Multi-Modal Target Detection for Autonomous Wide Area Search and Surveillance

Toby P. Breckon, Anna Gaszczak, Jiwan Han, Marcin L. Eichner, Stuart E. Barnes
School of Engineering, Cranfield University, Bedfordshire, UK

ABSTRACT

Generalised wide are search and surveillance is a common-place tasking for multi-sensory equipped autonomous systems. Here we present on a key supporting topic to this task - the automatic interpretation, fusion and detected target reporting from multi-modal sensor information received from multiple autonomous platforms deployed for wide-area environment search. We detail the realization of a real-time methodology for the automated detection of people and vehicles using combined visible-band (EO), thermal-band (IR) and radar sensing from a deployed network of multiple autonomous platforms (ground and aerial). This facilities real-time target detection, reported with varying levels of confidence, using information from both multiple sensors and multiple sensor platforms to provide environment-wide situational awareness. A range of automatic classification approaches are proposed, driven by underlying machine learning techniques, that facilitate the automatic detection of either target type with cross-modal target confirmation. Extended results are presented that show both the detection of people and vehicles under varying conditions in both isolated rural and cluttered urban environments with minimal false positive detection. Performance evaluation is presented at an episodic level with individual classifiers optimized for maximal each object of interest (vehicle/person) detection over a given search path/pattern of the environment, across all sensors and modalities, rather than on a per sensor sample basis. Episodic target detection, evaluated over a number of wide-area environment search and reporting tasks, generally exceeds 90%+ for the targets considered here.

1. INTRODUCTION

We present a real-time approach for the detection of people and vehicle targets using combined visible-band (EO), thermal-band (IR) and radar sensing from a deployed network of multiple autonomous platforms with results proven over extended wide-area evaluation trials. Autonomous target detection is an important aspect of autonomous platform deployment for wide-area search and surveillance.

Despite a range of prior work on both target detection and automated sensor understanding [1], wide-scale multi-sensor platform deployment raises a number of empirical and theoretical challenges largely unaddressed in prior studies [2–4]. Here, we present the autonomous detection of {people|vehicles} using an autonomous sensing network of aerial (UAV) platforms [5–7] and ground (UGV) platforms [8]. This is used as an empirical test case to explore the optimization of an observation model for the environment when distributed over multiple sensors, platforms and resulting target detection signatures. With an ever increasing use of remote deployed autonomous systems the problem of reviewing, processing and effectively reporting the sensor information they gather is one of growing significance with notable parallels in ground-based sensor networks [9, 10]

Prior work in the field is generally limited to the case of automated ground surveillance (the camera network case [9, 11–14]), isolated to the UAV based detection [5, 7, 15–18] with only very limited wide-scale consideration of multi-platform dual aerial and ground platforms within the same sensing scenario [19, 20]. Notably prior work does not address autonomy within this context [19, 20] and work considering multi-modal detection is in its infancy [7, 17, 18]. Of those aerial techniques addressing autonomous target detection most borrow heavily from the state of the art in generalised object detection such as [7, 17, 21, 22] but with often aerial detection specific enhancements (e.g. [7, 22]). Many state of the art approaches offer far from real-time performance as was recently noted in the pedestrian specific survey of [4] and also noted within the wider overviews offered by [1, 2].

By contrast here we address the key challenge of real-time detection within a deployed network of autonomous sensing platforms that transit a given environment (both ground and aerial). Furthermore we challenge prior evaluation approaches based on per-sample or per-signature performance (e.g. [2–4]) to consider the concept of episodic evaluation (i.e. target detection over the entire task/mission). Essentially we address the following key issue with relation to such multi-platform deployment:- given an search path and sampling strategy for the platform within the

environment (including multi-passes/samples per target), the characteristics of the sensors (constrained by weight/power/bandwidth) and *a priori* precision/recall characteristics of multiple classifiers how can episodic performance be maximized (Section 2).

Our people and vehicle targets are automatically detected using a multi-stage classification approach combining information from on-board {visible, thermal} (Unmanned Air Vehicle, UAV) and {visible, thermal, radar} (Unmanned Ground Vehicle, UGV) sensing using an ensemble of variable signature classifiers (Section 2.1 and 2.2). These detections are combined temporally and spatially to facilitate consistent wide-area target detection and real-time situational awareness (Section 2.3). Each target detection is reported with a varying detection confidence, derived directly from multi-modal signature classification, further facilitating target prioritization and platform re-tasking where further localized sampling for target confirmation may be required (Section 2.3). Extended results are presented from three wide-area system trails carried out in the UK, including the MoD Grand Challenge (2008) and subsequent wide-area testing (Section 3). Our approaches are shown to be robust to highly variable imagery and variable conditions resulting in high accuracy target detection and localization within the environment (Section 4).

Overall, the under-taking of multi-platform search and surveillance encompasses a wide-range of related system technologies relating to the platforms in use, wide area communications, autonomous platform tasking and effective visualization of detected targets for effective environmental awareness. This paper solely concentrates on the aspects of multi-modal target detection as developed, and effectively demonstrated for use in such a system. These other aspects of the system are not detailed in this study other than the brief overview provided in Section 2.1.

2. WIDE AREA SEARCH AND SURVEILLANCE

We present our work on this topic in three stages. First, we present a brief overview of the autonomous sensor platforms in use and specifically their multi-modal sensing capabilities (Section 2.1). The focus of this paper, namely the autonomous target detection within this wider system, is presented in Section 2.2 followed by an overview of how this capability fits into target reporting and analysis over a wide-area search and surveillance task (Section 2.3).

2.1 Sensor Platforms

The demonstrator system used for this work comprises of one Unmanned Ground Vehicle (UGV) and up to several Unmanned Air Vehicle (UAV) platforms communicating to a central ground control station via a combination of a high-power 2.4GHz radio network configured for ad-hoc data routing in an inter-platform over a mesh topology [23] and non-line of sight Coded Orthogonal Frequency Division Multiplexing (COFDM) dedicated network links [24].

Figure 1. Marshall System Design Group UGV (left); Blue Bear Systems Research UAV (right)

The UGV, developed by Marshall System Design Group (Petersfield, UK), comprises a 250Kg ruggedized platform capable of 3-6.5mph speed and operational endurance (time on mission) $>$ 90 min from on-board electrical power (Figure 1 left). The platform is designed to follow GPS way-points tasked via the communications network and operate using a pathway detection obstacle avoidance capability [8]. For such collision avoidance the platform is equipped with both a forward facing drive camera and time-of-flight planar laser scanner (Figure 1 left). For multi-modal target detection the platform is equipped with an un-cooled far infrared camera (*Thermoteknix Miricle 307k*, spectral range: 8-12μm), a visible-band colour camera (*Visionhitech VC57WD-24*, spectral range: $\sim$400-700nm) and a forward-facing TRW AC100 medium range radar (as per [25]).

