PUGTIFs: Passively User-Generated Thermal Invariant Features

EDWARD JACKSON, (Student Member, IEEE), AND LOUNIS CHERMAK, (Member, IEEE)
Centre of Electronic Warfare Information and Cyber (CEWIC), Cranfield University, Cranfield SN6 8LA, U.K.

Corresponding author: Edward Jackson (e.jackson@cranfield.ac.uk)

ABSTRACT Feature detection is a vital aspect of computer vision applications, but adverse environments, distance and illumination can affect the quality and repeatability of features or even prevent their identification. Invariance to these constraints would make an ideal feature attribute. Here we propose the first exploitation of consistently occurring thermal signatures generated by a moving platform, a paradigm we define as passively user-generated thermal invariant features (PUGTIFs). In this particular instance, the PUGTIF concept is applied through the use of thermal footprints that are passively and continuously user generated by heat differences, so that features are no longer dependent on the changing scene structure (as in classical approaches) but now maintain a spatial coherency and remain invariant to changes in illumination. A framework suitable for any PUGTIF has been designed consisting of three methods: first, the known footprint size is used to solve for monocular localisation and thus scale ambiguity; second, the consistent spatial pattern allows us to determine heading orientation; and third, these principles are combined in our automated thermal footprint detector (ATFD) method to achieve segmentation/feature detection. We evaluated the detection of PUGTIFs in four laboratory environments (sand, grass, grass with foliage, and carpet) and compared ATFD to typical image segmentation methods. We found that ATFD is superior to other methods while also solving for scaled monocular camera localisation and providing user heading in multiple environments.

INDEX TERMS feature detector, image segmentation, monocular scaled localisation, thermal footprint.

I. INTRODUCTION

Feature detection is a fundamental and widely-studied aspect of computer vision systems. Feature detectors aim to describe image content by computing abstractions such as points, edges/lines or image patches representing structure within scenes [1]. They are used in numerous applications, including the localisation and tracking of people or ground, naval and space vehicles, obstacle avoidance, and facial recognition. The repeatability and robustness of features as scenes change (or become obscured by obstacles or challenging illumination conditions such as glare and darkness) are two of the most desirable qualities of a feature detector.

We have therefore developed a new approach based on the detection of passively user-generated thermal invariant features (PUGTIFs), in this case the PUGTIF concept (patent pending, BAE Systems) is applied through thermal footprints

generated by a walking person. Thermal footprints are consistently available and independent of the changing scenes that typify classical feature detection methods. State-of-the-art feature detectors use classical features that are dependent on the changing scene structure and are based on methods that detect local points (e.g., SIFT [2], SURF [3], ORB [4], BRISK [5] and FAST [6]), edges (e.g., Canny [7] and Sobel), or regions (e.g., MSER [8] and Salient Regions [9]).

The performance and repeatability of feature detection methods have been discussed extensively in the literature, but published comparisons focus mainly on evaluations based on the performance of image features when the image is transformed, scaled or rotated [10], [11]. In contrast, the effect of the scene content has been investigated only recently. A comparison of image feature detectors using a large database of 20,482 images with 539 scenes revealed that feature detectors are influenced by the scene content [12], and further analysis indicated that the scene content influences feature detectors using the same database [13]. The authors

The associate editor coordinating the review of this article and approving it for publication was Christopher H. T. Lee.
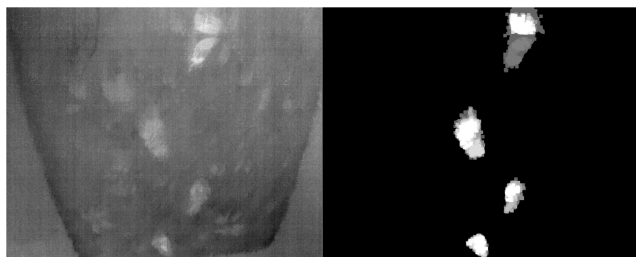
FIGURE 1. Footprints segmented using our new automated thermal footprint detector (ATFD) method. The original image is shown on the left and the image segmented using the ATFD is shown on the right.

FIGURE 2. The three major challenges in camera-based localisation. Features are matched with either temporally variant images or known world objects to estimate for scale/camera position. The quality of the estimation is strongly influenced by the selected feature. The ability to match this feature consistently determines the amount of drift the system will develop over time.
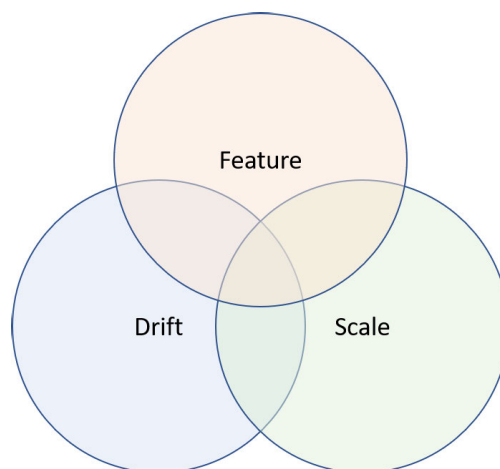
concluded that feature detectors have preferences for types of scene content that determine their repeatability scores, and the proposed solution was to use detectors that perform well in particular scenes [13].

In contrast to the classical approach, we set out to remove the dependency on scene content by detecting PUGTIFs (here, thermal footprints). The repeatability of such features is inherent because they are generated by the heat difference between the platform and the environment, and they occur whenever the platform is moving. Therefore, the feature is constantly available as long as the heat difference can be measured. The challenge in the exploitation of PUGTIFs is not only the segmentation of the feature from the image but also the estimation of the scaled monocular position and the user's heading orientation. This is possible due to the unique attributes of PUGTIFs. To our knowledge, the use of thermal footprints as features is completely novel and methods that achieve segmentation while estimating the monocular scaled position and user's heading have yet been published.

Accordingly, we developed a framework that first incorporates our separate methods to estimate monocular scaled camera position and a user heading orientation and then combines them into our automated thermal footprint detector (ATFD) to achieve segmentation (Fig. 1). We evaluated the performance of the new methods in four laboratory environments (grass, grass with foliage, sand, and carpet) in order to assess the repeatability and robustness of the method framework and PUGTIFs under diverse conditions. We compared ATFD to three of the most common thresholding segmentation techniques: Otsu [14], *k*-means [15] and adaptive thresholding [16].

