IEEE Sensors Journal, Volume: 15, Issue 11, November 2015
DOI: 10.1109/JSEN.2015.2456337

# Thermal Stereo Odometry for UAVs

Tarek Mouats, Nabil Aouf, *Member, IEEE*, Lounis Chermak, and Mark A. Richardson

***Abstract*— In the last decade, visual odometry (VO) has attracted significant research attention within the computer vision community. Most of the works have been carried out using standard visible-band cameras. These sensors offer numerous advantages but also suffer from some drawbacks such as illumination variations and limited operational time (i.e., daytime only). In this paper, we explore techniques that allow us to extend the concepts beyond the visible spectrum. We introduce a localization solution based on a pair of thermal cameras. We focus on VO and demonstrate the accuracy of the proposed solution in daytime as well as night-time. The first challenge with thermal cameras is their geometric calibration. Here, we propose a solution to overcome this issue and enable stereopsis. VO requires a good set of feature correspondences. We use a combination of Fast–Hessian detector with for Fast Retina Keypoint descriptor for that purpose. A range of optimization techniques can be used to compute the incremental motion. Here, we propose the double dogleg algorithm and show that it presents an interesting alternative to the commonly used Levenberg–Marquadt approach. In addition, we explore thermal 3-D reconstruction and show that similar performance to the visible-band can be achieved. In order to validate the proposed solution, we build an innovative experimental setup to capture various data sets, where different weather and time conditions are considered.**

***Index Terms*— Visual odometry, thermal imagery, infrared, geometric calibration, motion estimation, stereo vision system, 3D reconstruction.**

## I. Introduction

VISUAL odometry (VO) has received significant attention in the computer vision community. In the last decade, numerous contributions have been made to further improve the algorithms of the different building blocks of VO. However, most of these efforts have focused on *standard* visible-band cameras. In this work, we explore techniques that allow us to extend the VO concepts beyond the visible spectrum. We introduce a localisation solution based on a pair of thermal cameras in order to broaden the field of application of most computer vision algorithms. Here, we focus on VO and demonstrate the accuracy of the proposed system in day-time as well as night-time conditions. In contrast to visible-band cameras, infrared provide inherent robustness against illumination variations known to induce serious challenges to standard cameras [1]. Thermal sensors capture temperature variations within the scene and hence can be used in low-light and dark environments without requiring additional lighting. Despite the aforementioned advantages, thermal cameras also come with some challenges such as low image resolution, relatively low signal to noise ratios and non-uniformity noise. These can render some computer vision algorithms non-usable. The first challenge when using thermal cameras as a stereo setup is their geometric calibration. This procedure has been shown rather complicated for thermal modality. We studied different calibration strategies and propose a simple solution to overcome this issue.

Visual odometry, can be decomposed into a series of subsequent tasks [2]. In order to obtain accurate estimates of the camera motion, VO requires a set of good feature correspondences. In general, *classical* feature detectors/ descriptors or tracking algorithms are used with visible cameras. However, as illustrated in [3], different results may be expected when used with thermal imagery. For this reason, we conducted a performance analysis of common feature detectors/descriptors in thermal imagery where a benchmark similar to [4] and [5] was used as explained in Section IV-A. A range of optimisation techniques can be used to compute the incremental camera motion. We investigate the behaviour of various approaches namely Gauss-Newton, Levenberg-Marquadt and the Double Dogleg. We show that the latter presents an appealing alternative compared to the former. Thermal 3D reconstruction is also explored within the scope of this work, through the adaptation of existing computer vision algorithms.

The rest of the paper is organised as follows: Section II details the related works. The thermal geometric calibration is presented in Section III. The different sub-tasks of the visual odometry pipeline are introduced in Section IV. The experimental setup, datasets and results are discussed in detail in Section V. Conclusive remarks and highlights into future works are provided in Section VI.

## II. Related Works

There have been various works investigating visual odometry. Surveys and tutorials have been produced to explain the different sub-tasks that are involved [2], [6], [7]. Some works have been conducted on navigation problems beyond the visible spectrum. Jung et al. proposed an algorithm for egomotion estimation from a monocular infrared camera [8]. They used the focus of expansion for feature matching and reprojection errors for egomotion estimation. However, they did not include estimated trajectories with their results. A similar approach was proposed in [9] using

Manuscript received May 20, 2015; revised July 8, 2015; accepted July 8, 2015. Date of publication July 14, 2015; date of current version September 4, 2015. The associate editor coordinating the review of this paper and approving it for publication was Prof. Jun Ohta.

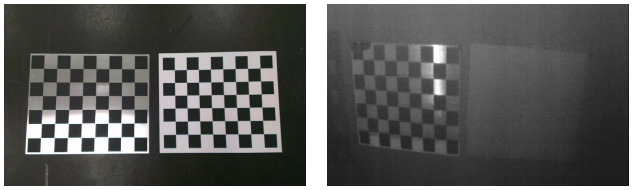
The authors are with the Centre for Electronic Warfare, Cranfield University, Defence Academy of the United Kingdom, Shrivenham SN6 8LA, U.K. (e-mail: t.mouats@cranfield.ac.uk; l.chermak@cranfield.ac.uk; n.aouf@cranfield.ac.uk; m.a.richardson@cranfield.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSEN.2015.2456337

two cameras - thermal and visible. A handover mechanism was introduced but each camera was used separately for monocular simultaneous localisation and mapping (SLAM). In our previous work [10], we made the first attempt to estimate the egomotion of a vehicle from a pair of heterogeneous cameras where promising results were obtained. A study was conducted in [11] to investigate the suitability of various types of infrared cameras deployed on unmanned ground vehicles for night-time stereo vision. However, the outcomes of this study have become obsolete due to recent advances in infrared technology. A thermal camera was used in [12] to enhance the estimates of the vehicle egomotion and indirectly the road geometry in 3D. The authors fused proprioceptive sensors with thermal imagery in order to improve the egomotion estimation. However, limited results were shown in their experimental results. Rankin et al. [13] carried out several works with a thermal stereo vision system on board an autonomous ground vehicle. Notably, they investigated the problem of dense depth maps from thermal stereo. The latter was used in the context of human and vehicle detection and classification. However, VO was not investigated in their work. A similar work was also conducted in [14]. An investigation was carried out in [15] where various sensor configurations were considered for pedestrian detection. These encompass colour, thermal and multi-spectral vision systems. The latter were shown to provide better detection accuracy.

Geometric calibration is an important prerequisite for using a set of cameras in navigation applications. It is a well-studied problem in the computer vision community. However, most of the works address visible-band cameras. Lately, with the proliferation of infrared sensors, there have been some efforts with thermal cameras [13], [16]–[20]. These resulted in a variety of calibration targets and tools with varying accuracies and manufacturing difficulties. The latter generally give an indication of the time and cost required to build the calibration board. A common observation with thermal cameras is that their lenses cause relatively large radial distortion due to a design trade-off for which the lenses are optimised for radiometric resolution rather than geometric. Zelek et al. attempted thermal camera calibration nearly a decade ago [19]. Luhmann et al. [16] produced a calibration board using self-adhesive foil augmented with various targets. They showed better performance than those obtained with an elaborate calibration board - a planar test-field. A variant of this test-field board was used to calibrate a multi-modal setup in [21] and a thermal camera in [22]. A complex target was manufactured in [13] as calibration was unsuccessful with simpler boards. Engstrom et al. relied on the difference of thermal emissivity to build calibration boards [17]. Essentially, these were made using aluminium plates with high emissivity materials taped on. A similar solution was introduced in [23] where a saddle point detector was used. Vidas et al. proposed a different approach coined *mask-based* calibration [18]. It was suggested as alternative to the classical heated chessboard (with flood lamps) as used in [24] and [25]. The board was designed as a mask that is placed in front of a heat source (e.g. monitor) to improve the contrast.



(a) (b)

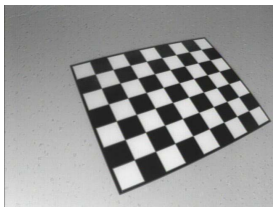
Fig. 1. Aluminium and paper chessboards. (a) Captured with visible-band camera. (b) Captured with thermal camera without heating.

## III. Geometric Calibration of Thermal Cameras

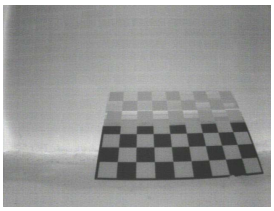
In order to use two infrared cameras in a stereo setup, their intrinsic and extrinsic camera parameters need to be determined accurately. This is usually done through the geometric calibration of the system which allows the correction of lens distortion and the determination of the rigid motion transformation relating the left and right cameras. The *classical* printed chessboard pattern cannot be used in this context due to its uniform temperature which yields a blurry image (Fig. 1b). There have been ongoing efforts to tackle this issue resulting in a variety of calibration tools and targets with various accuracies and manufacturing difficulties. Depending on the end application, the calibration accuracy may be traded against ease of manufacturing. Here, the aim is to use the thermal stereo vision system in a localisation framework and therefore low accuracies are not acceptable. For this reason, we investigate different approaches to seek an acceptable solution with respect to accuracy and time scale. The first approach (Section III-A) is based on the method generally adopted for visible-band imagery through the use of a *special calibration chessboard* [26]. The second approach (Section III-B) is based on the method proposed in [18]. The calibration was carried out both manually and automatically. The former means that the user is asked to intervene, particularly in the extraction of the chessboard corners, whereas the latter does not require any user interaction. Both approaches were considered to allow a fair comparison with the *mask-based* approach where the process is fully independent from user interactions.

### A. Special Chessboard Calibration

In order to use the standard calibration tools, generally based on chessboards, we had to produce one that allows the extraction of the corners in thermal imagery. We studied different options such as heating a printed chessboard but without success. As outlined in [18], the major issue when using heated printed chessboards is their relatively short cooling time i.e. the contrast created by heating the chessboard does not last long enough. Therefore, we had to look at the properties of different materials and their behaviour in thermal imagery. A readily available material that attracted our attention due to its reflective properties is aluminium. It was reported to have a reflectance of around 95% in the far infrared [27]. Consequently, we produced an aluminium-based chessboard pattern by coating a polished aluminium plate (1.6*mm* thick) with matt black squares (Fig. 1a). This process was carried out by a specialised manufacturer to guarantee high accuracy.