The Unmanned Air Vehicle (UAV) platform(s), developed by Blue Bear Systems Research (Bedford, UK), is a variant on their Blackstart platform with a 1m+ wingspan, operating weight of $<$1.8kg and operational duration $>$ 40 minutes flight duration [6]. Control is via an on-board auto-pilot control system from which the platform can be tasked via GPS way-point control including autonomous launch and recovery. For target detection the current platform is equipped with dual gimbled cameras and a bespoke sensor selection:- an un-cooled far infrared (thermal-band) camera (*Thermoteknix Miricle 307k*, spectral range: 8-12μm) and a

visible-band colour camera (*PCB-685B 1/3″ Sony Interline CCD*, spectral range: ∼400-700nm). Earlier versions of the platform were equipped with a fixed visible-band colour camera (*Sony 1/3″ Sony Interline CCD*) positioned at 45° angle to the horizontal in the direction of flight [5]. The UAV operates at approximately a 60m altitude.

Both platforms provide 1Hz image feeds from both sensors at PAL resolution, compressed to a 32Kb per image via JPEG 2000 [26], for transmission back to the ground station for processing using either ground or aerial specific autonomous target detection approaches.

2.2 Autonomous Target Detection

Autonomous target detection is performed on the image feeds received from the remote platforms using a combination of object detection approaches specifically targeted to the detection of people and vehicles from both a ground-based (UGV) and aerial (UAV) perspective view over the varying sensor modalities. The approaches in use are specifically selected due to their real-time performance on commodity hardware.

2.2.1 Ground Detection - People and Vehicles

Our ground detection approaches, operating from the UGV platform for people and vehicle detection, operate using a common two stage approach of 1) fast candidate localization using a cascaded Haar classifier and 2) secondary target confirmation using Support Vector Machine (SVM) classification. This approach is used for images from both the visible-band and thermal-band cameras on the UGV.

Fast Candidate Localization first identifies potential candidate targets within sub-regions of the image using cascaded Haar classifiers [27, 28]. Cascaded Haar classifiers were firstly proposed sometime ago by [27] with later improvement by [29] with the primary aim of face detection of faces [28]. Despite their maturity against more recent contemporary techniques [2, 3, 30, 31], they remain one of the few real-time detection approaches [27, 32] capable of operating without prior foreground segmentation and hence from a moving platform [4, 33].

The concept is to use a conjunctive set of weak classifiers to form a strong classifier - in this instance, a cascade of boosted classifiers applying Haar-like features. These Haar features are essentially drawn from the spatial response of Haar basis functions and derivatives (hence Haar-like features [29]) to a given type of feature at a given orientation. In practice these features are computed as the sum of differences between varying rectangular sub-regions at a localised scale which although limited in scope as individual features can be computed extremely efficiently (relying only on integer mathematics). Individually, they are weak discriminative classifiers but when combined as a conjunctive cascade a powerful discriminative classifier can be constructed capable of recognising common structure over varying illumination, base colour and scale [28, 33]. The cascaded classifier is trained via boosting, specifically AdaBoost [34], to select a maximally discriminant subset of these Haar-like features from the exhaustive and over-complete set. This subsequently acts as a multi-stage cascade [27]. In this way, the final cascaded Haar classifier consists of several key simpler (weak) classifiers that all form a stage in the resultant complex (strong) classifier. These simpler classifiers are essentially degenerative decision-tree classifiers that take the Haar-like feature responses as input to the weak classifiers and return a boolean pass/reject response. A given region within the image must then achieve a pass response from all of the weak classifiers in the cascade to be successfully classified as an instance of the object the overall strong classifier has been trained upon. The classifier is then evaluated over a query image at multiple scales and multiple positions using a sliding search window approach over the image [27]. Despite this apparent exhaustive search element of the classifier, the nature of the cascade (sorted in order of most discriminative feature) allows the early rejection of the majority of such search windows with only a minimal sub-set of the features present in the cascade being evaluated. As a result we achieve the *"fast rejection"* of the majority of image search windows using only minimal computational effort. In this way the Haar cascade classifier thus combines successively more classifiers in a cascade structure which eliminates negative regions as early as possible during detection but focuses attention on promising regions of the image. This detection strategy dramatically increases the speed of generalised object detection whilst providing an underlying robustness to changes in scale and maintaining achievable real-time performance [27, 28].

Secondary Target Confirmation takes every search window identified as target candidate in the first stage and uses a secondary SVM classifier to perform target confirmation. Each search window is first re-sized to common patch size, $r \times c$, dependent on the target type, *{people, vehicle}*. From this re-sized patch, a feature vector, $\vec{v_i}$, constructed of the Laplacian filter response at each pixel location, p_i, for a given patch

Figure 2. Example people target detection in thermal-band imagery under varying ambient thermal conditions.

size, $r \times c$, such that $\vec{v_i} = \{f_{laplace}(p_i), ..., f_{laplace}(p_{rc})\}$ where $f_{laplace}()$ is the 2D Laplacian response at a given pixel location using a 3×3 filter kernel [35]. This feature vector forms the input to a two-class SVM classifier, $\{target, !target\}$, for a given target type (following the approach of [36]). This SVM classifier is trained, using a RBF kernel, with grid-based kernel parameter optimization, within a cross-validation based training regime [34].

Classifier Generation is performed, following the aforementioned methodology, for both people and vehicle by performing training over a set of manually labeled positive and negative examples. For people detection, we use a data-set of approximately 2000 positive examples (people) and approximately 11,000 negative (non-people) examples randomly selected from the same source imagery. The negative set is again randomly sub-sampled to generate 2000 negative (non-people) examples for each stage of the cascaded Haar classifier training whilst it is used in full for training the secondary SVM classifier with the addition of another 2400 negative examples. This training procedure is performed for both visible-band and thermal-band imagery to generate a trained classifiers for detection that operates independently in each modality. For thermal imagery an additional two-stage classifier pair is generated using only the upper-torso portion of the human body to facilitate detection in some partial occlusion cases. A subset of the training examples used for people detection in the thermal-band imagery case is shown in Figure 3 (left) where we see a variety of body pose, environment clutter and ambient thermal conditions over the imagery.

Vehicle detection is handled slightly differently whereby we use a part-model based approach, similar in principle to that of [3] but realized using a set of disjoint two-stage classifier pathways, based on the same *Fast Candidate Localization [28]* $\rightarrow$ *Secondary Target Confirmation* [36] format outlined previously, for each of a number separate vehicle parts. Vehicle detection is based on the discrete detection of one or more of the set of vehicle sub-parts *{wheel, front, rear, side}*. As discussed later in Section 2.3, detection of one or more of these parts is used translated as varying confidence of vehicle present.

For vehicle detection, we use a data-set of approximately 600 positive examples (per sub-part) and approximately 12,000 negative (non-vehicle) examples randomly selected from the same source imagery. The negative set is again randomly sub-sampled to generate 2000 negative (non-vehicle) examples for each stage of the cascaded Haar classifier training whilst it is used in full for training the secondary SVM classifier with the addition of another 2400 negative examples. Here, this training procedure is performed for only visible-band imagery to generate a trained classifiers for this modality only. A subset of the training examples used for the sub-part detection is shown in Figure 3 (right) where we see examples of wheels, front/rear bumpers and vehicle sides.

Figure 2 illustrates the use of this approach for thermal-band people detection under varying ambient thermal conditions, with minimal false-positive detection, whilst part-wise vehicle detection in visible-band imagery is shown in Figure 4.

2.2.2 Ground Detection - People in Buildings

In addition to generalized people detection (Section 2.2.1), we employ an additional specific detection technique targeting the detection of people within open building orifices (apertures) occurring in a generalized urban environment. We use a two stage approach to first detect potential building apertures within the environment using visible-band sensing and then check for the presence of either a human torso or a complete human body outline using thermal-band sensing.