ATFD benefits from the characteristics that are unique to thermal footprints. Specifically, when two initial footprints have been segmented, size can be inferred from the known dimensions of the user's foot, after which the monocular scale ambiguity problem can be solved. Furthermore, the segmentation of the subsequent pair of thermal footprints reveals the user's heading, based on the natural stance with the centre of mass between the two legs. This provides a valuable opportunity to exploit applications such as people tracking, dead reckoning, self-localisation, and GPS-denied navigation.

As an example, navigation using a monocular camera raises three major challenges: feature availability, scale

ambiguity, and relative drift (Fig. 2). Given that classical feature detectors depend on illumination and structured content in the environment, the lack of (or inability to detect) such features would severely degrade the performance of subsequent processes, resulting in drift. Using a monocular camera in the ideal case, where the scene presents advantageous characteristics for a detector and features are detected consistently, transformation from two-dimensional (2D) image pixels into three-dimensional (3D) point world coordinates would involve an unknown scale. For classical approaches, detected features belong to objects of unknown size, so there is no direct way to solve the scale ambiguity problem. In contrast, a passively user-generated, periodic feature with known real-world dimensions independent of the scene content and related drift issues (i.e., PUGTIFs) can overcome these challenges.

The work discussed in this paper makes the following contributions to the field:

- Utilisation of PUGTIFs, specifically the instance of a user's footprint observed in the thermal spectrum, as a passively self-generated feature.
- A solution to monocular scale estimation, available for every frame based on the utilisation of *a priori* knowledge about the user's real-world footprint size.
- Development of a method to estimate user heading based on knowledge of the natural human stance and gait.
- Development of a segmentation and detection method (ATFD) that uses both the user's heading and monocular scale estimation to segment all thermal footprints.

The remainder of this paper is structured as follows. Section II introduces the three methods of the PUGTIF framework and the calculations required to determine scale and the user's heading. Section III analyses and evaluates the performance of the ATFD segmentation method in different
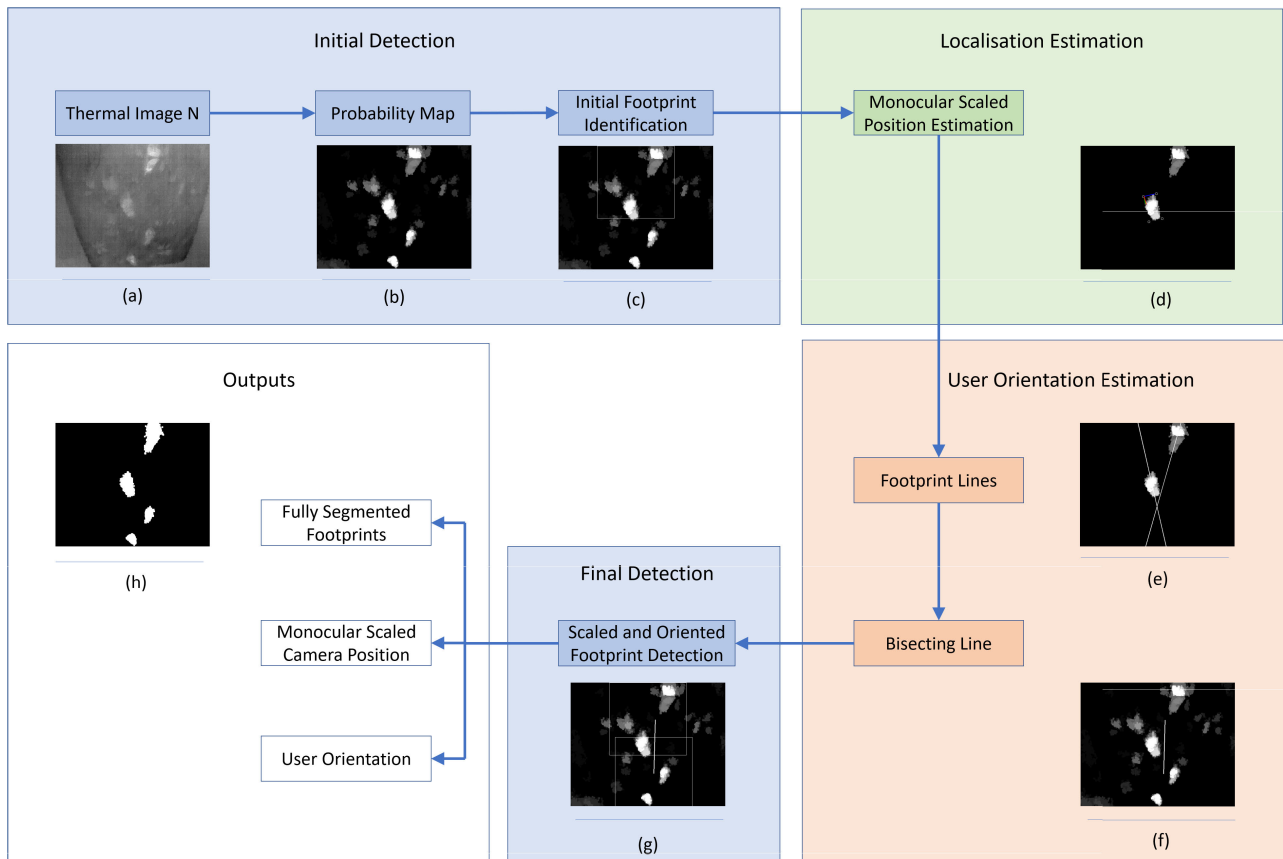
FIGURE 3. Overall template framework for PUGTIF showing the monocular scaled position (green), orientation (orange), and the initial and final detection (blue) to achieve ATFD. The outputs of the system are the segmented footprints, a monocular scaled camera position and a user orientation. The specific instance of PUGTIF using thermal footprints is shown in a flow diagram within each boxed section: (a) original thermal image; (b) probability map of footprints based on the matrix of contours that satisfied the contour constraints filter; (c) rectangle search area of the probability map to identify the two most likely footprints immediately in front of the camera; (d) footprint monocular scale estimation showing the axes (red, green and blue lines) projected onto the footprint using the estimated transform; the white circles are the 2D points corresponding to the 3D model we used to determine the scale; (e) user orientation, with two lines used to calculate the bisecting line drawn from the centre and angle of each footprint; (f) user orientation, with the probability map showing a scaled and oriented search line (white) identifying the target area for the next set of footprints; (g) scaled and oriented box search showing the final search box for the next set of footprints; (h) segmented footprints, showing all of the footprints that were identified.

laboratory environments. Section IV discusses our findings and how they relate to the methods of monocular scale estimation, determination of user heading, and segmentation. Finally, Section V presents our conclusions and recommends some areas of further work.