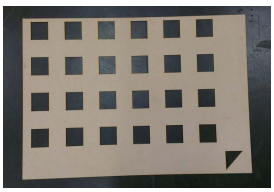


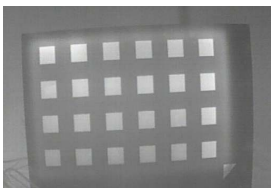
(a)



(b) (c)

Fig. 2. Good vs bad examples of calibration images using the *cold sky effect*. (a) Good image, (b) (c) Bad images.





(a) (b)

Fig. 3. Images of the *mask* calibration board. (a) Visible-band. (b) Thermal.

In order to obtain a reasonable contrast in thermal imagery, we exploited the reflectance property of aluminium and the *cold sky effect*. This is done by setting the calibration board outdoors in a position where it reflects the cold sky (Fig. 2). Some precautions need to be taken though when setting up the board to avoid reflecting hot objects (Fig. 2b and 2c) and yield crisp noise-free calibration images (Fig. 2a). This is largely due to the fact that, in thermal imagery, aluminium acts as a mirror in the visible-band. As shown in Fig. 2a, setting the calibration pattern outdoor in a clear day allows the acquisition of remarkable images where a sharp contrast between the chessboard elements can be observed. This contrast comes from the temperature difference between the reflected cold sky and the adhesive black squares printed on the aluminium plate which have different emissivity properties (with respect to aluminium).

### *B. Mask-Based Calibration*

We reproduced a similar calibration pattern to the one introduced in [18] on an A2 board where $40 \times 40mm$ squares were professionally laser-cut. Instead of the cardboard, we opted for Medium-Density Fibreboard (MDF), as shown in Fig. 3. The motivation was that MDF provides a better thermal isolation than cardboard when in contact with a source of heat. It requires relatively longer to heat-up and therefore allows more time for acquisition of the calibration images.

### *C. Comparison*

In addition to the produced calibration boards, we explored different software options. The first tool is the well-documented Caltech Camera Calibration Toolbox proposed by **Bouguet** [26]. The second is the Automatic Multi-Camera Calibration (**Amcc**) toolbox [28] (adaptation of Bouguet's software). It allows automatic extraction of the chessboard corners and implements automated monocular and stereo calibration procedures [28]. The third tool is from Mathworks who issued a calibration tool (**calibrator**) in their 2014b MATLAB release (also an adaptation of Bouguet's toolbox). Our aluminium-based chessboard target was used with these tools. The *mask-based* approach was tested using the chessboard (**mm-chess**) and the mask pattern (**mm-mask**).

TABLE I
STEREO PARAMETERS OBTAINED USING THE STUDIED CALIBRATION APPROACHES

| Parameters | $T = [t_x t_y t_z]$ | $R = [r_x r_y r_z]$ |
|---|---|---|
| Bouguet | [-267.8445 0.4383 0.8668] | [-0.0154 -0.0117 0.0037] |
| Amcc | [-267.8441 0.4702 0.5850] | [-0.0145 -0.0121 0.0037] |
| Calibrator | [-267.8802 0.1741 -4.1270] | [0.0069 0.0179 -0.0028] |
| mm-chess | [-267.1987 0.1666 -10.2591] | [0.0126 0.0141 -0.0027] |
| mm-mask | [-270.5508 -3.7881 -5.9573] | [0.0009 0.0129 -0.0025] |

It uses an automatic corner extraction. In contrast to the others, an automatic selection of the best set of calibration images is implemented. The mean reprojection error (MRE) is used to assess the quality of camera calibration. It is computed as follows:

$$\text{MRE} = \frac{1}{K \times M} \sum_{i=1}^{M} \sum_{j=1}^{K} \|\mathbf{x}(i,j) - \hat{\mathbf{x}}(i,j)\| \qquad (1)$$

where $\mathbf{x}$ are 2D points extracted from the calibration images and $\mathbf{X}$ their 3D location. $\hat{\mathbf{x}}$ are the reprojected 2D points using the estimated camera calibration matrices. $K$ represents the total number of points and $M$ the total number of images. In order to set all the algorithms on an equal footing, we used the same set of calibration images ($M = 22$). The results are summarised in Tables I and II for monocular and stereo parameters, respectively.

In Table II, $(f_x, f_y)$ represent the focal lengths, $(u_0, v_0)$ is the centre point. $k_1$ and $k_2$ correspond to the radial distortion coefficients whereas $p_1$ and $p_2$ are the tangential coefficients. In addition to the MRE, we also use the *epipolar geometry* to evaluate the accuracy of the stereo calibration [29]. More specifically, we plot the epipolar lines of corresponding points in two stereo images in order to analyse the epipolar errors. The latter corresponds to the distance between the feature location and the epipolar line corresponding to its match. In the ideal case (perfect calibration), this error should be equal to zero. Corresponding features in the left and right images should be crossed by the same line. Table III shows the epipolar errors for the studied calibration algorithms. Error_1-2 and Error_2-1 correspond to the epipolar errors when using the left image and right images as origin. Fig. 4 shows a stereo image pair where three epipolar lines corresponding to three feature locations are plotted.

Comparing the epipolar errors in Table III, and looking at Fig. 4, we can state that **calibrator** (Mathworks 2014)

TABLE II

CALIBRATION RESULTS

| Parameters | Bouguet | | Amcc | | Calibrator | | mm-chess | | mm-calib | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Left | Right | Left | Right | Left | Right | Left | Right | Left | Right |
| $f_x$ | 522.1570 | 521.8004 | 522.3360 | 521.7557 | 516.8043 | 514.1008 | 524.8815 | 521.3998 | 535.1738 | 528.4455 |
| $f_y$ | 519.9779 | 518.9811 | 520.0211 | 518.7962 | 510.2553 | 507.6046 | 521.3927 | 517.0899 | 534.4184 | 527.9880 |
| $u_0$ | 317.9416 | 319.5849 | 317.8889 | 319.7698 | 316.5400 | 321.5900 | 326.1222 | 328.9086 | 313.8362 | 316.6676 |
| $v_0$ | 261.3959 | 248.3533 | 261.0548 | 248.4461 | 253.3017 | 244.8522 | 253.4161 | 242.1563 | 246.1570 | 243.9634 |
| $k_1$ | -0.2850 | -0.2918 | -0.2851 | -0.2901 | -0.2787 | -0.2949 | -0.2757 | -0.3157 | -0.3240 | -0.3109 |
| $k_2$ | 0.0859 | 0.1014 | 0.0864 | 0.0966 | 0.0739 | 0.1250 | -0.0496 | 0.1692 | 0.2335 | 0.1967 |
| $p_1$ | -0.0035 | -0.0031 | -0.0033 | -0.0032 | -0.0017 | -0.0035 | -0.0019 | -0.0021 | -0.0013 | -0.0037 |
| $p_2$ | 0.0012 | 0.0011 | 0.0011 | 0.0011 | 0.0029 | $2.16e^{-04}$ | 0.0017 | $2.18e^{-04}$ | 0.0035 | $-5.0e^{-04}$ |
| **MRE** | **0.1749** | **0.1884** | **0.1679** | **0.1786** | **0.2205** | **0.2172** | **1.2399** | **1.2432** | **0.5323** | **0.7364** |

TABLE III

EPIPOLAR ERRORS OF THE COMPARED CALIBRATION APPROACHES

| Error | Bouguet | Amcc | Calibrator | mm-chess | mm-mask |
|---|---|---|---|---|---|
| Error_1-2 | 0.7010 | 0.6906 | 9.7247 | 0.8506 | 1.2072 |
| Error_2-1 | 0.7088 | 0.6983 | 9.6560 | 0.8578 | 1.2150 |

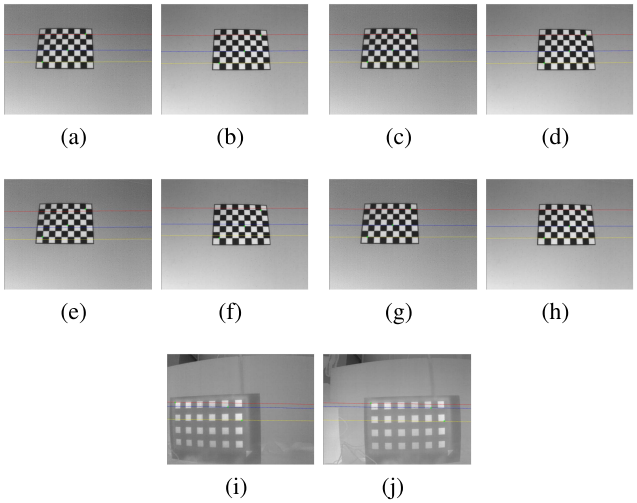



Fig. 4. Thermal stereo calibration accuracy. Lines of the same colour should pass through the same point in the left and right images. (a)(b) **Bouguet**, (c)(d) **Amcc**, (e)(f) **calibrator**, (g)(h) **mm-chess**, and (i)(j) **mm-mask**.

provides the least reliable calibration parameters of the thermal vision system. This indicates that obtaining low MRE during the calibration procedure (0.2205) does not necessarily yield to accurate calibration parameters. Vidas' algorithm (**mm-chess** & **mm-mask**) provided better error rate with the chessboard pattern than with the custom made mask pattern. **Bouguet** software and **Amcc** provided similar outcomes. The main difference between the two algorithms is that **Amcc** was designed to alleviate the user interaction during the calibration process. Indeed, in contrast to **Bouguet** software, **Amcc** extracts the chessboard corners and the calibration parameters automatically [28]. Looking at the MRE and the epipolar error figures (Tables II and III), we can conclude that the calibration parameters provided by **Amcc** algorithm using our calibration board are the most accurate. The monocular and stereo calibration parameters estimated using the **Amcc** approach using our calibration board are used in the remainder of this work.