Building Aperture Detection is performed on the visible-band image via a combination of edge detection [37], image morphology, geometric reasoning and colour-difference based candidate rejection. Firstly edge detection performed via [37] and post-processed

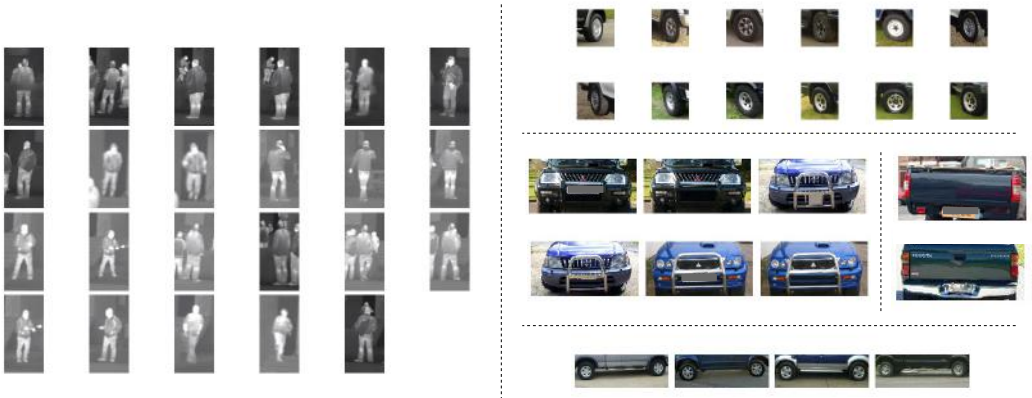


Figure 3. Example labeled training data examples for thermal-band people detection (left) and visible-band vehicle subpart detection (right)

Figure 4. Example part-wise vehicle detection in visible-band imagery under varying poses.

with traditional image morphology [35]. A set of straight lines are then extracted from this set using the Principle Components Analysis (PCA) driven approach of [38]. Following this approach, the post-processed edge detection results are split into connected segments and for each segment a covariance matrix is calculated from its pixel positions. PCA on this covariance matrix is used to determine whether the segment is a straight line, independent of its orientation, based on its second eigenvalue being less than a given threshold, $\tau_{straight}$. The resulting set of straight lines are divided into two sets based on their angle to the image horizontal, θ, as either a horizontal line relative to likely orientation of a building to the camera ($150 \leq \theta \leq 165$) or vertical ($80 \leq \theta \leq 100$). These two sets of lines, horizontal and vertical, are then searched to find a maximally consistent set of parallel vertical line pairs that additionally have a maximum of two co-joining horizontal lines to form a complete or semi-complete rectangular/parallelogram. This set is finally filtered based on size (relative to the image dimensions) and on the ratio of height to width, $r_{aperture} = height : width$, such that $r_{aperture} \leq 1.50$ for window apertures and $r_{aperture} \leq 1.50$ and $r_{aperture} \geq 2.3$ for doorways. The result is a list of scene rectangles representing all of the potential building aperture candidates in the scene

Figure 5. Example scene rectangles (left) and identified building apertures (right)

as shown in Figure 5. In this example (Figure 5) we see a number of scene rectangles corresponding both to building features as well as open and closed building apertures identified by varied coloured line overlays.

Subsequently, this set of scene rectangles (e.g. Figure 5) is further filtered based on an calculating the mean, $m_{aperture}$, and standard deviation, $\sigma_{aperture}$, of brightness (greyscale intensity) for the pixels on the two diagonal lines joining opposite corners. If the corresponding scene region appears two bright, $m_{aperture} > \tau_{mean}$, or the distribution of values is to wide, $\sigma_{aperture} > \tau_{\sigma}$, then this rectangle is rejected as a potential open aperture in the building. A resulting subset of open building apertures is identified in the scene as illustrated by the white (4 sides detected) and blue (three sides detected) rectangular overlays in Figure 5 (right).

Figure 6. Identified building apertures and corresponding human traces (thermal) used for training

Human Trace Detection makes use of the co-registered thermal-band image to detect the presence of a person within the identified building aperture (details Section 2.3.1). The sub-region in the thermal-band image corresponding to each building aperture is extracted, thresholded using the mean pixel value of the region and re-scaled to $width \times height = 50 \times 50$ pixels. From this re-sized patch, a feature vector is constructed to capture both the shape distribution of any thermal trace within the building aperture and the aspect ratio of that aperture (prior to re-scaling, $r_{aperture}$). This feature vector, $\vec{v_i}$, is constructed by concatenating both the vertical (column-wise) and horizontal (row-wise) histogram projections of this re-scaled thesholded image region with the aspect ratio for each detected aperture i, $\vec{v_i} = \{\vec{histogram}_{vertical}, \vec{histogram}_{horizontal}, r_{aperture}\}$. $\vec{histogram}_{vertical}$ and $\vec{histogram}_{hoziontal}$ are both 50 element histogram projections resulting in a feature vector representation of 101 for each aperture. This feature vector forms the input to a four-class neural network classifier, $\{full-body, half-body, empty, non-human\}$ detecting full or partially occluded people whilst ignoring apertures which are empty (no thermal trace) or contain a non-human (shaped) thermal trace. This neural network classifier is trained, using a standard sigmoid activation function via backpropogation, within a k-fold cross-validation based training regime ($k = 5$). [34]. The cross-validation based training approach is used to select a maximally performing three-layer network topology with a single hidden layer comprising of 28 hidden nodes. Training was performed using approximately 3,500 examples split evenly over the four possible classes. A subset of the examples used for training are shown in Figure 6 (inset) together with a range of varying building apertures from which the training set where taken (Figure 6). Example detections of humans within building apertures are illustrated in Figure 7 where we see a range of such detections in both windows and doorway apertures.

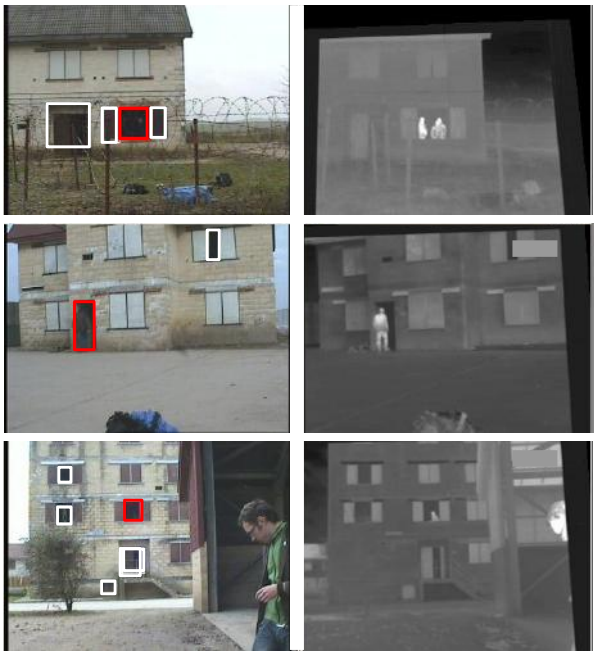


Figure 7. Identified building apertures (visible-band, white boxes) and those apertures thermal-band human traces (left) shown within the original aperture (visible-band, red).