II. METHODOLOGY

We present a template PUGTIF framework that can identify a known thermal signature within an image while predicting subsequent instances of the same feature based on the inherent nature of the moving platform. The template comprises three methods: estimation of scaled monocular camera position, estimation of orientation, and the combination of these methods (ATFD) to achieve full segmentation (Fig. 3).

The first stage of the methodology involves the creation of a probability map using thresholding along with parameters to identify multiple potential footprints. The two most recent footprints are then identified by drawing a bounding box positioned close to the camera (at the top of the image) and

selecting the two most likely candidates from within. The search for the next pair of footprints is accomplished by estimating the scale of the search area and also the user's heading direction. The monocular scale is estimated using the larger of the two initial footprints identified within the bounding box. The user orientation is then calculated using the first two footprints, as shown in Fig. 4(a). The scale is then used to set the search direction to find the third footprint, as shown in Fig. 4(b). After searching in the direction calculated from the first pair of footprints at the appropriate scale, all the footprints in the scene can be identified.

A. PROBABILITY MAP CONSTRUCTION

The statistical distribution of pixels in a thermal footprint by intensity is shown in Fig. 5. We started thresholding at the highest point in the histogram (threshold iteration $it1$) and continued until itN. Following each iteration, the image was dilated to reduce the speckled noise. After dilation, we searched for contours, applying a filter to constrain the
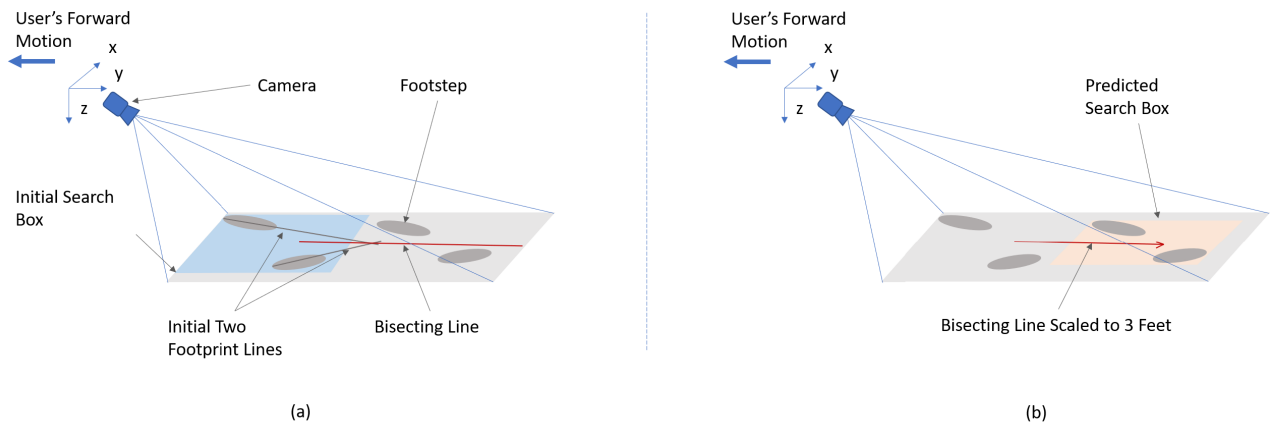
FIGURE 4. Simplified representation of the footprint detection principles. (a) The bisecting line between two footprints can be used to calculate the search direction for the next two. (b) Calculating the scale allows us to set the search bounding box to 75 cm distance assuming 25 cm is equal to one user footprint in this application.
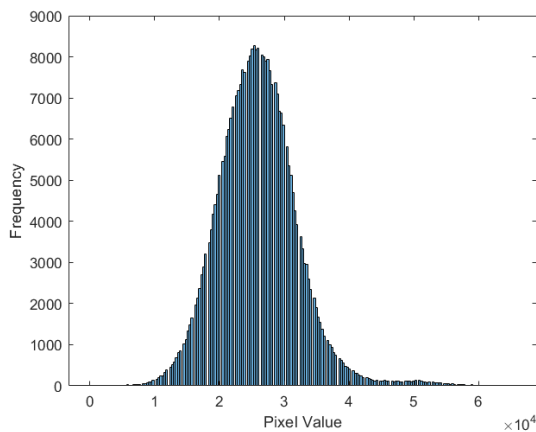
FIGURE 5. Histogram of a 16-bit thermal footprint image based on pixel intensity.

footprints using an empirical area of $[50 \times 10 - 100 \times 100]$ pixels, with anything outside this range discarded as noise. Due to the shape of the footprints and their similarity to ellipse shapes, we fitted ellipses to the contours as shown in Table 1.

The ellipses represented all possible footprints. We filtered the ellipses using three additional parameters: the proportion of the area overlap between the contour and the ellipse, the size of indents into the ellipse from the contour, and a similarity metric between the contour and the ellipse. The red outlines represent outliers that were rejected using these constraints and the green outlines represent footprints that passed the filter. This process was repeated for *itN* iterations (as shown for four iterations in Table 1), and during each iteration the contours that met the criteria were added to a matrix storing the locations of all the pixels that were most likely to belong to footprints. We therefore named this matrix a probability map, as shown in Fig. 3(b).

The contours in the probability map were then located using a border-following algorithm [17]. The application had

