *Robotics & Automation Magazine, IEEE*, vol. 19, no. 2, pp. 78–90, 2012.

[13] S. Vidas and S. Sridharan, "Hand-held monocular SLAM in thermal-infrared," in *International Conference on Automation, Robotics, Control and Vision (ICARCV)*, 2012.

[14] W. Maddern and S. Vidas, "Towards robust night and day place recognition using visible and thermal imaging," in *Robotics Science and Systems (RSS): Workshop on Alternative Sensing Modalities*, 2012.

[15] S. Sobarzo, J. Pezoa, and S. Torres, "Real-time kalman filtering for nonuniformity correction in infrared focal-plane arrays," *Pattern Recognition, Image Analysis and Applications*, vol. 3773, pp. 752–761, 2004.

[16] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, and T. Wark, "An exploration of feature detector performance in the thermal-infrared modality," in *Digital Image Computing Techniques and Applications (DICTA)*, 2011, p. 217224.

[17] A. Kumar, S. Sarkar, and R. P. Agarwal, "Correcting infrared focal plane array sensor non uniformities based upon adaptive filter," 2006, pp. 1537–1540.

[18] H. Kong, J.-Y. Audibert, and J. Ponce, "Vanishing point detection for road detection," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 96–103.

[19] Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 5, no. 4, pp. 309–318, 2004.

[20] H. Kong, J.-Y. Audibert, and J. Ponce, "General road detection from a single image," *Image Processing, IEEE Transactions on*, vol. 19, no. 8, pp. 2211–2220, 2010.

[21] P. Borges and P. Moghadam, "Combining motion and appearance for scene segmentation," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, May 2014, pp. 1028–1035.

[22] I. Vanhamel, I. Pratikakis, and H. Sahli, "Multiscale gradient watersheds of color images," *Image Processing, IEEE Transactions on*, vol. 12, no. 6, pp. 617–626, 2003.

[23] G. Farneback, "Fast and accurate motion estimation using orientation tensors and parametric motion models," in *Pattern Recognition, 2000 IEEE International Conference on*, vol. 1. IEEE, 2000, pp. 135–139.

[24] ——, "Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field," in *Computer Vision (ICCV), 2001 IEEE International Conference on*, vol. 1. IEEE, 2001, pp. 171–177.

[25] C. Kondermann, D. Kondermann, B. Jähne, and C. Garbe, "An adaptive confidence measure for optical flows based on linear subspace projections," *Pattern Recognition*, pp. 132–141, 2007.

[26] N. Nourani-Vatani, P. V. K. Borges, J. M. Roberts, and M. V. Srinivasan, "On the use of optical flow for scene change detection and description," *Journal of Intelligent & Robotic Systems*, pp. 1–30, 2013.

[27] C. Fraley and A. E. Raftery, "How many clusters? which clustering method? answers via model-based cluster analysis," *The computer journal*, vol. 41, no. 8, pp. 578–588, 1998.

[28] V. S. Bottazzi, P. V. Borges, and J. Jo, "A vision-based lane detection system combining appearance segmentation and tracking of salient points," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 443–448.

[29] S. Vidas, R. Lakemond, S. Denman, C. Fookes, S. Sridharan, and T. Wark, "A Mask-Based approach for the geometric calibration of Thermal-Infrared cameras," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 1625–1635, 2012.

[30] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, "Ros: an open-source robot operating system," in *ICRA workshop on open source software*, vol. 3, no. 3.2, 2009.

[31] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, 2010, pp. 486–492.

[32] A. Kelly, "Linearized error propagation in odometry," *The International Journal of Robotics Research*, vol. 23, no. 2, pp. 179–218, 2004.

[33] A. E. Johnson, S. B. Goldberg, Y. Cheng, and L. H. Matthies, "Robust and efficient stereo feature tracking for visual odometry," in *IEEE International Conference on Robotics and Automation*, May 2008, pp. 39–46.

[34] P. Buddharaju, I. T. Pavlidis, P. Tsiamyrtzis, and M. Bazakos, "Physiology-based face recognition in the thermal infrared spectrum," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 613–626, 2007.

[35] C. Ibarra-Castanedo, D. Gonzalez, M. Klein, M. Pilla, S. Vallerand, and X. Maldague, "Infrared image processing and data analysis," *Infrared physics & technology*, vol. 46, no. 1, pp. 75–83, 2004.

[36] P. Campos, M. Forkin, and X. Xu, "Evaluating gabor preprocessing for sift-based ocular recognition," in *Proceedings of the 49th Annual Southeast Regional Conference*. ACM, 2011, pp. 365–366.

[37] C. Tsai, "Effects of 2-d preprocessing on feature extraction," *2D Imaging Project Report, Department of Electrical Engineering, Stanford University*, 2008.

Paulo Vinicius Koerich Borges received the B.E. and M.Sc. degrees in electrical engineering from Federal University of Santa Catarina (UFSC), Brazil, in 2002 and 2004, respectively. In 2007 he received his Ph.D. degree from Queen Mary, University of London (QMUL). From 2007 to 2008, he was as post-doc researcher at QMUL. Since 2009, he has been in the Autonomous System Laboratory at CSIRO in Brisbane, Australia, where he is currently a Senior Research Scientist and leads the Robotics Perception Team. His current research focuses on visual-based robot localization, obstacle detection and tracking, and multi-sensor information fusion. In 2012-13, he held a visiting scientist appointment at ETH Zurich in Switzerland. Dr. Borges is also an Adjunct Senior Lecturer with the School of Information Technology and Electrical Engineering at the University of Queensland, Australia.

Stephen Vidas Stephen Vidas received the B.Eng. (First Class Hons.) degree in electrical engineering from the Queensland University of Technology (QUT), Australia, in 2009, as part of the Deans Scholars Program, and the Ph.D. degree from QUT in 2014, with a thesis entitled Handheld 3-D Thermography Using Range Sensing and Computer Vision. During the development of this paper he held a Post-Doctoral Fellowship with the Intelligent Robotics Laboratory, Nanyang Technological University, Singapore. He currently works as a freelance Software Engineer in Perth, Australia. His research interests include camera calibration, computer vision, SLAM, 3D modeling, thermography, and robot control.