2.2.3 Aerial Detection - People and Vehicles

Our corresponding aerial detection approaches, for both people and vehicle targets, have been previously reported in depth in [5, 7]. This work [5, 7] uses a similar two stage approach of 1) fast candidate localization using a cascaded Haar classifier and 2) secondary target confirmation using either multivariate Gaussian shape matching (people) or thermal trace detection (vehicles). This approach is used for images from both the visible-band (vehicles) and thermal-band cameras (people) on the UAV.

The automatic detection of vehicles is based on using multiple trained cascaded Haar classifiers, each following the approach outlined in Section 2.2.1, each trained on varying vehicle orientations with secondary confirmation in thermal imagery based on segmented *"thermal race"* analysis. Furthermore, the related approach for people detection is similarly based on multiple cascaded Haar feature classifiers (Section 2.2.1) with secondary multivariate Gaussian shape matching on the thermal contour of the person. Further detail is reported fully in [5, 7].

The results presented show the successful detection of vehicle and people under varying conditions in both isolated rural and cluttered urban environments with minimal false positive detection (Figures 11 - 15).

2.3 Target Analysis

Based on our target detection approaches over varying modalities we now address the key issues of a) multi-modal integration within the sensor suite of a given sensor platform (Section 2.3.1) and b) target position estimation relative to a given sensor platform deployed within the environment (Section 2.3.2).

2.3.1 Multi-modal Integration

Multi-modal integration occurs at two levels across the two sensor platform types, UGV and UAV. The visible-band and thermal-band imagery are co-registered using a 2D homography mapping (for UGV and UAV) whilst the radar field-of-view is calibrated against that of the imaging sensors (UGV only). This approach is similar to that used in the related dual visible/thermal band work of [7, 39] and contemporary work of [17].

Following the approach outlined in [39], we recover the transformation between the two image planes, for visible-band camera and thermal-band camera, permitting the use of common reference frames for the two camera sensors in terms of *"in image"* target positioning. This transformation is denoted as the *planar homography* that projects one image plane (thermal) to the other (visible) and is readily expressed in terms of the matrix transformation:

$$a_V = sHa_I \tag{1}$$

where a_{I_i} is a point in the thermal (infrared) image and a_{V_i} is the corresponding point in the visible image (i.e. same point in the scene). Parameter s is an arbitrary scale factor, H is a 3×3 transformation matrix and both points are homogeneous coordinates, $a_{V_i} = \{x_{V_i}, y_{V_i}, 0\}^T$ $a_{I_i} = \{x_{I_i}, y_{I_i}, 0\}^T$. In order to compute homography matrix H, four planar points in both views are selected (i.e. four correspondent points in the thermal image, $\{a_{I_i} : i \in 1..4\}$ and in the visible image, $\{a_{V_i} : i \in 1..4\}$) from which H is recovered by solving the resulting least-squares optimization posed by Eqn. 1 over the two correspondent point sets [40].

In Figure 8 we can both visible-band and thermal-band images from the UGV platform (left/middle) and the resulting overlay of one image onto the other (thermal onto visible) using an image to image plane homography mapping (right). This effectively facilitates the creation of a resulting four channel multi-modal, or extending spectral range, image comprising of the original three RGB colourspace channels of the visible-band image (spectral range: ∼400-700nm) and an additional fourth thermal channel (spectral range: 8-12μm) (see Figure 8 (right)). This allows targets detected within the image plane geometry of either sensor (Section 2.2) or specific image regions (Section 2.2.2) to be readily considered across these two sensing modalities in subsequent analysis (Section 2.3.1). Looking at the detail of the resulting homography image registration of Figure 8 (left), we can see that the visible and thermal images do not completely align for objects at all scene depths due to parallax between the images [39, 40] (e.g. Figure 8, left - edges of scene buildings). This limitation is similarly noted in earlier work [7, 17, 39] but empirically this approach is sufficient for level of spatial accuracy required for such cross-spectral sensor mapping in this application.

On the UGV platform, in addition to visible and thermal-band sensor co-registration, the radar sensor is similarly calibrated to the reference frame of the visible-band camera using the sensor fusion technique described in [25] (TRW, Solihull, UK). This sensor is used to both confirm target presence at given scene location (based on radar return) and for distance to target estimation.

2.3.2 Target Position Estimation

Based on automated detection (Section 2.2), integrated over multiple sensing modalities (Section 2.3.1), target position is initially known with *"sensor space"* (i.e. position within the image or radar field of view). Whilst target position can be recovered from the radar sensor on the UGV platform this offers only a ground plane distance capability (not for targets above ground level, e.g. Section 2.2.2) and is not available for target positioning from the UAV platform. Consequently, target position is estimated based on the principles of photogrammetry together with knowledge of the perspective transform under which targets are imaged and an assumption on the physical (real-world) dimension of a target in one plane.

All targets are imaged under a the standard perspective projection [35] as follows:

$$x = f\frac{X}{Z},\ y = f\frac{Y}{Z} \tag{2}$$

where real-world object position, (X, Y, Z), in 3D scene co-ordinate space is imaged at image pixel position, (x, y), in pixel co-ordinate space for a given camera focal length, f. We assume both positions are the

Figure 8. Visible-band image (left), thermal-band image (middle) and resulting image registration (visible-thermal overlay) (right).

centroid of the object with (x, y) being the centre of the bounding box, of the image sub-region, for a target (object) detected in the scene (Section 2.2, e.g. Figure 2).

With knowledge of the camera focal length, f, the original object (target) position, (X, Y, Z), can be recovered based on (assumed) knowledge of either object width, $\triangle X$, or object height, $\triangle Y$ (i.e. the difference in minimum and maximum positions in each of these dimensions for the object). From the bounds of the detected targets (Section 2.2) we can readily recover the corresponding object width, $\triangle x$, and object height, $\triangle y$, in the image. Based on this knowledge, rearranging and substituting into Eqn. 2 we can recover the depth (distance to target, Z) of the object position as follows:

$$Z = f' \frac{\triangle Y}{\triangle y} \tag{3}$$

Knowing Z via Eqn. 3, we can now substitute back into Eqn. 2 and with knowledge of the object centroid in the image, (x, y), we can recover both X and Y resulting in full recovery of real-world target position, (X, Y, Z), relative to the sensing platform. In Eqn. 3, f' represents focal length, f, translated from standard units, mm, to focal length measured in pixels:-

$$f' = \frac{width_{image} \cdot f}{width_{sensor}} \tag{4}$$

where $width_{image}$ represents the width of the image (pixels), $width_{sensor}$ represents the camera CCD sensor width (mm).

Crucially, if we now assume a fixed width, $\triangle X$, or height, $\triangle Y$, for our object we can recover complete 3D scene position relative to the sensor platform. For people detection we can assume average adult human height based on available medical statistics [41] whilst for vehicle detection, the dimensions of road vehicles have evolved to a relatively steady state in terms of both wheel-base and width [42, 43] facilitating generalised assumption for one or both of these dimensions. Despite the crudeness of this assumption, empirically it appears to work well for target positioning in the absence of radar based position estimation (Section 3). Furthermore it offers a passive, as opposed to active radar-based, position estimation for detected targets. Figure 9 illustrates the application of this approach to the position estimation, showing distance to target only, with an example human target that is detected using the approach outlined in Section 2.2.