some fixed parameters, such as the camera location. The camera was held behind the user, so we knew that two footprints would always be located below the top centre of the image. Therefore, a rectangle area was created in which two of the most likely footprints could be found, as shown in Fig. 3(c). The size of this rectangle was determined empirically but it can be adjusted if necessary.
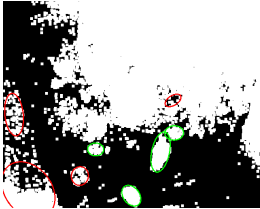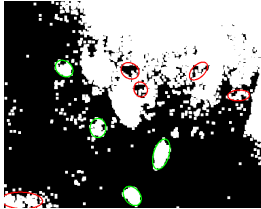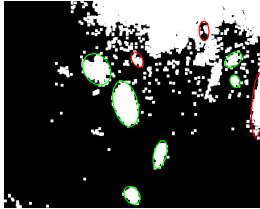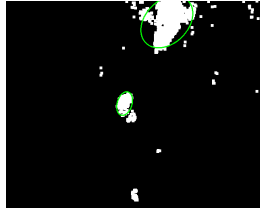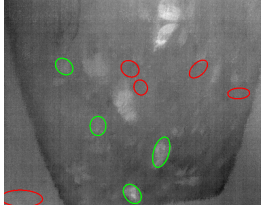
B. ESTIMATION OF MONOCULAR SCALE
Having found the two initial footprints, it was necessary to estimate the scale based on *a priori* parameters. The first of these were the measured length and width of the user's foot. The length and width were used to construct a 2D rectangle, once constructed the corners were stored. The four corners representing the 2D rectangle were then used to build a 3D planar cuboid using the camera calibration parameters and 3D perspective transform. The second known parameter was the camera matrix and distortion coefficients, which were acquired by calibrating the camera using a metal chessboard. Given these known parameters, the larger of the two footprints inside the rectangle area was selected because it was more likely to be completely in view of the camera. A rectangle was fitted to the footprint and the four corners of the rectangle were identified. This generated a 2D model of the entire segmented footprint with four corners, shown as circles in Fig. 3(d). Using the correspondences between the 2D model and the 3D planar model, we could solve for the object pose using Perspective-n-Point with Levenberg-Marquardt optimisation. The object points, in our case were the 3D footprint model points and the corresponding in the 2D image plane. To visually represent the quality of the scale estimation, we then projected the axis to the footprint as shown in Fig. 3(d).

C. ESTIMATION OF USER ORIENTATION
The next step was to find the remaining footprints in the image. To achieve this, we found the bisecting line between

TABLE 1. An example of four threshold iterations as applied to thermal footprints. The red outlines represent outliers that were rejected using ellipse contour constraints and the green outlines represent inlier footprints that satisfied this filter.

Type	Threshold 1	Threshold 2	Threshold 3	Threshold 4
Threshold				
Original Image				

the first two footprints and used it to determine the walking direction and thus the likely position of the next two footprints. As shown in Fig. 3(e), the two footprints closest to the camera were both fitted with ellipse shapes, and the centres and angles θ_0 and θ_1 were then used to calculate the bisecting line. The calculation shown in (1) gave the direction vector $\mathbf{v}$ which we used to search for the next pair of footprints. Although this determined the direction of the search, the scale (to limit the search along this vector) was unknown. The missing scale value was determined using the scale we had previously calculated by applying Perspective-n-Point.

$$a_0 = -\tan\frac{\theta_0\pi}{180} \quad a_1 = -\tan\frac{\theta_1\pi}{180} \quad b_0 = 1\ b_1 = 1$$
$$ab = \frac{a_0}{\sqrt{a_0^2+b_0^2}} - \frac{a_1}{\sqrt{a_1^2+b_1^2}}$$
$$bb = \frac{b_0}{\sqrt{a_0^2+b_0^2}} - \frac{b_1}{\sqrt{a_1^2+b_1^2}} \quad \mathbf{v} = \begin{bmatrix} ab \\ bb \end{bmatrix} \quad (1)$$

The line was drawn at an arbitrary length in the 2D image plane from the midpoint of the two footprints. To switch the line between the 2D image plane and the 3D coordinates so that it could be scaled, it was necessary to construct a homography matrix. The extrinsic matrix $[R|\mathbf{t}]$ was created by concatenating the rotation and translation matrices. The extrinsic matrix was then multiplied by the camera matrix K to give the projection matrix P. The projection matrix was used to construct the homography matrix as shown in (2).

$$P = K \times [R|\mathbf{t}]$$
$$\mathbf{H} = \begin{pmatrix} p_{01} & p_{02} & p_{04} \\ p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \end{pmatrix} \quad (2)$$

The inverse homography matrix was then used to transform the start and end points of the line to equivalent 3D points.

The new 3D points of the start and end of each line were used to calculate the equation of the line in the 3D coordinate system. This 3D line vector could then be scaled using the known size of footprints. We restricted the length of the line by using the size of three footprints, or 75 cm in real units (25 cm = one footprint). The start and end points of the new line in the 3D world frame were then converted back into the 2D image plane using the homography matrix. A line was drawn to visualise the direction and scale of the new line, as shown in Fig. 3(f).

The target area for the next footprints was therefore fixed to the directional vector of the two initial footprints and was limited to a distance of up to three footprints away. The end of the line was consequently used to centre another rectangle box, within which the most likely footprints were selected from the probability map shown in Fig. 3(g). This led to the identification of the next pair of footprints as shown in Fig. 3(h).

III. EVALUATION AND ANALYIS
A. EXPERIMENTAL SETUP
The new method was tested in four different laboratory environments: grass, grass with foliage, sand, and carpet (Fig. 6). We used a FLIR Tau 2 thermal camera with a resolution 640 × 512, a spectral range of 7.5-13.5μm and a noise-equivalent temperature differential (NETD) of less than 50mK. The NETD is important because it determines how easily the footprints can be identified. Using the Tau 2 it was possible to detect footprints given a temperature difference of 8.6 °C. This value was calculated by taking the average differences between material temperature and foot temperatures in Table 2.

An Oregon Scientific WMR300 Weather Station was used to measure the ambient temperature and humidity, and a digital thermometer and infrared (IR) temperature probe and gun were used to measure the temperature of the target and

TABLE 2. Initial temperature and humidity parameters for the target and environment before the detection of thermal footprints.

Environment	Material temperature °C		Foot temperature °C		Room temperature °C	Humidity %
	Probe	IR gun	Probe	IR gun	Weather station	Weather station
Grass	18	17.5	27.3	26.5	19.2	61
Grass Foliage	18.3	17.5	24.3	26.5	19.2	61
Sand	17.9	18.3	27.3	27	19.1	61
Carpet	18.2	18.3	27.8	28	19	62

FIGURE 6. Setup of the four different test environments in the laboratory: (a) sand, (b) grass, (c) grass with foliage, (d) carpet.

the surface. The temperature and humidity values are shown in Table 2.