## IV. VISUAL ODOMETRY

### A. Feature Extraction, Description and Matching

At each time step, the visual odometry pipeline is fed a pair of thermal images $(t-1, t)$. The first step in VO is the extraction of stable and tractable features from images. We carried out a performance analysis[1] of popular feature detection and description algorithms in order to study their behaviour in the thermal modality. More specifically, we studied the repeatability and matching scores of feature detectors namely DoG [30], Fast-Hessian [31], FAST [32], Harris [33], Shi-Tomasi [34] and CenSurE [35]. In addition, we evaluated the *recall*/1-*precision* curves of description algorithms namely SIFT [30], SURF [31], LIOP [36], ORB [37], BRISK [38] and FREAK [39]. For that purpose, we generated a thermal dataset encompassing various environments and image transforms. Building on this analysis, we adopted a mixed feature detection/description scheme. First, we detect Fast-Hessian features from the acquired images. These are then described using the binary FREAK descriptor. FREAK was preferred to other descriptors as it provided good performance at a fraction of their computational cost. One of the advantages when using binary descriptors is the gain in matching speed. Indeed, the bit string descriptors can be matched using simple XOR and POPCNT instructions from SSE4.2 [40]. In addition, the coarse-to-fine approach of the FREAK descriptor allows to further reduce the computational burden. It was claimed that this coarse-to-fine approach allows to discard 90% of the candidates therefore accelerating the matching process [39]. Overall, Fast-Hessian provided the best scores in our performance analysis. Furthermore, it was adopted in this work for its relatively low computational requirements.

The matching process is carried out in a loop fashion [10]. This ensures more reliable and accurate feature correspondences. Note that to obtain a decent number of extracted features with Fast-Hessian, the *standard* thresholds used with visible-band imagery need to be considerably lowered. Notably, the Hessian threshold has to be set to as low as

[1]Currently under review in the International Journal of Computer Vision.

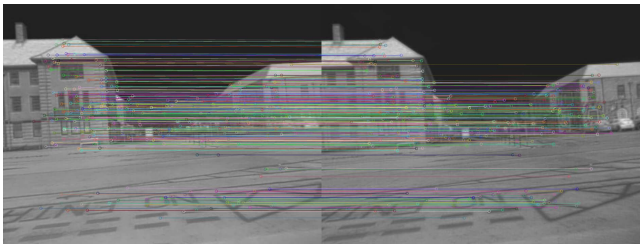


Fig. 5. Illustration of stereo matching using Fast-Hessian features and FREAK descriptors.

5 to yield a decent number of features. Fig. 5 shows a stereo matching example using Fast-Hessian features described using FREAK.

### *B. Motion Estimation*

The camera motion is estimated from a set of 3D-2D feature matches by minimising the sum of the reprojection errors. A frame to frame approach is adopted where the six motion parameters defining the inter-frame rotation and translation are computed. Traditionally, Gauss-Newton (GN) and Levenberg-Marquadt (LM) [41], [42] optimisation schemes are used. Here, we explore another algorithm, called the Double Dogleg (DDL) [43], and show that it presents an interesting alternative. Let us consider the objective function formulated in Eq. 2. Let $f$ be the projection function that maps the 3D points $\mathbf{X}_{\mathbf{pr}}^{(i)}$ of the previous frame (obtained through the parameters vector $p$) to the 2D coordinates $\mathrm{x}_{c\mathrm{L}}^{*}$ and $\mathrm{x}_{c\mathrm{R}}^{*}$ in the current left and right images, respectively.

$$\min \sum_{i=1}^{N} \left\| \mathrm{x}_{c\mathrm{L}}^{*} - f\left(\mathbf{X}_{\mathbf{pr}}^{(i)}; p\right) \right\|^{2} + \left\| \mathrm{x}_{c\mathrm{R}}^{*} - f\left(\mathbf{X}_{\mathbf{pr}}^{(i)}; p\right) \right\|^{2} \tag{2}$$

In order to retrieve the motion parameters $p$, Eq.2 is minimised using GN, LM or DDL. To make the paper self-contained, we provide a short description of the LM and DDL algorithms. More details can be found in [41]–[43]. GN was implemented in [10] where a RANdom SAmpling Consensus (RANSAC) framework [44] was adopted for the outlier rejection scheme based on the reprojection errors.

*1) Levenberg-Marquadt:* also known as the damped least-squares (DLS) method, LM is one of the most used optimisation algorithms to solve non-linear least squares problems in computer vision applications. LM can be thought of as a combination of the steepest descent and the Gauss-Newton method. When the current solution is far from a local minimum, the algorithm behaves like a steepest descent method: slow, but guaranteed to converge. When the current solution is close to a local minimum, it becomes a Gauss-Newton method and exhibits fast convergence [45]. Similarly to Gauss-Newton, LM is an iterative procedure. Given an initial estimate of the motion parameters $p = (\phi, \theta, \psi, t_x, t_y, t_z)$ and a measurement vector $\mathbf{x} = (\mathbf{x_{cL}}, \mathbf{x_{cR}})$, it computes a series of parameters $p_k$ that converge to the minimiser $p_{optim}$ of $f$. The LM algorithm solves the equation given by

$$\left(J^{\mathrm{T}} J - I\mu\right) \delta_{LM} = J^{\mathrm{T}} r \tag{3}$$

**Algorithm 1**: Pseudo-Code for Levenberg-Marquardt

**Input**: $\mathbf{x_{cL}}, \mathbf{x_{cR}}, \mathbf{X_p}, \mathbf{p_0}$
**Output**: $\mathbf{p_{optim}}$
**Set**: $k = 0; maxIter = 100; \epsilon_1 = \epsilon_2 = 10^{-3}; \nu = 2;$
**algorithm:**
$\mathbf{p_k} = \mathbf{p_0}$; $\mathbf{A} = \mathbf{J}^{\mathrm{T}}\mathbf{J}$; $\mathbf{r_{p_k}} = \mathbf{x} - f(\mathbf{X_p}, \mathbf{p_k};)$ $\mathbf{g} = \mathbf{J}^{\mathrm{T}}\mathbf{r_{p_k}}$;
converged := $\|\mathbf{g}\| \leq \epsilon_1$;
**while** *(not converged) and* $(k \leq maxIter)$ **do**
  k:= k+1;
  **repeat**
    Solve $(\mathbf{A} + \mu\mathbf{I})\,\delta_{LM} = \mathbf{g}$;
    **if** $||\delta_{LM}|| \leq \varepsilon_2 ||\mathbf{p_k}||$ **then**
      converged:= $true$;
    **else**
      $\mathbf{p_{new}} := \mathbf{p_k} + \delta_{LM}$;
      $\rho := \frac{||\mathbf{r_{p_k}}||^2 - ||\mathbf{x} - f(\mathbf{X_p}, \mathbf{p_{new}})||^2}{\delta_{LM}^{\mathrm{T}}(\mu\delta_{LM} + \mathbf{g})}$;
      **if** $\rho > 0$ **then**
        $\mathbf{p_k} = \mathbf{p_{new}}$;
        $\mathbf{A} := \mathbf{J}^{\mathrm{T}}\mathbf{J}$; $\mathbf{r_{p_k}} := \mathbf{x} - f(\mathbf{X_p}, \mathbf{p_k})$; $\mathbf{g} := \mathbf{J}^{\mathrm{T}}\mathbf{r_{p_k}}$;
        converged:= $(||\mathbf{g}||_\infty \leq \varepsilon_1)$;
        $\mu := \mu \times max\left(\frac{1}{3}, 1 - (2\rho - 1)^3\right)$; $\nu$:= 2;
      **else**
        $\mu := \mu \times \nu$; $\nu := 2 \times \nu$;
  **until** $\rho > 0$ *or (converged)*;
$\mathbf{p_{optim}} := \mathbf{p_k}$;

where $I$ is the identity matrix, $J$ is the Jacobian matrix, $r \in \mathbb{R}^n$ is the residual vector and $(J^{\mathrm{T}}J)$ is an approximation of the Hessian matrix [46]. The strategy of altering the diagonal elements of $J^{\mathrm{T}}J$ is called damping and $\mu$ is the damping parameter. It allows LM to alternate between a slow descent approach when it is far from the minimum by increasing $\mu$ and a fast, quadratic convergence when being near the minimum's neighbourhood by decreasing $\mu$. In each iteration of LM, $\mu$ is adjusted to achieve the best possible update. Algorithm 1 illustrates the necessary steps for the LM method.

*2) Double Dogleg:* the DL algorithm, as well as its Double Dogleg variation [43], belong to the trust region optimisation methods. Similarly to the LM algorithm, the Dogleg combines Gauss-Newton with the steepest descent techniques. The Dogleg method was shown in [45] to provide similar results to LM at a fraction of the computational cost for the 3D reconstruction bundle adjustment formulation. It was also concluded that the trust region method could be used for constrained versions of bundle adjustment. In the Dogleg algorithm, the objective function is approximated by a quadratic model function L which is trusted only for points within a region of radius $\Delta$ centred at the current point. Finding the candidate step $\delta$ corresponds to solving the following constrained equation

$$\min_{\boldsymbol{\delta}} L(\boldsymbol{\delta}) \quad subject\ to \quad \|\boldsymbol{\delta}\| \leq \Delta \tag{4}$$

The radius of the trust region is of crucial importance. In practice, it is based on the success of the model in the previous approximations of the objective function. If the model is reliable, the radius is increased allowing the test of larger steps.