As sensor platform position is known from GPS (Section 2.1) the target position relative to the sensor recovered using this approach, (X, Y, Z), is readily transformed into global position coordinates for target reporting within the search environment. For aerial detection (UAV platform) position is estimated, based on the same key principles but largely assuming a point target, using techniques previously reported in [7].

2.4 Target Reporting

Detected targets are reported based on position within the environment (Section 2.3.2), the sensing platform that obtained the detection and a derived confidence value associated to that target detection, δ_t. Recognising that even current state of the art techniques in generalised object detection are not perfect an generally incur both a false detection (false positive) and missed detection (false negative) rate that is non-zero [1–3, 28], this confidence value plays a crucial role within effective target reporting. It is based upon detection across multiple spatially integrated (co-registered) modalities (Section 2.3.1) each with their own associated classifier specific to a given modality on a given platform.

Multi-modal Target Co-occurrence: Target detections in any two modalities, m_i and m_j, are assumed to relate to the same physical scene target occurrence when the spatial overlap of the target detections is greater than a given overlap criteria, $\alpha_{(m_i,m_j)}$, calculated as the spatial union of both target detection sub-regions, $region(m, t)$, in modality m of target t as a fraction of largest such sub-region:

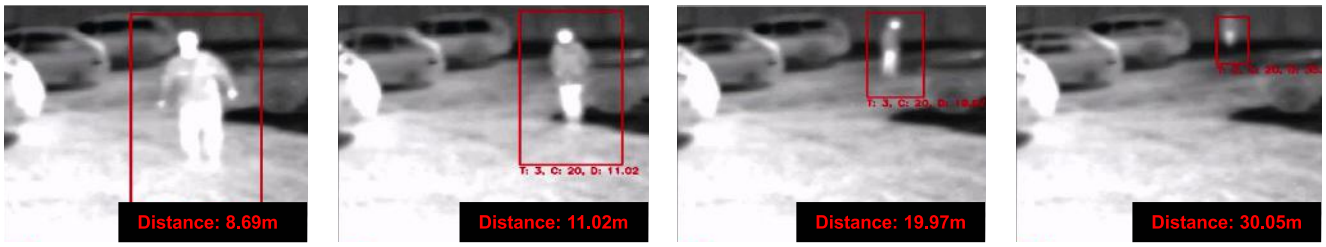


Figure 9. Photogrammetry facilitates the approximate recovery of a sensor to target distance for an example target (person) without any need for additional (active) range sensing.

$$\alpha_{(m_i,m_j)} = \frac{region(m_i,t) \cap region(m_j,t)}{max(region(m_i,t), region(m_j,t))} \quad (5)$$

where $\alpha_{(m_i,m_j)} = 1$ for perfect spatial overlap of one sub-region with another and $\alpha_{(m_i,m_j)} = 0$ for non-overlapping threats. This assumes spatial co-registration of all sensor modalities on a given platform using the approach outlined in Section 2.3.1. Sets of target detections found to be spatial coincident in the scene all contribute to the confidence value, δ_t, of the same target occurrence within the scene (where $\alpha_{(m_i,m_j)} \gtrsim 0.6$) . Otherwise they are assumed to represent multiple, physically separate scene targets (e.g. multiple people or vehicles in the scene).

Target detection confidence: The confidence value, δ_t, is associated to target t based on a weighted combination of classifier response, over all modalities of detection, m, associated with the target from a given sensor platform, p, as follows:

$$\delta_t = \sum_m \frac{w_{c_{(p,m\rightarrow t)}}}{\sum_m w_{c_{(p,m\rightarrow t)}}} f_{c_{(p,m\rightarrow t)}}(t) \quad (6)$$

where weight $w_{c_{(p,m\rightarrow t)}}$ is the relative weight of a given classifier, $c_{(p,m\rightarrow t)}$, trained for the detection of target, t, using modality, m. The classification response function, $f_{c_{(p,m\rightarrow t}}(t)$, representing the returned value from classifier $c_{(p,m\rightarrow t)}$ is bounded to range $\{0 \rightarrow 1\}$ following a probabilistic interpretation of target likelihood. The set $\{c_{(p,m\rightarrow t)} | m \in sensors(p)\}$ thus represents the set of all such classifiers, across all modalities m, for a given sensor platform, p, that detect target type, t. For example, people detection from the UGV platform would comprise of two Haar cascade classifiers (one thermal-band, one visible-band), two SVM classifiers (one thermal-band, one visible-band) and one radar response.

The target signature of t, comprising the sensor information from the platform, is assumed to encompass all available sensing modalities from a given platform synchronized to the same point in time (within a marginal time offset <100ms). Where a given sensor modality is not available on a given platform due to technical failure or deactivation its weight in the Eqn. 6 remains unchanged, and thus the influence of each modality in overall target detection remains unchanged. However, in this case the maximally achievable confidence of detection is now lower as could reasonably and practically be expected in cases where a target is detected but only by a subset of the potentially available sensing modalities.

Processing and Spatial Visualization: All sensor information (images/signals) are transmitted from the sensor platforms to the Ground Control Station (GCS) (Selex Galileo, Luton, UK) where they are processed in parallel prior to visualization as a real-time situational awareness map of the environment. The GCS is implemented using a 4-core/8-thread multi-core CPU (Intel Core i7) using explicit OpenMP task parallelization.

Target presence are recorded based on spatial position within the environment using the concept of a spatial occupancy grid [44, 45]. Spatially co-occurring targets, re-detected by the same platform over multiple sensor samples or from multiple sensor platforms traversing the same region of the environment, are merged to form a single target record. This follows a spatial Gaussian cell-weighting of each target occurrence to individual cell occupancy in the grid. The confidence associated to multiple spatially co-occurring target detections are similarly integrated to form high confidence target occurrences on the environment target map. This occurs unless multiple targets are explicitly detected at the same time by given platform (e.g. group of people). Target occupancy of the grid decays temporally, configurable by the user. Detected targets are displayed in real-time on the GCS as a series of icons with associated confidence and raw sensor infor-

Figure 10. Example target visualization showing both example threat icons (pink) and example target information (right).

mation, in addition to current sensor platform position, for operator review as illustrated in Figure 10.

Multi-platform Search and Tasking: The specified environment is searched either to provide an initial target location or longer-term target surveillance. Each platform performs a specified search pattern tasked over the communications networks as a series of GPS way-points. Based on a search area specified interactively via the GCS, this way-point tasking is automatically derived as a search pattern incorporating additional constraints such as road/ground layout (UGV) and launch/recovery points (UAV). These search taskings are designed to maximize effective environment sampling given platform speed and sensor sample rate in relation to environment coverage using techniques such as [46]. Furthermore platforms can be re-tasked to re-sample given areas of the environment or to obtain additional confirmation of target presence with additional sensing capabilities.

3. EVALUATION

Large-scale system-level evaluation is carried out over a series of wide-area evaluation trials facilitated at varying locations in the UK (UK MoD / Stellar Research Services Ltd, Southampton, UK). The work on automatic target detection is integrated into the wide-area sensing and surveillance system for these trials and this evaluation concentrates solely upon the evaluation of this system component within the system. Details of these trials are shown in Table 1 where we see the range of conditions and environments under which the autonomous target detection approaches proposed were evaluated. Each evaluation trial consisted of multiple large area search experiments with ground truth targets (people / vehicles) emplaced throughout the search environment at known (GPS recorded positions).