After measuring the temperatures before each test, the user walked forwards, holding the thermal camera behind at waist level to record the footprints across each environment. The process was carried out four times for each environment. Ten images from each environment were selected and manually labelled with the 3-4 ground truth footprints present in each case. The 10 images were segmented using Otsu [14] which automatically computes a global threshold, k-means [15] using 2 clusters to represent the footprint and background, adaptive thresholding [16] with a sensitivity of 0.4 and ATFD, the results were compared using the evaluation metrics described below. The first evaluation metric we used

was the intersection over union (IoU), which is commonly used to evaluate segmentation. The MeanIoU score gives a statistical measurement that penalizes false positives. This is an important attribute in our application because we are attempting to identify only the footprint feature in the image, and any pixels beyond this structure should be classed as noise or an outlier. The score was determined using the ratio of correctly classified pixels to the total number of ground truth and predicted pixels, as shown in (3), where TP = true positive, FP = false positive and FN = false negative.
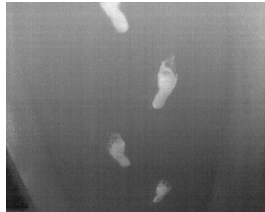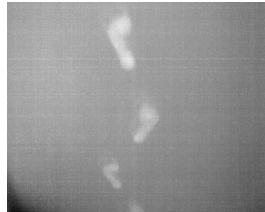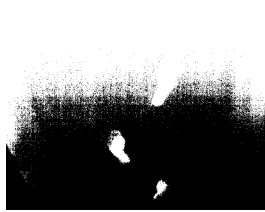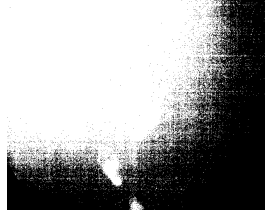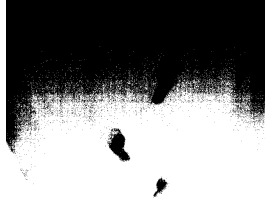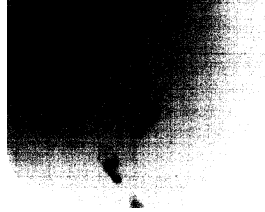
$$IoU = \frac{TP}{(TP + FP + FN)} \quad (3)$$

The second evaluation metric we used was the boundary F1 (BF) contour-matching score shown in (4). This indicates how well the predicted boundary of the footprint aligns with the boundary of the ground truth. The method is described as the harmonic mean (F1-measure) of the precision and recall alongside a distance error to determine whether a point on the predicted boundary has a match on the ground truth boundary. The boundary score was used because it is thought to correlate better with human qualitative assessments [18] than the IoU, which is advantageous in some computer vision applications.

$$BF(BoundaryF1)Score = \frac{2 \times Precision \times Recall}{(Recall + Precision)} \quad (4)$$

The performance of the thresholding segmentation techniques in the four different environments are compared in Table 3. Otsu and k-means preformed similarly across the different environments with comparable amounts of dense noise, making it difficult to identify the footprints in each environment. Adaptive thresholding produced segmented foot- prints coupled with sparse areas of noise. The clustered points that identify each footprint showed the location of each footprint well but did not fully identify the footprint feature. The footprint was only partially represented by points that met the adaptive thresholding criteria, thus resulting in clustered points. The problem of partly segmented footprints is highlighted when looking at the surrounding noise because this limits the ability to further populate the partly segmented footprints. ATFD segmented the footprints with the least amount of sparse noise. However, ATFD produced areas that overlapped the boundary of the identified footprints, which could also be considered noise. The location of the footprints, the density of the clustered points representing them, and the lack of sparse noise facilitated visual identification.

TABLE 3. Comparison of different segmentation techniques for the detection of thermal footprints.

Therefore, Otsu, *k*-means and adaptive thresholding generally included too much noise or showed insufficient accuracy in the location of the segmented footprints to identify them as features. In comparison, across all four environments, ATFD identified the footprints with the greatest location accuracy and least sparse noise. The challenges associated with each type of environment are discussed in more detail below.

B. GRASS
The segmentation of the footprints on grass using Otsu and *k*-means showed that some of the footprints were segmented enough for visual identification, but others were surrounded by too much noise. Adaptive thresholding produced some clustered points that could be identified as footprint features, but there were very few points in each cluster, resulting in a spread of speckled noise. ATFD achieved the segmentation

of the footprints on grass with no sparse points, which can be attributed to the parameters used to assess each footprint. In terms of overall performance, Otsu and *k*-means showed the poorest results on grass, with no footprints identified in images 6 or 7, and overall averages of 41% and 43%, respectively (Table 4). Adaptive thresholding identified three or four of the footprints in most images but only two in image 6, resulting in an average of 82.5%. In contrast, ATFD identified all footprints in all 10 images of the grass environment.

Otsu and *k*-means achieved low MeanIoU scores, with averages of 0.0827 and 0.0908, respectively, due to the penalisation of the many false positives (Table 5). Adaptive thresholding performed better, with an average score of 0.11, but ATFD scored an average of 0.71 with a high of 0.83, highlighting the much better classification of true positives and the elimination of false negatives.

TABLE 4. Number of footprints identified on grass (expressed as a percentage, for 10 individual images and the average).

Image number	Otsu	k-means	Adaptive	ATFD
1	75	75	100	100
2	50	50	100	100
3	75	75	100	100
4	50	50	75	100
5	50	50	75	100
6	25	0	50	100
7	0	25	100	100
8	50	25	75	100
9	25	25	75	100
10	25	25	75	100
Average	43	41	84	100

TABLE 5. MeanIoU of footprints identified on grass.

Image number	Otsu	k-means	Adaptive	ATFD
1	0.091375	0.10096	0.14878	0.7527
2	0.088949	0.09312	0.1832	0.8041
3	0.093718	0.093718	0.13572	0.78781
4	0.12585	0.12585	0.16349	0.69732
5	0.13172	0.13172	0.13703	0.5853
6	0.048831	0.039877	0.058263	0.40424
7	0.062706	0.060761	0.1961	0.82968
8	0.083644	0.084288	0.05389	0.73101
9	0.091405	0.0067627	0.041286	0.75992
10	0.090135	0.090135	0.048747	0.75146
Average	0.0908333	0.08271917	0.1166506	0.710354

TABLE 6. MeanBFScore of footprints identified on grass.