**Algorithm 2**: Pseudo-Code for Double Dogleg

**Input**: $\mathbf{x_{cL}}, \mathbf{x_{cR}}, \mathbf{X_p}, \mathbf{p_0}$
**Output**: $\mathbf{p_{optim}}$
**Set**: $k = 0$; $\Delta = 1$; $maxIter = 100$; $\varrho_1 = 0.15$; $\varrho_2 = 0.75$; $\chi_1 = 0.5$; $\chi_2 = 2$; $\epsilon_1 = \epsilon_2 = 10^{-4}$
**algorithm:**
$\mathbf{p} = \mathbf{p_0}$; $\mathbf{A} = \mathbf{J}^{\mathrm{T}}\mathbf{J}$; $\mathbf{r_{p_k}} = \mathbf{x} - f(\mathbf{X_p}, \mathbf{p_k})$; $\mathbf{g} = \mathbf{J}^{\mathrm{T}}\mathbf{r_{p_k}}$;
converged := $\|\mathbf{g}\| \leq \epsilon_1$; **while** *(not converged) and* $(k \leq maxIter)$ **do**
  $k = k + 1$;
  $\delta_{CP} = \frac{\|\mathbf{g}^{\mathrm{T}}\mathbf{g}\|}{\|\mathbf{g}^{\mathrm{T}}\mathbf{A}\mathbf{g}\|}\mathbf{g}$; $GN = false$;
  **repeat**
    **if** $\|\delta_{CP}\| \geq \Delta$ **then**
      $\delta_{DDL} = \frac{\Delta}{\|\delta_{CP}\|}\delta_{CP}$;
    **else**
      **if** *not GN* **then**
        $\delta_{GN} = \mathbf{A}^{-1}\mathbf{g}$; $GN = true$;
      **if** $\|\delta_{GN}\| \leq \Delta$ **then**
        $\delta_{DDL} = \delta_{GN}$;
      **else**
        $\delta_{DDL} = \delta_{CP} + \beta(\delta_{CP} - \delta_{GN})$;
    **if** $(\|\delta_{DDL}\| \leq \epsilon_2\|\mathbf{p}\|)$ **then**
      converged = true;
    **else**
      $\mathbf{p_{new}} = \mathbf{p_k} + \delta_{DDL}$;
      $\xi = \frac{\|\mathbf{x} - f(\mathbf{X_p}, \mathbf{p_k})\| - \|\mathbf{x} - f(\mathbf{X_p}, \mathbf{p_{new}})\|}{L(0) - L(\delta_{\mathrm{DDL}})}$;
      **if** $\xi > 0$ **then**
        $\mathbf{p} = \mathbf{p_{new}}$;
        $\mathbf{A} = \mathbf{J}^{\mathrm{T}}\mathbf{J}$; $\mathbf{r_p} = \|\mathbf{x} - f(\mathbf{X_p}, \mathbf{p_k})\|$; $\mathbf{g} = \mathbf{J}^{\mathrm{T}}\mathbf{r_p}$;
        converged := $(\|\mathbf{g}\| \leq \epsilon_1)$;
      update $\Delta$ using Eq. 5;
      converged := $(\Delta \leq \epsilon_2\|\mathbf{p}\|)$;
  **until** *(*$\xi > 0$*) or (converged)*;
$\mathbf{p_{optim}} = \mathbf{p}$;

If the model fails, the radius is reduced accordingly and Eq. 4 is solved again over a smaller region. Powell suggested that Eq. 4 can be decomposed into two line segments: follow the steepest descent direction to reach the Cauchy Point (CP) and then converge to the Newton Point (NP) through the Dogleg step. The Dogleg path intersects the trust region boundary at most once. The variation proposed by Dennis and Mei [43] introduces a bias towards the Gauss-Newton direction through an intermediate Newton step between the CP and the actual NP resulting in improved performance.

Once the new parameter vector is computed, the trust region has to be updated according to the gain ratio ($\xi_k$ in Algorithm 2) The update equations of the trust region are then given by

$$\Delta_{k+1} = \begin{cases} \chi_1 * \Delta_k & if \quad \xi_k < \varrho_1 \\ \Delta_k & if \quad \varrho_1 < \xi_k < \varrho_2 \\ \chi_2 * \Delta_k & if \quad \xi_k > \varrho_2 \end{cases} \tag{5}$$

where $0 < \chi_1 < 1 < \chi_2$ and $0 < \varrho_1 < \varrho_2 < 1$. Algorithm 2 highlights the different steps of the Double Dogleg technique where $\boldsymbol{\delta}_{\mathrm{CP}}$ and $\boldsymbol{\delta}_{\mathrm{GN}}$ correspond to the Cauchy point and the Gauss-Newton steps. $\beta$ must achieve $\|\boldsymbol{\delta}_{\mathrm{DDL}}\| = \Delta$. $\gamma = 0.8 * \kappa + 0.2$ ($\kappa \in [0, 1]$) represents an adjusting factor which sets the position of the intermediate NP step in the DDL algorithm.

## V. Experimental Results and Discussion

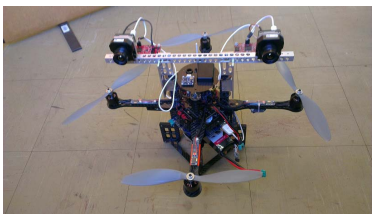
### A. Quadrotor UAV and the Vision System

In this section, we present the hardware components of the experimental setup. The quadrotor platform used in our experiments is the Pelican from Ascending Technologies.[2] It weighs $1.5\,kg$ and spans a diameter of $0.65\,m$ with a payload capacity of $0.65\,kg$. The quadrotor is equipped with an on-board Intel Core i7-3612QE Mastermind processor. It also comes with a GPS receiver which allows the acquisition of ground truth (GT) information when performing outdoor experiments.

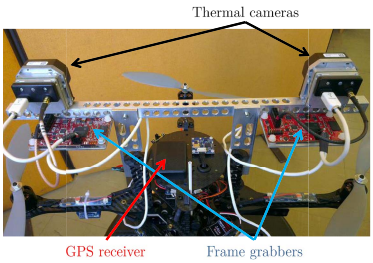
We installed two FLIR Tau2 infrared cameras on a front-looking stereo rig with a $27\,cm$ baseline i.e. distance between the left and right camera centres. The FLIR Tau2 cameras are based on uncooled Vanadium Oxide (VOx) micro-bolometers and capture thermal radiation in the spectral band $7.5 - 13.5\,\mu m$, which corresponds to the far infrared region. The cameras have a Noise-Equivalent Differential Temperature (NETD) below $50\,mK$. The $9\,mm$ lenses mounted on the camera offer a generous $69° \times 56°$ field of view when combined with the 17 micron detectors (pixel size). We installed two Sensoray frame grabbers, one for each camera, to stream 8/14-bit monochromatic images. This was due to the fact that the cameras do not provide a digital output without the use of another device from the manufacturer. The thermal vision system is able to capture stereo images of $640 \times 480$ pixels at approximately $30\,fps$. The synchronisation between the GPS device and the frame-grabbers was ensured using individual timestamps. This allowed the comparison of the estimated and measured trajectories i.e. VO vs GT. The quadrotor MAV and the thermal stereo system are shown in Fig. 6. In general, uncooled infrared cameras require thermal calibrations to prevent non-uniformities from building-up. In the case of the acquired TAU2 cameras, this involves presenting a material/shutter with uniform temperature (flat field) to the detector elements (pixels) - this operation is coined flat field correction (FFC). Therefore, a data interruption (up to 1 second) happens every time this operation is performed. However, FFCs are necessary to correct for temperature drift in the camera to ensure that pixel values (i.e. intensities) do not drift away from the real thermal radiance.

In visual odometry, such interruptions may cause estimation failures and are therefore unacceptable. On the other hand, thermal cameras that operate for long periods of time without FFC may also be detrimental. A solution to accommodate the inherent thermal imagery issue must be found. One way to alleviate this is to increase the time interval between consecutive FFC operations. However, there is no guarantee that they would not happen at crucial times e.g. turning MAV. For this reason, we adopted another solution where a

[2] http://www.asctec.de/uav-uas-drohnen-produkte/asctec-pelican/



(a)




(b)

Fig. 6. MAV quadrotor with the thermal stereo system. (a) Front view of the system. (b) Close-up rear-view of the system with labeled components.

TABLE IV
THERMAL STEREO SEQUENCES

| # | Sequence | travelled distance | Time of day | Weather conditions |
|---|---|---|---|---|
| 1 | **Seq1** | 687m | daytime (11am) | cloudy and cold |
| 2 | **Seq2** | 497m | daytime (10am) | foggy and relatively warm |
| 3 | **Seq3** | 701m | daytime (9am) | foggy and relatively warm |
| 4 | **Seq4** | 460m | daytime (11am) | cloudy and cold |
| 5 | **Seq5** | 882m | daytime (10am) | cloudy and warm |
| 6 | **Seq6** | 489m | night-time (9pm) | clear sky and very cold |
| 7 | **Seq7** | 399m | night-time (10pm) | cloudy and cold |
| 8 | **Seq8** | 314m | night-time (10.30pm) | cloudy and cold |
| 9 | **Seq9** | 655m | night-time (10pm) | cloudy and cold |

shutterless FFC-like operation is performed. This is possible through thermal calibration of the cameras. Using cold and warm blackbodies, non-volatile flat fields are created and stored in the memory of the camera. These will be used instead of the *standard* FFC which will have to be deactivated to ensure operational continuity.

### B. Thermal Sequences

Here, we introduce the stereo thermal dataset that was captured using the vision system described in Section V-A. Various scenarios with different weather and time conditions were considered. Table IV provides a description of the generated test sequences which include daytime as well as night-time scenarios and cover an overall distance of over $5\,km$. The top five rows in Table IV correspond to daytime sequences whereas the four bottom rows were captured during night-time. The generated dataset provides a variety of trajectories with different lengths ($300\,m - 900\,m$) and shapes. Ground truth information was recorded for each sequence to enable quantitative as well as qualitative evaluations.

(a) (b)

Fig. 7. Illustration of the ROI influence on the quality of the captured images. (a) ROI = full image. (b) Sky excluded from the ROI.

In order to obtain the best image quality, different camera settings and improvements were explored. This *pre-processing* stage is necessary to enhance the images that are fed to the visual odometry pipeline. The FLIR TAU2 cameras have various parameters that require tuning to obtain usable images (depending on the scene type). In particular, scaling the raw 14-bit data to 8-bit images causes a loss in dynamic range. This is aggravated when objects exhibiting a large temperature difference are imaged (e.g. cold sky and hot cars in a sunny day). Fig. 7 shows an example where setting an appropriate region of interest (ROI) for the automatic gain control algorithm (AGC) improves the image quality. In this example, the ROI was set to exclude the upper part of the image, which in many instances corresponds to the sky. This causes the AGC to ignore that part when computing the image histogram. Consequently, the dynamic range of the *interesting* scene content is increased.