To this end, we consider evaluation on an episodic basis by aiming to optimize target detection performance for optimal detection of targets at least once in the environment, given an adequate sampling strategy of that environment by the platforms (Section 2.3). Essentially, we are considering a success criteria such that targets present are detected and mapped with reasonable confidence (Section 2.3), at least once per search mission (i.e. sensor platform deployment). The episodic approach follows [5, 7] rather than the traditional per-sample performance consideration of detecting every object in each sensor sample [2]. Individual classifier performance is optimized to maximize true positive detection with the consideration that false positives are tolerable provided they can be characterized as infrequent and randomly distributed in the environment. This facilitates the filtering out of such spurious false positive detections based on their associated confidence value when compared to that associated to a true positive target detected over multiple sensor samples, from multiple platforms, and spatially integrated upon the environment occupancy grid (Section 2.3).

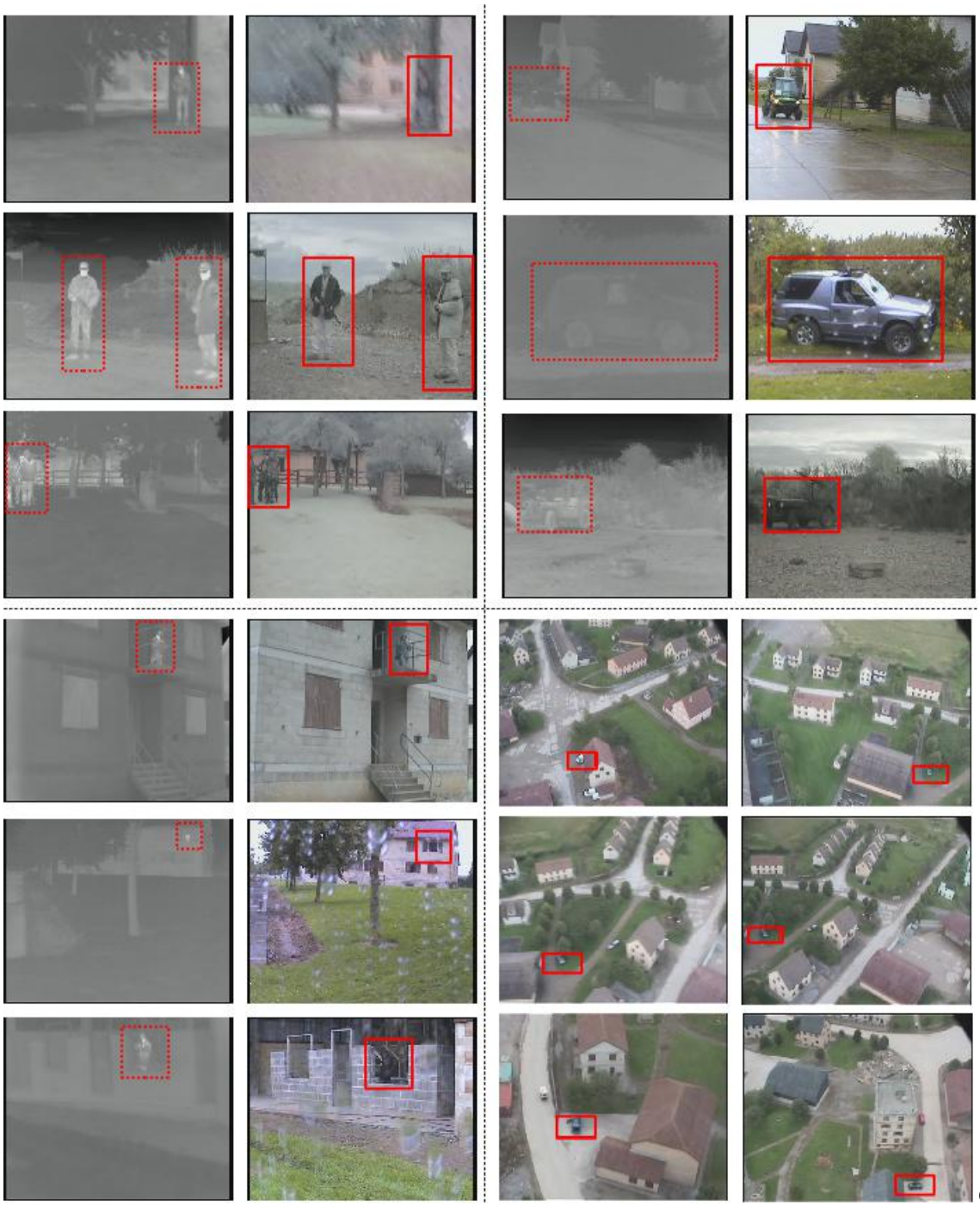
A sub-set of the sample detection results for these wide-area evaluation trials are shown in Figures 11 - 15 as a set of four quadrants in each case (which we will reference as Q1→Q4 clockwise from top left, i.e. $\frac{Q1}{Q4} \mid \frac{Q2}{Q3}$).

Figure 11 (Q1,Q2,Q4) shown target detection results for both people (Q1), vehicles (Q2) and specifically people within building apertures (Q4) in visible/thermal image pairs from the UGV platform. We see the detection of people in various poses around the environment, at different scales and levels including under varying weather conditions (Figure 11 Q1/Q4). In these examples, which form a sub-set of the target detections obtained in the MoD Grand Challenge (finale and proving events, August 2008), target detections are highlighted (in red) in both the visible (left) and thermal (right) imagery (Figure 11). Interestingly, we can see the thermal image driven person detector is robust to very poor quality optical imagery (Figure 11, Q1 - top right). A range of 4×4 type vehicles are detected in the examples of Figure 11, Q2) where we see detection independent of scale, orientation and make/model/configuration variants. Figure 11 (Q4) shows people detection within building apertures and in one such instance (middle) a detection at 40m+ distance in poor weather conditions. Figure 11 (Q3) shows the corresponding set of aerial vehicle detections, from an earlier single-sensor variant of the UAV platform, for this evaluation trial (Section 2.2.3) as reported earlier in [5].

Following the same quadrant layout as Figure 11, Figures 12 and 13 now present further supporting results from the second wide-area evaluation trial under

Evaluation Trial	Location	Search Area	Ambient Conditions	Environment
Site #1 (CDV/GC)	Wiltshire, UK	$\sim 0.5km^2$	$11-19^oC$, rain / clear	urban / suburban
Site #2 (NZF)	Wiltshire, UK	$1-2km^2$	$10-25^oC$, clear	rural / agricultural
Site #3 (CDV/CAT)	Wiltshire, UK	$2km^2+$	$2-10^oC$, overcast	dense urban / suburban

Table 1. Evaluation trials - environmental conditions



C

Figure 11. A range of both ground and aerial targets detected over Evaluation Trial 1, presented in {visible, thermal} imagery pairs as appropriate (UGV ground detection examples, top left/right and bottom left) and visible imagery only (UAV aerial detection examples, bottom right).

differing environmental conditions (Table 1). Target detections are highlighted (in red) in both the visible (left) and thermal (right) imagery (Figure 11). In Figures 12/13 (Q1,Q2,Q4) we can see target detection results for both people and vehicles from the UGV platform as visible/thermal imagery pairs. Figure 14 and 13 (Q3) shows the corresponding set of aerial people and vehicle detections, from the dual-sensor UAV platform (Section 2.1), for this evaluation trial (Section 2.2.3) as reported earlier in [7]. Furthermore, we can see both examples of multiple co-occurring target detection from the same platform (Figure 12 Q1 + Q 2 top, UGV) and between platforms (Figure 12 Q4, UGV + Figure 12 Q3 bottom, UAV and Figure 14 Q4 top/middle, UGV + Figure 14 Q3 top/middle, UAV).