Image number	Otsu	k-means	Adaptive	ATFD
1	0.065135	0.075259	0.16935	0.7868
2	0.061498	0.064014	0.15014	0.87116
3	0.06055	0.06055	0.15168	0.86224
4	0.084471	0.084471	0.1364	0.68831
5	0.076121	0.076121	0.079682	0.48559
6	0.044335	0.032353	0.0626	0.56276
7	0.03504	0.032469	0.27319	0.9345
8	0.12474	0.1246	0.21125	0.83273
9	0.1324	0.12358	0.13747	0.84247
10	0.14515	0.14515	0.18966	0.81584
Average	0.082944	0.0818567	0.1561422	0.76824

Otsu and k-means achieved low average MeanBFScores of 0.082 and 0.081, respectively, showing that neither method is suitable for the identification of clear footprint boundaries (Table 6). Adaptive thresholding scored 0.15, although poor results were achieved in images 5 and 6, with scores of 0.07 and 0.06, respectively. In contrast, ATFD achieved an average of 0.76 and a high of 0.87, showing a clear increase in performance.

C. GRASS WITH FOLIAGE

The segmentation of the footprints on grass with foliage using Otsu and k-means again showed the effect of excessive noise, and the foliage added further clustered noise making it even harder to visually distinguish the footprints (Table 3). Adaptive thresholding produced speckled noise but also clustered segments representing the foliage. Some footprints were identified, but were often broken into smaller

TABLE 7. Number of footprints identified on grass with foliage (expressed as a percentage, for 10 individual images and the average).

Image number	Otsu	k-means	Adaptive	ATFD
1	50	25	75	100
2	25	25	75	100
3	75	50	100	100
4	75	75	100	100
5	25	25	100	100
6	0	0	100	100
7	75	75	100	100
8	75	75	100	100
9	75	75	100	100
10	75	50	100	100
Average	55	47.5	94	100

TABLE 8. MeanIoU of footprints identified on grass with foliage.

Image number	Otsu	k-means	Adaptive	ATFD
1	0.063068	0.059636	0.141	0.75745
2	0.058533	0.058533	0.13517	0.76465
3	0.07478	0.07478	0.17416	0.7563
4	0.088697	0.085373	0.16064	0.79995
5	0.081025	0.078782	0.12763	0.74684
6	0.069534	0.067487	0.14041	0.74474
7	0.082946	0.086779	0.20815	0.72892
8	0.083129	0.083129	0.2248	0.76599
9	0.1104	0.10545	0.17189	0.79297
10	0.10129	0.096339	0.16199	0.78767
Average	0.0813402	0.0796288	0.164584	0.764548

segments making it challenging to distinguish a full footprint. ATFD achieved segmentation without speckled or clustered noise, and fully identified the footprints without breaking them into smaller segments. These results are reflected in the average scores, where Otsu and k-means again showed a poor performance (55% and 47.5% on average, respectively), adaptive thresholding achieved an average of 94%, but ATFD again achieved 100% in all images (Table 7).

Otsu and k-means achieved low mean IoU scores of 0.08 and 0.07, respectively, with corresponding high values of 0.11 and 0.10, again due to a large number of false positives (Table 8). Adaptive thresholding scored an average of 0.16 with a high of 0.22, whereas ATFD scored an average of 0.76 and a low of 0.72, again highlighting the accurate identification of true positives without false negatives.

Otsu and k-means achieved consistently low average MeanBFScores of 0.05 (Table 9). Adaptive thresholding scored an average of 0.14 and a high of 0.17, whereas ATFD scored a much higher average of 0.81 with consistent performance across all images, the lowest score being 0.76.

D. SAND

The segmentation of the footprints on sand using Otsu and k-means showed dense noise originating from the upper part of the image. The footprints in one half of the image were therefore easily identifiable with little noise, whereas the other half presented dense noise with no footprints. Adaptive thresholding segmented all the footprints but incorrectly identified the edge of the sandpit. Although the actual footprints

TABLE 9. MeanBFScore of footprints identified on grass with foliage.

Image number	Otsu	k-means	Adaptive	ATFD
1	0.041249	0.033743	0.17187	0.82573
2	0.024373	0.024373	0.17793	0.84235
3	0.049606	0.049606	0.13554	0.81587
4	0.060319	0.057558	0.12066	0.84351
5	0.040125	0.036448	0.14394	0.78562
6	0.044132	0.039746	0.15554	0.79465
7	0.065155	0.069885	0.1203	0.76493
8	0.059343	0.059343	0.13841	0.83586
9	0.08109	0.076538	0.10605	0.83254
10	0.054024	0.050953	0.14393	0.83714
Average	0.0519416	0.0498193	0.141417	0.81782

TABLE 10. Number of footprints identified on sand (expressed as a percentage, for 10 individual images and the average).

Image number	Otsu	k-means	Adaptive	ATFD
1	75	75	100	100
2	75	75	100	100
3	75	75	100	100
4	75	75	100	100
5	75	75	100	100
6	75	75	100	100
7	75	75	100	100
8	75	75	100	100
9	75	75	100	100
10	75	75	100	100
Average	75	75	100	100

TABLE 11. MeanIoU of footprints identified on sand.

Image number	Otsu	k-means	Adaptive	ATFD
1	0.072902	0.071299	0.46733	0.76697
2	0.066161	0.068994	0.43826	0.79141
3	0.072387	0.073711	0.45257	0.78481
4	0.069135	0.070203	0.46062	0.81781
5	0.065018	0.065811	0.43326	0.82003
6	0.062645	0.063527	0.41306	0.73161
7	0.058573	0.059171	0.42072	0.70066
8	0.066949	0.066949	0.37567	0.73613
9	0.068189	0.068984	0.38237	0.80458
10	0.061956	0.061956	0.33445	0.83449
Average	0.0663915	0.0670605	0.417831	0.77885

TABLE 12. MeanBFScore of footprints identified on sand.

Image number	Otsu	k-means	Adaptive	ATFD
1	0.070867	0.073834	0.30508	0.87026
2	0.07289	0.077384	0.30162	0.90607
3	0.079538	0.0808	0.26083	0.91794
4	0.09678	0.0942	0.2978	0.96621
5	0.10195	0.10383	0.38384	0.93489
6	0.088584	0.08581	0.36647	0.81404
7	0.096311	0.095043	0.40756	0.80135
8	0.10192	0.10192	0.2741	0.8294
9	0.10992	0.10566	0.33543	0.92743
10	0.082072	0.082072	0.21795	0.95452
Average	0.0900832	0.0900553	0.315068	0.892211

TABLE 13. Number of footprints identified on carpet (expressed as a percentage, for 10 individual images and the average).