In order to avoid illumination-variation-like problems, which are common with standard cameras, the automatic gain control threshold needs to be appropriately tuned. The value of this threshold determines how fast the AGC is allowed to vary when the scene content changes. Indeed, setting the value close to 255 causes abrupt intensity changes when a relatively hot object is introduced in the imaged scene. These sudden variations can fail the temporal feature matching and hence the trajectory estimates. If set too low (close to zero), the captured intensity values may not correspond to the real thermal radiance. Therefore, one must find a trade-off value to account for both aspects. We illustrate an example in Fig. 8 where we show the effect of introducing a hot object in the imaged scene for two threshold values (1 and 100). We can observe from Fig. 8a and 8b that introducing a hot object (the user's arm) in the scene has virtually no impact when the threshold is set to 1 as the AGC algorithm requires a longer period of time to adapt to the scene content. In contrast, setting a relatively higher threshold value (Fig. 8c and 8d) induces instantaneous intensity changes when a hot object enters/leaves the scene.

### C. Results and Discussion

*1) Stereo Thermal Odometry:* in this section we present the thermal visual odometry results corresponding to the dataset presented in Section V-B. For each sequence, we evaluate the quality of the trajectories estimated using the

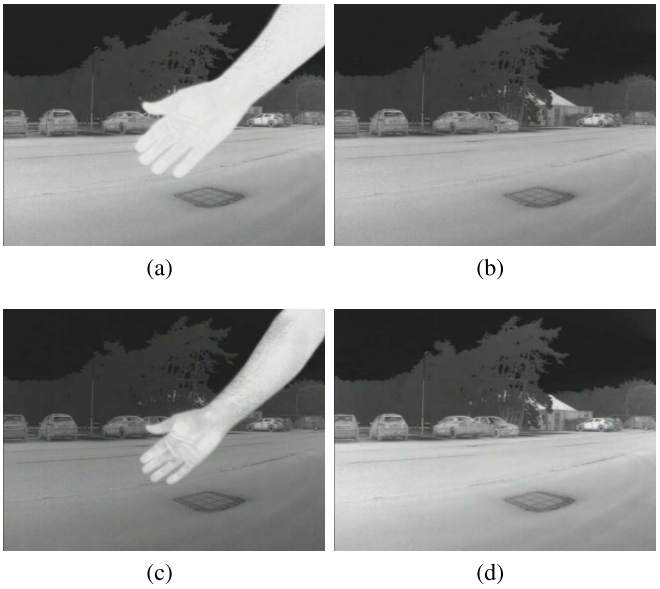


Fig. 8. Illustration of the influence of the AGC threshold on the quality of the captured images. Both sets of images were acquired in the same conditions. A similar time gap elapsed between introducing and removing the user's arm in both sequences (top and bottom row). (a) (b) correspond to a threshold value of 1 (c) (d) corresponds to a value of 100. Note that introducing a hot object (the human arm) in the scene has virtually no impact when the threshold is set to 1 as the AGC requires a longer period of time to adapt. In contrast, setting a relatively high threshold value induces instantaneous intensity changes when a hot object enters/leaves the scene.

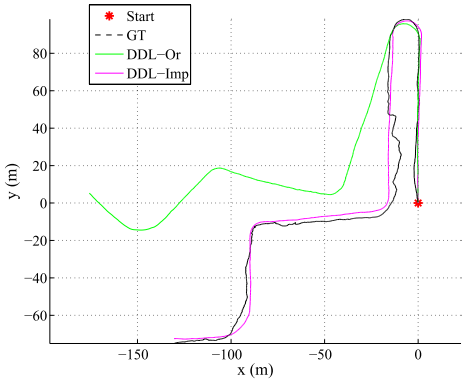



Fig. 9. Impact of the quality of the captured images on visual odometry (red star: starting point; black line: GT; green line: DDL on original images; magenta line: DDL on enhanced images).

different optimisation algorithms i.e. Gauss-Newton (GN), Levenberg-Marquadt (LM) and Double Dogleg (DDL). Each algorithm has been tuned separately. The selected parameters can be found in Algorithm 1 and Algorithm 2 for LM and DDL, respectively.

First, we illustrate the impact of image quality on visual odometry. As discussed in Section V-B, the captured images need to be pre-processed in order to enhance their quality. Fig. 9 shows the estimated trajectories using DDL from the original and improved images (contrast enhancement). AGC was performed on the lower part of the image to exclude the sky. The same parameters were used in both runs to ensure that only image quality varies. As it can be seen from Fig. 9, feeding unprocessed images to the VO pipeline can result in large errors in terms of trajectory estimation.

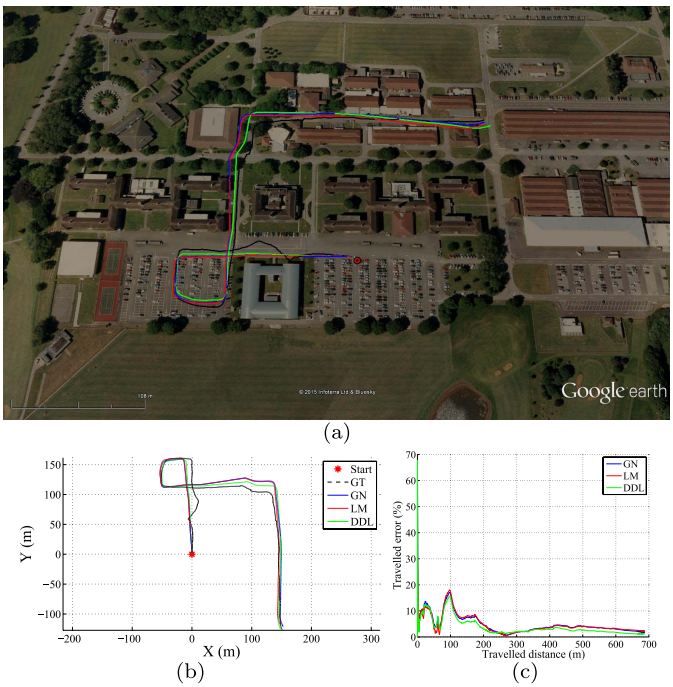



Fig. 10. Computed trajectories and travelled errors for Seq1. (a) Computed trajectories overlaid on Google Earth maps (red star: starting point; black line: GT; blue line: GN; red line: LM; green line: DDL) (b) computed trajectories (same legend applies) (c) corresponding travelled errors (same legend applies).

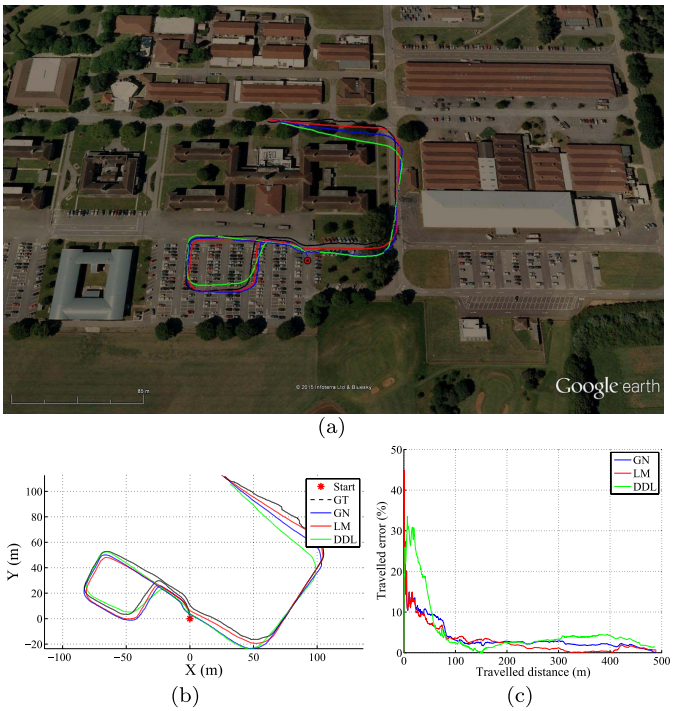



Fig. 11. Computed trajectories and travelled errors for Seq2. (a) Computed trajectories overlaid on Google Earth maps (red star: starting point; black line: GT; blue line: GN; red line: LM; green line: DDL) (b) computed trajectories (same legend applies) (c) corresponding traveled errors (same legend applies).

Fig. 10-18 show the computed trajectories using the different optimisation algorithms for all sequences (Table IV). These figures also include the relative travelled errors obtained using each algorithm. The travelled errors at a given frame

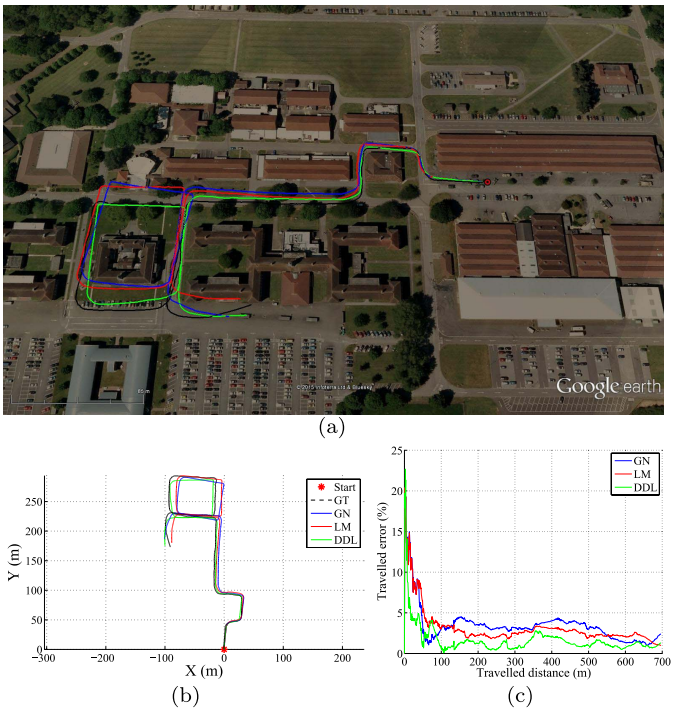



Fig. 12. Computed trajectories and travelled errors for Seq3. (a) Computed trajectories overlaid on Google Earth maps (red star: starting point; black line: GT; blue line: GN; red line: LM; green line: DDL) (b) computed trajectories (same legend applies) (c) corresponding travelled errors (same legend applies).