The same quadrant layout is used again in Figures 14 and 15 which present further results from the third wide-area evaluation trial under differing environmental conditions (Table 1). Here target detections are highlighted (in red, green or yellow) in either the visible (left) and thermal (right) imagery depending on the primary sensing source of the target detection (Figure 14/15). In Figures 14/15 (Q1,Q2,Q4) we see target detection results for both people and vehicles from the UGV platform as visible/thermal imagery pairs. Figure 14 and 15 (Q3) shows the corresponding set of aerial people and vehicle detections, from the dual-sensor UAV platform (Section 2.1), as reported earlier in [7]. Despite the illustrated success of detection within this environment we can note the issue of *"thermal white-out"* occurring under certain environmental conditions (Figure 15 Q1,Q2) which give rise to reliance on visible-band target detection (prone to camouflage fabrication) or to strong thermal outlines for successful detection with the thermal sensor (Figure 15 Q1,Q2). Notably, this issue appears isolated to the UGV perspective view of the environment as UAV target detection under the same conditions appear unaffected (Figure 15 Q3 / 14 Q3).

Overall, we achieve an episodic target detection rate of $\sim 90 - 100\%$ based on the use of individual classifiers optimized as outlined previously. False positive detections are generally spurious and readily filtered out based on spatial cohesiveness and limited co-occurrence with the overall target reporting structure for environment search (Section 2.3). Whilst this detection rate may appear high, if we consider each individual target detection classifier (Section 2.2) having a true positive rate which is at worst only$\geq$70% it can be readily formulated that, over multiple sensor samples of the any given target in the environment, the resulting probability of non-detection asymptotically tends to zero. Each target is sampled multiple times, by multiple platforms, and on each sample occurrence the probability of true positive detection is $\geq$70% and similarly false positive detection $\lesssim$ 10%. Similarly, each sample is at least slightly *"different"* in terms of view onto the target, platform, angle, distance to target etc. If we thus assume loose sample independence under these conditions, the greater the number of sensor samples of a given single target that are *"presented"* for classification, the lower the resulting probability of not detecting the target becomes (as every independent classification has a chance of success $\geq$70%). This is argument is more strongly supported in cases where the target is highly representative of the class of targets (e.g. people) upon which the specific classifier was trained. Conversely, false positive detections over a random sequence of sensor samples will occur with low probability ($\lesssim$ 10%) and are therefore most likely to be spurious and spatially inconsistent with regard to environment target mapping (Section 2.3).

Given a sufficient sampling of the environment (Section 2.3), the known precision/recall characteristics of our chosen classification approaches [5, 7, 28] and our two-stage candidate detection / secondary confirmation approach (Section 2.2) we can thus achieve target detection to a relatively high level of episodic accuracy. These target detection results are subsequently visualized as set out in Section 2.3 where target position accuracy is generally found to be in the range $\pm 5m$ against ground truth using the techniques outlined in Section 2.3.2 (considering GPS error).

4. CONCLUSIONS

We demonstrate the integration of autonomous target detection for wide-area search and surveillance using multi-modal sensing across both aerial and ground sensor platforms. Based on a range of established automatic classification approaches, we detail both multi-modal integration (post-classification) to produce a per target confidence value and its spatial integration/mapping to a current visualization of the environment. A range of results are detailed from wide-area evaluation trials, under varying environmental conditions, that produce highly accurate target reporting within the proposed episodic evaluation framework ($\sim 90 - 100\%$ detection, spatial location $\pm 5m$). The use of multi-modal sensing, from multiple autonomous sensor platforms, is demonstrated and proven *"in the large"* extending prior work in the field for this type of wide-area search and target mapping activity.

Future work will investigate both the use of saliency for initial candidate target identification [47, 48], target

Figure 12. A range of both ground and aerial targets detected over Evaluation Trial 2, presented in {visible, thermal} imagery pairs as appropriate (UGV ground detection examples, top right/left and bottom left) and visible/thermal imagery only (UAV aerial detection examples, bottom right).