Image number	Otsu	k-means	Adaptive	ATFD
1	25	25	100	100
2	50	50	100	100
3	50	50	100	100
4	33	33	100	100
5	25	25	100	100
6	0	0	66	100
7	0	0	75	100
8	0	0	0	100
9	0	0	66	100
10	25	25	100	100
Average	21	21	81	100

of 0.31 with a low of 0.21, whereas ATFD excelled with an average of 0.89 and a high score of 0.95.

E. CARPET

The segmentation of the footprints on carpet using Otsu and *k*-means showed dense noise originating from the top left corner of the image, making it impossible to fully identify all the footprints (although two were visible). Adaptive thresholding located all three or all four footprints (depending on the image) and represented them well in the form of sparsely clustered points, but also displayed sparse noise around the image particularly in the lower left corner. ATFD again yielded densely segmented footprints in the correct positions with no sparse noise.

Otsu and *k*-means both achieved an average of 21% detections based on the number of footprints identified per image, with a failure to identify any footprints in four of the images (Table 13) and a generally inconsistent performance (ranging from 0% to 50% across different images). Adaptive thresholding scored an average of 81% but was also very inconsistent, ranging from 0% in one image to 100% in six of the images. Again, ATFD was 100% successful in all images.

Otsu and *k*-means both achieved low average MeanIoU scores of 0.059 and 0.054, respectively, again due to excess noise (Table 14). In comparison, adaptive thresholding scored an average of 0.19 but over a wide range, from a high of 0.36 to a low of 0.00032. ATFD scored an average of 0.78 with a high of 0.82.

The MeanBFScore values of the Otsu and *k*-means methods were 0.0330 and 0.0318, respectively, with zero scores for images 6 and 8 reflecting its inability to detect any

were segmented, they were presented as a few clustered points that did not fully segment the whole footprint despite the correct location. In comparison, ATFD identified and densely segmented all the footprints, although the segmentation borders slightly overlapped the original. Both Otsu and *k*-means consistently identified 75% of the footprints in all images, whereas adaptive thresholding and ATFD identified all footprints in all images (Table 10).

The average MeanIoU scores for the Otsu and *k*-means methods were 0.066 and 0.067, respectively, due to the excess noise and consequential high rate of false positives (Table 11). Adaptive thresholding scored an average of 0.41 with a low of 0.33, whereas ATFD achieved an average score of 0.77 with a low of 0.7.

Both the Otsu and *k*-means methods achieved a MeanBF-Score of 0.09, as expected given the amount of noise in the images (Table 12). Adaptive thresholding scored an average

TABLE 14. MeanIoU of footprints identified on carpet.

Image number	Otsu	k-means	Adaptive	ATFD
1	0.04564	0.10157	0.30185	0.82321
2	0.063724	0.062687	0.35024	0.77738
3	0.064337	0.057271	0.36675	0.74733
4	0.051626	0.076259	0.21708	0.80407
5	0.030631	0.062273	0.24361	0.81408
6	0	0.045147	0.043924	0.78093
7	0.016423	0.045808	0.11174	0.79407
8	0.02112	0.04581	0.00032165	0.73505
9	0	0.044225	0.056743	0.78708
10	0.037037	0.00046244	0.30433	0.77238
Average	0.05946	0.054151	0.199658	0.783558

TABLE 15. MeanBFScore of footprints identified on carpet.

Image number	Otsu	k-means	Adaptive	ATFD
1	0.04564	0.04564	0.25708	0.92168
2	0.063724	0.060716	0.30113	0.9239
3	0.064337	0.057712	0.29877	0.89429
4	0.051626	0.057475	0.2173	0.85905
5	0.030631	0.030631	0.2487	0.9095
6	0	0	0.052591	0.85928
7	0.016423	0.016423	0.19161	0.91175
8	0.02112	0.02112	0	0.84157
9	0	0	0.050444	0.87705
10	0.037037	0.028557	0.32488	0.89744
Average	0.0330538	0.0318274	0.1942505	0.889551

contour boundaries (Table 15). Adaptive thresholding scored an average of 0.194 (also with a zero score for image 8), whereas ATFD averaged 0.889 with no individual zero scores.

IV. DISCUSSION

The concept of PUGTIFs was demonstrated through the instance of thermal footprints, which highlighted several advantages. The features were shown to be passively user-generated as well as inherently non-reliant on the scene content and illumination in several environments. The ability of thermal footprints to allow for a scaled monocular position was confirmed, allowing us to develop a system to estimate the user heading. Finally, these processes were combined to develop a method for the segmentation of the PUGTIFs in four environments.

A. ENVIRONMENTS

We found that PUGTIFs were available in four diverse environments. The grass environment is representative of many situations, including the middle of an open grass field, in which the features are far away. This would present a challenge for classical feature detection methods because the scene contains very little content and the salient features are far away. Even so, we have shown that a thermal footprint feature is available. Similarly, the grass with foliage environment simulates more complex grass scenes, such as forests or fields with foliage, but we have shown thermal footprints are present and can be detected. The carpet environment can represent several classical feature scenes, including a plain wall in a room. The classical feature approach would pick up poor features and a lot of noise due to the absence of significant scene content. However, thermal footprints are suitable features even in this environment. Similarly, the sand

environment represents a desert scene, which provides very little content for classical feature detectors, but we have shown that thermal footprints are present in this context too. Representation in all these environments shows that PUGTIFs are available to be utilised. We have described three methods that are required to utilise thermal footprints.

B. MONOCULAR SCALED LOCALISATION

The first method is the estimation of monocular scaled position. The ability to estimate the position accurately and thus determine the scale is entirely dependent on the initial segmentation phase of the framework, hence the performance is reflected in the overall ATFD scores. The score assessing the boundary of the footprints gives a good indication as to the accuracy of localisation. This is because the estimation of monocular scaled position is dependent on the correspondence between a 3D world model of the foot and the associated 2D image of the thermal footprint. The average MeanBFScores across each environment were 0.77, 0.82, 0.90 and 0.89, showing that a strong 2D model was available to use as a representation of the 3D counterpart. Therefore, we can expect a good estimate of the monocular scaled position. This is supported by the detection rate of ATFD, which inherently uses the estimated scale and frequently achieves average scores of 100%.

C. USER HEADING

The second method is the estimation of user heading. The success of this method can be determined by looking at the average ATDF scores, which depend on the orientation estimation method. The scores, sometimes averaging 100%, reflect the ability of the method to accurately locate a search area for the next footprints.