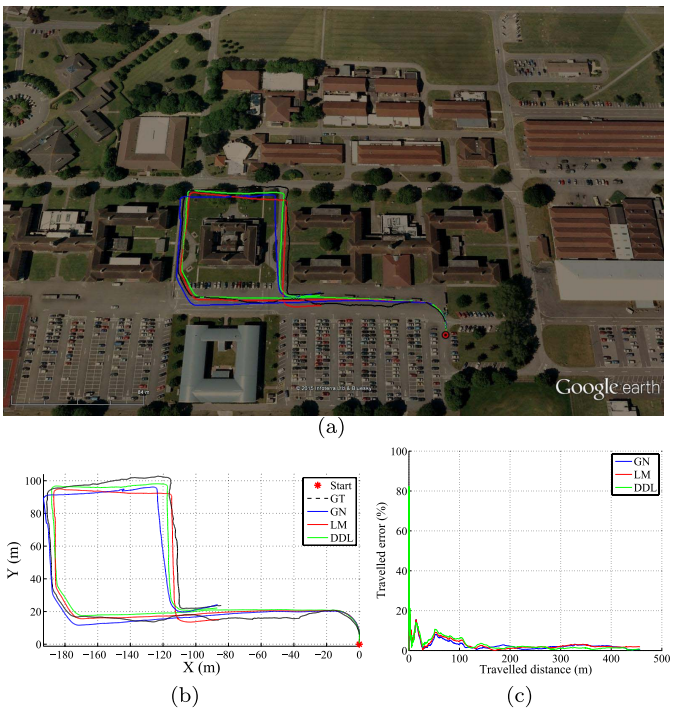



Fig. 13. Computed trajectories and travelled errors for Seq4. (a) Computed trajectories overlaid on Google Earth maps (red star: starting point; black line: GT; blue line: GN; red line: LM; green line: DDL) (b) computed trajectories (same legend applies) (c) corresponding travelled errors (same legend applies).

are calculated by (1), as shown at the bottom of this page, where $P = [X, Y, Z]^{\mathrm{T}}$ are the estimated camera poses and $GT = [GT_{\mathrm{X}}, GT_{\mathrm{Y}}, GT_{\mathrm{Z}}]^{\mathrm{T}}$ correspond to the ground truth. These travelled errors are plotted for the whole trajectory for each sequence. In addition, the average (Avg) and final errors (Fin.) are summarised in Table V for each sequence and each optimisation algorithm (GN, LM and DDL). The average errors were computed at predefined travelled distances in a similar manner to the Kitti Benchmark.[3] In contrast to the final errors, the average errors provide a better indication of the performance of the algorithms. We chose to illustrate both metrics for all algorithms/sequences. For each sequence in Table V, the best performance in terms of the average error is highlighted in **bold**. Similarly, the best final errors are highlighted in red.

Note that the GPS measurements are imprecise. This issue was also reported in [10] and [47]. Indeed, we can clearly notice from Fig. 10a that the ground truth presents some irregularities, especially at the beginning of the sequence. On the other hand, the estimated trajectories appear to be smoother. This indicates that the quality of the estimations might be better than the GPS measurements

[3]http://www.cvlibs.net/datasets/kitti/eval_odometry.php

TABLE V

AVERAGE AND FINAL TRAVELLED ERRORS FOR THE DIFFERENT SEQUENCES AND OPTIMIZATION ALGORITHMS

| Seq# | Travelled distance | GN | | LM | | DDL | |
|---|---|---|---|---|---|---|---|
| | | Avg | Fin. % (m) | Avg | Fin. % (m) | Avg | Fin. % (m) |
| **Seq1** | 687m | 3.79 | 1.78 (12.2m) | 3.53 | 2.24 (15.39m) | **3.07** | 1.19 (8.18m) |
| **Seq2** | 497m | 3.06 | 0.15 (0.75m) | **2.30** | 0.57 (2.83m) | 3.04 | 1.38 (6.86m) |
| **Seq3** | 701m | 3.05 | 2.44 (17.10m) | 2.85 | 1.04 (7.29m) | **1.14** | 1.33 (9.32m) |
| **Seq4** | 460m | 2.11 | 0.50 (2.3m) | 2.39 | 1.88 (8.65m) | **2.07** | 0.61 (2.81m) |
| **Seq5** | 882m | 3.59 | 1.29 (11.38m) | 3.87 | 1.52 (13.41m) | **2.17** | 1.53 (13.49m) |
| **Seq6** | 489m | 6.28 | 1.46 (7.13m) | 5.59 | 4.06 (19.85m) | **4.28** | 1.33 (6.5m) |
| **Seq7** | 399m | 2.08 | 2.26 (9.02m) | 2.00 | 1.99 (7.94m) | **1.36** | 0.67 (2.67m) |
| **Seq8** | 314m | 1.46 | 1.38 (4.33m) | **1.44** | 0.95 (2.98m) | 1.51 | 0.72 (1.32m) |
| **Seq9** | 655m | 1.91 | 1.65 (10.81m) | 1.91 | 2.13 (13.95m) | **1.41** | 0.45 (2.95m) |

(e.g. near buildings). This observation is valid for all the datasets.

The achieved results summarised in Table V are successful. Indeed, the final errors reached below 1% while the average errors were between 1% and 4%. These results are comparable to those obtained using standard visible-band stereo vision systems.

In general, the performance for daytime and night-time sequences is similar. This demonstrates the feasibility of

$$\varsigma(i) = \frac{\sqrt{(X(i) - GT_x(i))^2 + (Y(i) - GT_y(i))^2 + (Z(i) - GT_z(i))^2}}{\sqrt{(GT_x(i) - GT_x(i-1))^2 + \left(GT_y(i) - GT_y(i-1)\right)^2 + (GT_z(i) - GT_z(i-1))^2}} \tag{6}$$

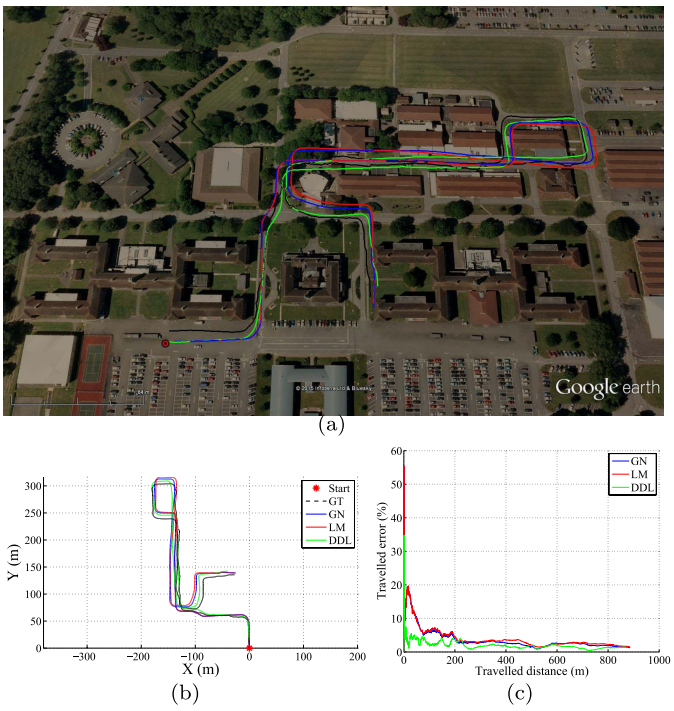



Fig. 14. Computed trajectories and travelled errors for Seq5. (a) Computed trajectories overlaid on Google Earth maps (red star: starting point; black line: GT; blue line: GN; red line: LM; green line: DDL) (b) computed trajectories (same legend applies) (c) corresponding travelled errors (same legend applies).

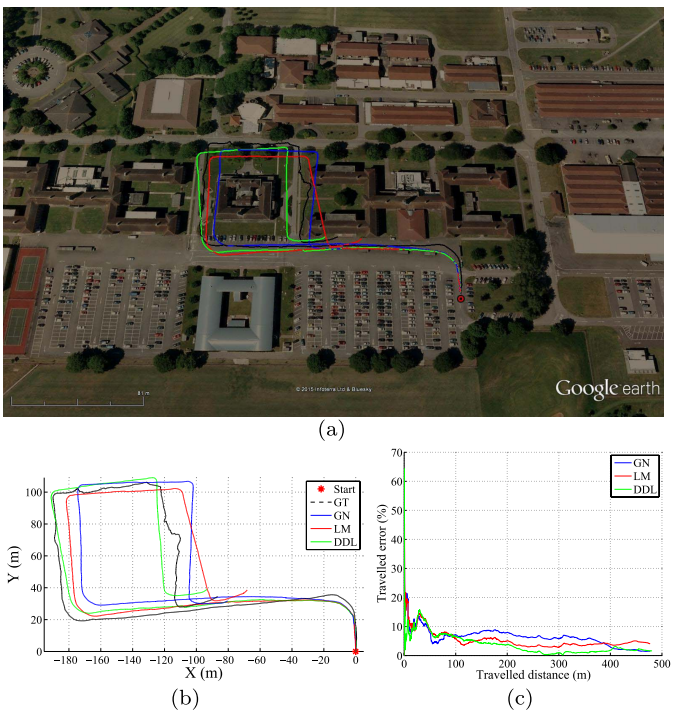



Fig. 15. Computed trajectories and travelled errors for Seq6. (a) Computed trajectories overlaid on Google Earth maps (red star: starting point; black line: GT; blue line: GN; red line: LM; green line: DDL) (b) computed trajectories (same legend applies) (c) corresponding travelled errors (same legend applies).

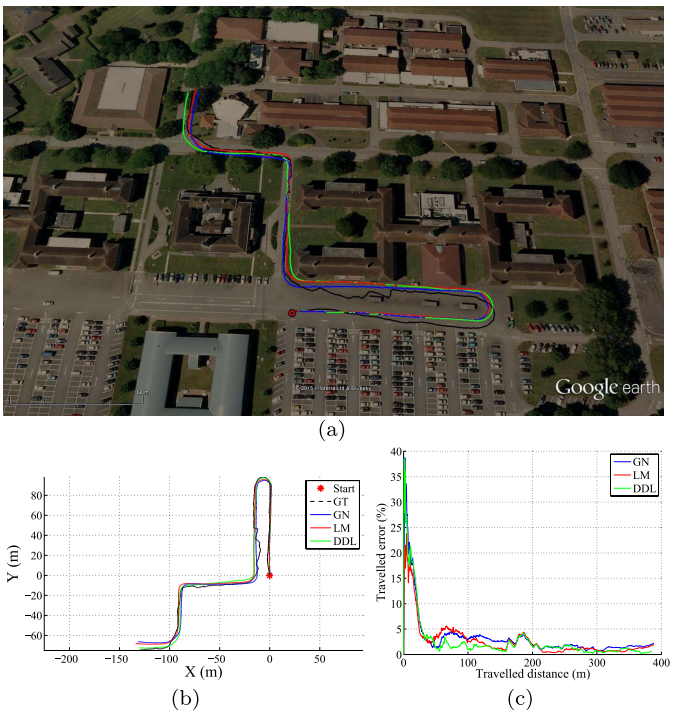



Fig. 16. Computed trajectories and travelled errors for Seq7. (a) Computed trajectories overlaid on Google Earth maps (red star: starting point; black line: GT; blue line: GN; red line: LM; green line: DDL) (b) computed trajectories (same legend applies) (c) corresponding travelled errors (same legend applies).

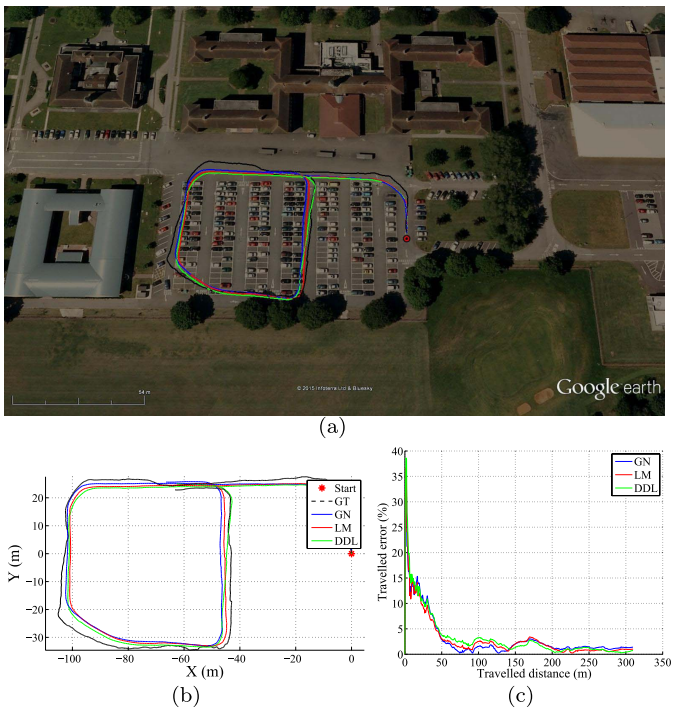



Fig. 17. Computed trajectories and travelled errors for Seq8. (a) Computed trajectories overlaid on Google Earth maps (red star: starting point; black line: GT; blue line: GN; red line: LM; green line: DDL) (b) computed trajectories (same legend applies) (c) corresponding travelled errors (same legend applies).

night-time navigation to a relatively high degree of accuracy. Consequently, it also shows that visual odometry concepts can be extended to night-time. This can be exploited in a variety of applications e.g. intelligent transportation systems, search and rescue, military operations to name a few.

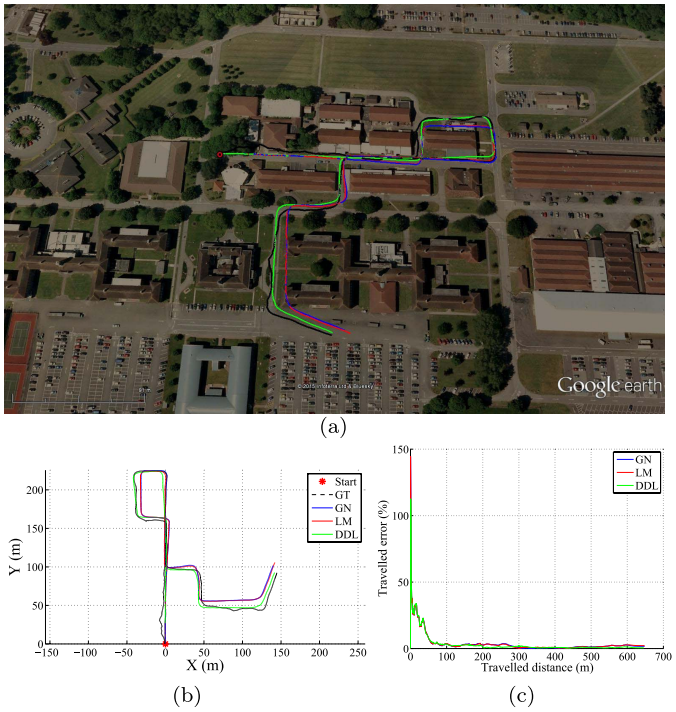



Fig. 18. Computed trajectories and travelled errors for Seq9. (a) Computed trajectories overlaid on Google Earth maps (red star: starting point; black line: GT; blue line: GN; red line: LM; green line: DDL) (b) computed trajectories (same legend applies) (c) corresponding travelled errors (same legend applies).

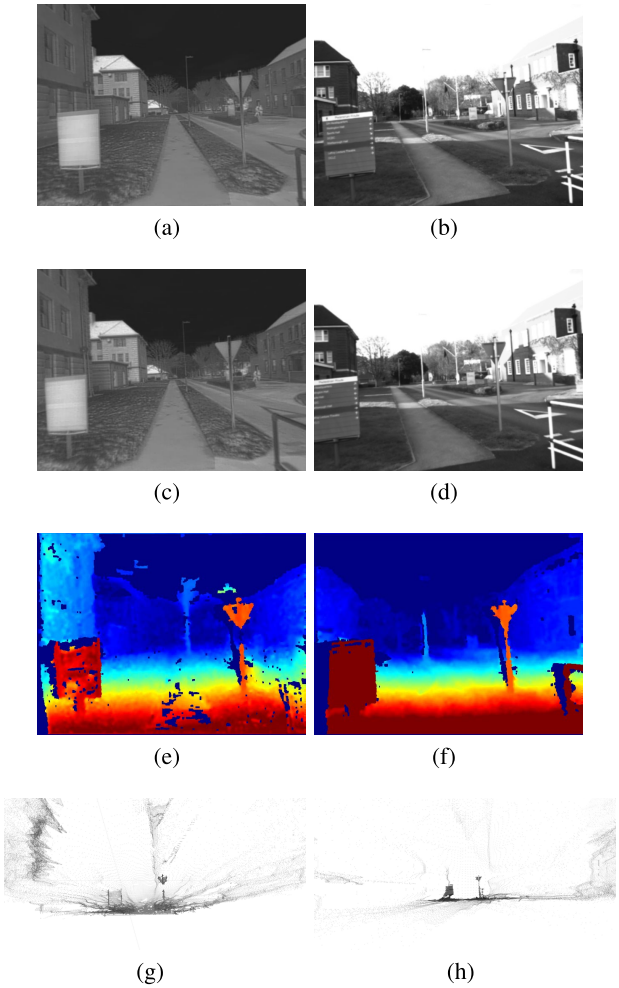



Fig. 19. Example showing thermal and visible disparity maps and 3D reconstruction of a similar scene (a) (c) thermal left and right images (b) (d) visible left and right images (e) (f) disparity map from thermal and visible images (g) (h) 3D point clouds from thermal and visible images. Note that we can identify the two signs (triangle and square) in the point clouds.

We can observe from Table V that the highest average error corresponds to **Seq6** which was captured during a very cold night with clear sky. The main effect of cold weather in thermal imagery is a loss in terms of contrast and texture as the difference in temperature between scene elements is reduced. This is mainly caused by the absence of a heat source (e.g. the sun) and clouds (i.e. clear sky). This has an effect on the matching sub-task and therefore the visual odometry process. A similar trajectory to **Seq6** was captured during daytime and used in the evaluation process (**Seq4**). We can note from Fig. 13 that the estimated trajectories during daytime (**Seq4**) are better than for the very cold night-time (**Seq6**). This observation is also valid for their corresponding travelled errors (Table V) which are higher for **Seq6**. This said, in warmer weather conditions, the performance in night-time is comparable to daytime.

With respect to the optimisation algorithms, we can clearly note from Table V and Fig. 10-18 that the general trend is that the Double Dogleg algorithm provides better trajectory estimates than Gauss-Newton and Levenberg-Marquadt. This correlates with the findings of [48] where the Dogleg approach was shown to perform better than LM for visible-band VO. Here, we illustrated that DDL presents an interesting alternative to LM, which is extensively used in the optimisation sub-task of visual odometry.

*2) Thermal 3D Reconstruction:* in this section, we show that 3D reconstruction can be achieved using a thermal stereo vision system in a similar fashion to *standard* stereo setups. Fig. 19 shows the disparity map along with the reconstructed 3D point cloud (PCL) computed from a pair of stereo thermal images. These were captured during daytime in cloudy and cold weather. We also plot the disparity map and 3D PCl from a pair of visible-band stereo images (Fig. 19b-19d) capturing the same scene using the vision system proposed in [1]. We used a specifically tuned version of the standard semi-global block matching algorithm [49] to generate the disparity map (Fig. 19e-19f). The 3D point cloud was then reconstructed using the stereo calibration parameters (Fig. 19g-19h). We can note from Fig. 19 that both disparity maps and 3D point clouds are comparable. This indicates that, similarly to the visible-band, the infrared modality can be used for mapping applications with an additional crucial advantage. Indeed, while visible-band cameras can only be used in daytime, thermal sensors extend the operability to night-time.

The computed 3D model can be augmented with thermal information. Indeed, the Sensoray frame grabbers introduced in Section V-A, allow the capture of up to two video feeds from the TAU2 cameras. The first feed can be used for VO whereas the second can be used to augment the 3D model of the scene. This enables thermal 3D modelling i.e. building

3D models with temperature information overlaid. This can be useful in many areas such as building inspection or search and rescue operations.