D. SEGMENTATION

The combination of the first two methods gives rise to the ATFD approach, which allows the segmentation of thermal footprints while also estimating the scale and orientation. We compared our method in terms of segmentation performance to three other common segmentation techniques in four different laboratory environments. We found that the new method was superior and also necessary to achieve the correct segmentation of thermal footprints.

In grass, the segmentation of this feature using current methods achieved a maximum MeanIoU score of 0.11, whereas ATFD achieved a score of 0.76 as well as higher identification and boundary scores. Similarly, the ATFD method outperformed its competitors in grass with foliage, achieving a MeanIoU score of 0.76 and a MeanBFScore of 0.81, compared to maximum values of 0.16 and 0.14, respectively, for the other methods. In the carpet environment, ATFD achieved a MeanIoU score of 0.78 compared to a maximum of 0.19 for the other methods, confirming that it is necessary for feature detection in an indoor carpeted environment. In the sand environment, ATFD segmented the thermal footprint feature with a MeanIoU score of 0.77, compared to a maximum of 0.4 for the other methods. By comparing

multiple threshold segmentation methods in several different environments, we confirmed that ATDF is necessary for the detection of thermal footprint features. The efficiency with which segmentation was achieved reflected the ability of the individual methods to estimate monocular scaled position and user heading.

We have demonstrated the application and feasibility of PUGTIFs, i.e. features that are consistently user generated as long as the sensor can measure the heat difference between the platform and environment. We used the instance of a thermal footprint to demonstrate the availability of PUGTIFs in four different environments. We confirmed that PUGTIFs can be used to estimate a scaled position using only one camera and from a single image, followed by the determination of user heading. Finally, these methods were combined to segment thermal footprints, achieving much higher MeanIoU and MeanBFScore values than three competing segmentation methods.

V. FUTURE WORK AND APPLICATIONS

PUGTIFs allow the use of passively user-generated features that are consistently present as long as the thermal sensor can identify the signature. The application of this approach could be extremely beneficial considering current methods rely on classical features that are dependent on the scene and thus not repeatable consistently, particularly in the case of localisation and tracking, where feature quality and repeatability are fundamental to success [19]. PUGTIFs can be used in any environment given the availability of a thermal sensor. We have confirmed that such features solve for monocular scaled position as well as user heading, already showing the application of localisation from a single image. We have therefore explored the wider field of localisation, mapping and navigation applications, but further work is necessary to evaluate the broader possibilities of PUGTIFs. Future research should therefore consider the diverse applications of PUGTIFs and focus on those with the greatest potential for exploitation.

ACKNOWLEDGMENT

REFERENCES

[1] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Found. Trends Comput. Graph. Vis.*, vol. 3, pp. 177–280, Nov. 2007.

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

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

[4] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput.*, Nov. 2011, pp. 2564–2571.

[5] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput.*, Nov. 2011, pp. 2548–2555.

[6] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Oct. 2005, pp. 1508–1515.

[7] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[8] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, 2004.

[9] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Computer Vision Systems* (Lecture Notes in Computer Science), vol. 5008, A. Gasteratos, M. Vincze, and J. K. Tsotsos, Eds. Berlin, Germany: Springer, 2008, pp. 65–75.

[10] M. El-Gayar, H. Soliman, and N. Meky, "A comparative study of image low level feature extraction algorithms," *Egyptian Informat. J.*, vol. 14, no. 2, pp. 175–181, 2013.

[11] J. Heinly, E. Dunn, and J.-M. Frahm, "Comparative evaluation of binary features," in *Computer Vision—ECCV*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Germany: Springer, 2012, pp. 759–773.

[12] B. Ferrarini, S. Ehsan, N. U. Rehman, and K. D. McDonald-Maier, "Performance comparison of image feature detectors utilizing a large number of scenes," *J. Electron. Imag.*, vol. 25, no. 1, 2016, Art. no. 010501.

[13] B. Ferrarini, S. Ehsan, A. Leonardis, N. U. Rehman, and K. D. McDonald-Maier, "Performance characterization of image feature detectors in relation to the scene content utilizing a large image database," *IEEE Access*, vol. 6, pp. 8564–8573, 2018.

[14] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[15] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms (SODA)*. Philadelphia, PA, USA: SIAM, 2007, pp. 1027–1035.

[16] D. Bradley and G. Roth, "Adaptive thresholding using the integral image," *J. Graph. Tools*, vol. 12, no. 2, pp. 13–21, 2007.

[17] S. Suzuki and K. Be, "Topological structural analysis of digitized binary images by border following," *Comput. Vis., Graph., Image Process.*, vol. 30, no. 1, pp. 32–46, 1985.

[18] G. Csurka, D. Larlus, and F. Perronnin, "What is a good evaluation measure for semantic segmentation?" in *Proc. BMVC*, 2013, pp. 1–11.

[19] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE Robot. Autom. Mag.*, vol. 18, no. 4, pp. 80–92, Dec. 2011.

EDWARD JACKSON received the B.Eng. degree (Hons.) in electrical and electronic engineering from Northumbria University, Newcastle, U.K., in 2016. He is currently pursuing the Ph.D. degree in computer vision with Cranfield University. He joined the Autonomy Group, Centre of Electronic Warfare Information and Cyber, School of Defence and Security, in 2016. Previously, he was with the Rutherford Appleton Laboratory Space, Imaging Systems Department. He currently works in computer vision for localization and autonomous navigation.

LOUNIS CHERMAK (M'12) received the Engineering degree in electrical engineering from Polytech Grenoble, Grenoble, France, in 2010, and the M.Sc. and Ph.D. degrees in computer vision from Cranfield University, Cranfield, Shrivenham, U.K., in 2011 and 2014, respectively. From 2014 to 2017, he was a Research Fellow, and since 2017, he has been an Assistant Professor in computer vision and autonomous systems with Cranfield University, where he currently leads the Autonomy Group and Autonomous Systems Laboratory, Centre of Electronic Warfare Information and Cyber, School of Defence and Security. He has authored or coauthored several technical papers, and has been a reviewer for high-impact journals and conferences. His research interests include development and implementation of techniques and algorithms for visual-based aerial, ground, and space applications toward autonomous mobility. He was a recipient of the Selwyn Award from the Royal Photography Society, in 2017. He has been appointed as a Technical Member and the Co-Chair of one of the Autonomous Vehicle Technology Exploratory Teams, NATO Science and Technology Organization, since 2018.

• • •