## VI. Conclusion

A thermal stereo odometry solution was proposed for the estimation of travelled trajectories using solely captured infrared images. The main objective of this work is to demonstrate the usability of thermal cameras in applications they were not specifically designed for. Notably, we were able to extend the concepts of visual odometry beyond the visible spectrum enabling a new range of crucial applications e.g. night-time navigation. In contrast to visible band cameras, the calibration of thermal sensors proved very challenging due to numerous reasons (Section III). However, these difficulties were overcome and the stereo vision system was successfully calibrated. Fast-Hessian interest points were combined with FREAK descriptors to enhance feature matching in thermal modality. The validity of the proposed approach has been extensively demonstrated using our own datasets where different weather conditions and time-of-day were considered. Daytime as well as night-time navigation capabilities were established. More specifically, we showed that using an alternative optimisation algorithm i.e. Double Dogleg allowed us to improve the quality of the estimated trajectories. Additionally, thermal 3D reconstruction was illustrated. This shows that despite the inherent problems of thermal imagery, many computer vision algorithms can be adopted to produce outcomes comparable to *standard* vision systems. For future work, we are looking at ways to further improve the proposed approach. One way would be the integration/fusion of other proprioceptive and/or exteroceptive sensors. For instance, filtering algorithms can be considered in a SLAM-like framework where IMU/image information is fused. Alternatively, 3D cameras or laser scanners could be used to further enhance the trajectory estimation accuracy.

## References


[1] T. Mouats, N. Aouf, and M. A. Richardson, "A novel image representation via local frequency analysis for illumination invariant stereo matching," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2685–2700, Sep. 2015.

[2] D. Scaramuzza and F. Fraundorfer, "Visual odometry part I: The first 30 years and fundamentals," *IEEE Trans. Robot. Autom.*, vol. 18, no. 4, pp. 80–92, Dec. 2011.

[3] 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 *Proc. Int. Conf. Digit. Imag. Comput. Techn. Appl.*, Dec. 2011, pp. 217–224.

[4] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[5] K. Mikolajczyk *et al.*, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, nos. 1–2, pp. 43–72, Oct. 2005.

[6] N. Süenderhauf, K. Konolidge, T. Lemaire, and S. Lacroix, "Comparison of stereovision odometry approaches," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA05)*, Apr. 2005.

[7] F. Fraundorfer and D. Scaramuzza, "Visual odometry: Part II: Matching, robustness, optimization, and applications," *IEEE Robot. Autom. Mag.*, vol. 19, no. 2, pp. 78–90, Jun. 2012.

[8] S.-H. Jung, J. Eledath, S. Johansson, and V. Mathevon, "Egomotion estimation in monocular infra-red image sequence for night vision applications," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Feb. 2007, p. 8.

[9] M. Magnabosco and T. P. Breckon, "Cross-spectral visual simultaneous localization and mapping (SLAM) with sensor handover," *Robot. Auto. Syst.*, vol. 61, no. 2, pp. 195–208, Feb. 2013.

[10] T. Mouats, N. Aouf, A. D. Sappa, C. Aguilera, and R. Toledo, "Multispectral stereo odometry," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1210–1224, Jun. 2015.

[11] K. Owens and L. Matthies, "Passive night vision sensor comparison for unmanned ground vehicle stereo vision navigation," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 1, 2000, pp. 122–131.

[12] T. B. Schon and J. Roll, "Ego-motion and indirect road geometry estimation using night vision," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2009, pp. 30–35.

[13] A. Rankin *et al.*, "Unmanned ground vehicle perception using thermal infrared cameras," *Proc. SPIE*, vol. 8045, pp. 804503-1–804503-26, May 2011.

[14] K. Hajebi and J. S. Zelek, "Structure from infrared stereo images," in *Proc. Can. Conf. Comput. Robot. Vis.*, May 2008, pp. 105–112.

[15] S. J. Krotosky and M. M. Trivedi, "On color-, infrared-, and multimodal-stereo approaches to pedestrian detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 4, pp. 619–629, Dec. 2007.

[16] T. Luhmann, J. Piechel, and T. Roelfs, "Geometric calibration of thermographic cameras," *Thermal Infr. Remote Sens.*, vol. 17, no. 1, pp. 27–42, May 2013.

[17] P. Engström, H. Larsson, and J. Rydell, "Geometric calibration of thermal cameras," *Proc. SPIE*, vol. 8897, p. 88970C, Oct. 2013.

[18] 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 Trans. Instrum. Meas.*, vol. 61, no. 6, pp. 1625–1635, Jun. 2012.

[19] J. S. Zelek, M. Holbein, K. Hajebi, D. C. Asmar, and D. Cheng, "IR depth from stereo for autonomous navigation," *Proc. SPIE*, vol. 5784, pp. 316–330, May 2005.

[20] J. Harguess and S. Strange, "Infrared stereo calibration for unmanned ground vehicle navigation," *Proc. SPIE*, vol. 9084, p. 90840S, Jun. 2014.

[21] M. Weinmann, J. Leitloff, L. Hoegner, B. Jutzi, U. Stilla, and S. Hinz, "Thermal 3D mapping for object detection in dynamic scenes," in *Proc. ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 1. Nov. 2014, pp. 53–60.

[22] A. Ellmauthaler, E. A. B. da Silva, C. L. Pagliari, J. N. Gois, and S. R. Neves, "A novel iterative calibration approach for thermal infrared cameras," in *Proc. IEEE 20th Int. Conf. Imag. Process. (ICIP)*, Sep. 2013, pp. 2182–2186.

[23] F. Barrera Campo, F. Lumbreras Ruiz, and A. D. Sappa, "Multimodal stereo vision system: 3D data extraction and algorithm evaluation," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 5, pp. 437–446, Sep. 2012.

[24] S. Prakash, P. Y. Lee, T. Caelli, and T. Raupach, "Robust thermal camera calibration and 3D mapping of object surface temperatures," *Proc. SPIE*, vol. 6205, pp. 62050J-1–62050J-8, Apr. 2006.

[25] M. Mohd Norzali, M. Kashima, K. Sato, and M. Watanabe, "Effective geometric calibration and facial feature extraction using multi sensors," *Int. J. Eng. Sci. Innov. Technol.*, vol. 1, no. 2, pp. 170–178, 2012.

[26] J.-Y. Bouguet, "Camera calibration toolbox for Matlab," 2008.

[27] J. Bartl and M. Baranek, "Emissivity of aluminium and its importance for radiometric measurement," *Meas. Phys. Quantities*, vol. 4, no. 3, pp. 31–36, 2004.

[28] M. Warren, D. McKinnon, and B. Upcroft, "Online calibration of stereo rigs for long-term autonomy," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2013, pp. 3692–3698.

[29] T. Emanuele and V. Alessandro, *Introductory for 3-D Computer Vision*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.

[30] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[31] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Imag. Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[32] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. 9th Eur. Conf. Comput. Vis.*, vol. 3951. May 2006, pp. 430–443.

[33] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vis. Conf.*, 1988, pp. 147–152.

[34] J. S. J. Shi and C. Tomasi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1994, pp. 593–600.

[35] M. Agrawal, K. Konolige, and M. R. Blas, "CenSurE: Center surround extremas for realtime feature detection and matching," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2008, pp. 102–115.

[36] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 603–610.
[37] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2564–2571.
[38] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2548–2555.
[39] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 510–517.
[40] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Proc. 11th Eur. Conf. Comput. Vis.*, vol. 6314, Sep. 2010, pp. 778–792.
[41] K. Levenberg, "A method for the solution of certain problems in least squares," *Quart. Appl. Math.*, vol. 2, pp. 164–168, 1944.
[42] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. Ind. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.
[43] J. E. Dennis, Jr., and H. H. W. Mei, "Two new unconstrained optimization algorithms which use function and gradient values," *J. Optim. Theory Appl.*, vol. 28, no. 4, pp. 453–482, Aug. 1979.
[44] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
[45] M. I. A. Lourakis and A. A. Argyros, "Is Levenberg–Marquardt the most efficient optimization algorithm for implementing bundle adjustment?" in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, Oct. 2005, pp. 1526–1531.
[46] B. Triggs, P. F. Mclauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—A modern synthesis," *Vision Algorithms: Theory and Practice*, vol. 1883. Berlin, Germany: Springer-Verlag, 2000, pp. 298–372.
[47] A. Geiger, J. Ziegler, and C. Stiller, "StereoScan: Dense 3D reconstruction in real-time," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2011, pp. 963–968.
[48] L. Chermak, "Standalone and embedded stereo visual odometry based navigation sensor," Ph.D. dissertation, Cranfield Univ., Bedford, U.K., 2014.
[49] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

**Tarek Mouats** received the Computer Science Engineering degree from the National Polytechnic School, Algiers, Algeria, in 2005, and the M.Sc. degree in defense sensors and data fusion from Cranfield University, Shrivenham, U.K., in 2008, where he is currently pursuing the Ph.D. degree with the Centre for Electronic Warfare.

His research focuses on image processing, multimodal image processing, intelligent transportation systems, localization techniques, and more specifically visual odometry.

**Nabil Aouf** is currently a Reader with the Centre of Electronic Warfare, Cranfield University, U.K. He has authored over 100 publications in high calibre in his domains of interest. His research interests are aerospace and defense systems, information fusion and vision systems, guidance and navigation, tracking, and control and autonomy of systems. He is an Associate Editor of the *International Journal of Computational Intelligence in Control.*

**Lounis Chermak** received the M.Sc. degree in computer vision and the Ph.D. degree in visual navigation from Cranfield University, U.K., in 2011 and 2014, respectively. He is currently a Research Fellow with the Centre of Electronic Warfare, Cranfield University. His research interests include 3-D visualization, computer vision, robotics, navigation systems, and embedded sensor.

**Mark A. Richardson** has over 30 years of experience in electro-optics and infrared systems and countermeasures in the defense industry and U.K. academia, and has written well over 200 classified and unclassified papers on these subjects. He is currently the Head of the Centre for Electronic Warfare and the Director of Research with the Defence Academy, U.K.