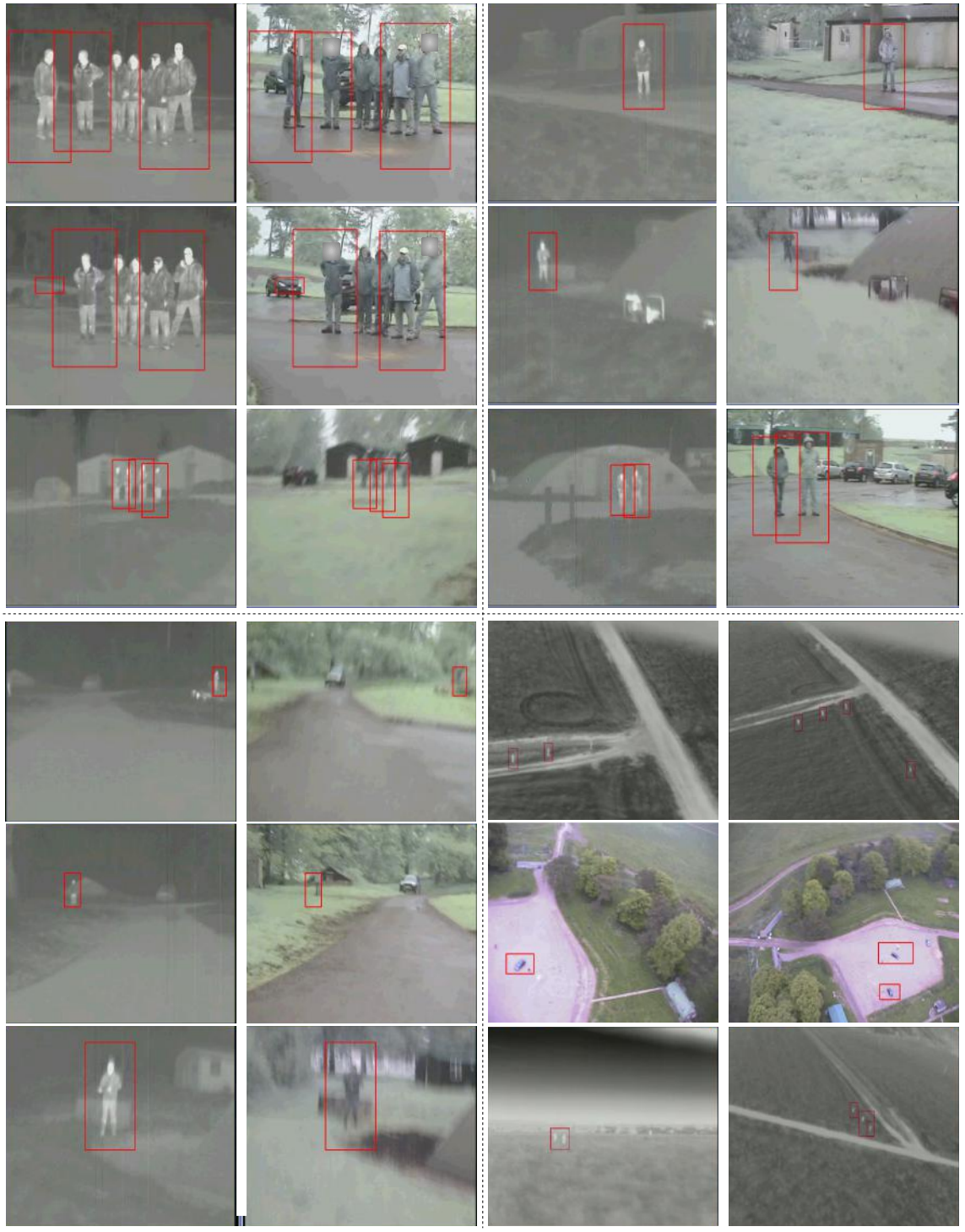


Figure 13. A range of both ground and aerial targets detected over Evaluation Trial 2, presented in {visible, thermal} imagery pairs as appropriate (UGV ground detection examples, top right/left and bottom left) and visible/thermal imagery only (UAV aerial detection examples, bottom right).

Figure 14. A range of both ground and aerial targets detected over Evaluation Trial 3, presented in {visible, thermal} imagery pairs as appropriate (UGV ground detection examples, top right/left and bottom left / UAV aerial detection examples, bottom right).

Figure 15. A range of both ground and aerial targets detected over Evaluation Trial 3, presented in {visible, thermal} imagery pairs as appropriate (UGV ground detection examples, top right/left and bottom left / UAV aerial detection examples, bottom right).

pose classification [49] and the use of 3D environment mapping via cross-spectral stereo vision [50] combined with cross-spectral Self Localization and Mapping (SLAM) [39]. Future development of the autonomous ground vehicle navigation capabilities may include both integration of terrain understanding [51], stereo vision based vehicle guidance [52, 53] and drive camera stabilization [54].

The work presented in this paper was carried out by the authors at the School of Engineering, Cranfield University as part of the SATURN (Sensing & Autonomous Tactical Urban Reconnaissance Network) project carried out by Stellar Team (2007-2009). The authors gratefully acknowledge the support of Stellar Research Services, Blue Bear Systems Research, Marshall SDG, TRW Conekt, Selex Galileo, DCMT Shivenham - Guidance and Control Group and the UK MoD in this research activity.

REFERENCES

1. A. Andreopoulos and J. K. Tsotsos, "50 Years of Object Recognition: Directions Forward," *Computer Vision and Image Understanding*, vol. 117, pp. 827–891, May 2013.
2. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *International Journal of Computer Vision*, vol. 88, pp. 303–338, Sept. 2009.
3. P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 99, pp. 1–1, 2010.
4. P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 743–761, July 2011.
5. T. Breckon, S. Barnes, M. Eichner, and K. Wahren, "Autonomous real-time vehicle detection from a medium-level UAV," in *Proc. 24th International Conference on Unmanned Air Vehicle Systems*, pp. 29.1–29.9, March 2009.
6. K. Wahren, I. Cowling, Y. Patel, P. Smith, and T. Breckon, "Development of a two-tier unmanned air system for the MoD grand challenge," in *Proc. 24th International Conference on Unmanned Air Vehicle Systems*, pp. 13.1 – 13.9, March 2009.
7. A. Gaszczak, T. Breckon, and J. Han, "Real-time people and vehicle detection from UAV imagery," in *Proc. SPIE Conference Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*, vol. 7878, 2011.
8. I. Katramados, S. Crumpler, and T. Breckon, "Real-time traversable surface detection by colour space fusion and temporal analysis," in *Proc. International Conference on Computer Vision Systems*, vol. 5815 of *Lecture Notes in Computer Science*, pp. 265–274, Springer, 2009.
9. H. K. Aghajan and A. Cavallaro, *Multi-camera networks: principles and applications*. Academic press, 2009.
10. G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *J. of Ambient Intelligence and Humanized Comp.*, vol. 2, no. 2, pp. 127–151, 2011.
11. F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 267–82, Feb. 2008.
12. O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views," *Computer Vision and Image Understanding*, vol. 109, pp. 146–162, Feb. 2008.
13. X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognition Letters*, 2012.
14. M. Taj and A. Cavallaro, "Distributed and Decentralized Multicamera Tracking," *IEEE Signal Processing Magazine*, vol. 28, pp. 46–58, May 2011.
15. M. Teutsch, W. Krüger, and N. Heinze, "Detection and classification of moving objects from UAVs with optical sensors," in *SPIE Defense, Security, and Sensing* (I. Kadar, ed.), pp. 80501J–80501J–14, International Society for Optics and Photonics, May 2011.
16. U. Zengin and A. Dogan, "Cooperative target pursuit by multiple UAVs in an adversarial environment," *Robotics and Autonomous Systems*, vol. 59, pp. 1049–1059, Dec. 2011.
17. H. Flynn and S. Cameron, "Multi-modal People Detection from Aerial Video," in *Proc. of the 8th Inter. Conf. on Computer Recognition Systems*, pp. 815–824, Springer, 2013.
18. D. C. Borghys, M. Idrissa, M. Shimoni, O. Friman, M. Axelsson, M. Lundberg, and C. Perneel, "Fusion of multispectral and stereo information for unsupervised target detection in VHR airborne data," in *SPIE Defense, Security, and Sensing*, pp. 874514–874514–12, May 2013.
19. D. Calisi, A. Farinelli, L. Iocchi, and D. Nardi, "Multi-objective exploration and search for autonomous rescue robots," *Journal of Field Robotics*, vol. 24, no. 8-9, pp. 763–777, 2007.
20. H. Surmann, D. Holz, S. Blumental, T. Linder, P. Molitor, and V. Tretyakov, "Teleoperated Visual Inspection and Surveillance with Unmanned Ground and Aerial Vehicles.," *International Journal of Online Engineering*, vol. 4, no. 4, pp. 26–38, 2008.
21. M. Andriluka, P. Schnitzspan, J. Meyer, S. Kohlbrecher, K. Petersen, O. von Stryk, S. Roth, and B. Schiele, "Vision based victim detection from unmanned aerial vehicles," in *Proc. International Conference on Intelligent Robots and Systems*, pp. 1740–1747, IEEE, Oct. 2010.

22. V. Reilly, B. Solmaz, and M. Shah, "Shadow Casting Out Of Plane (SCOOP) Candidates for Human and Vehicle Detection in Aerial Imagery," *International Journal of Computer Vision*, vol. 101, pp. 350–366, Oct. 2012.
23. G. Bourdon, "Compact integrated sensor processor: a common sensor processing core for the HYDRA unattended ground sensor system," in *SPIE Europe Security and Defence*, pp. 71120M–71120M–10, Oct. 2008.
24. W. Y. Zou and Y. Wu, "COFDM: An overview," *IEEE Transactions on Broadcasting*, vol. 41, no. 1, pp. 1–8, 1995.
25. J. Burlet and M. Dalla Fontana, "Robust and efficient multi-object detection and tracking for vehicle perception systems using radar and camera sensor fusion," in *Proc. Conference on Road Transport Information and Control*, pp. 24–24, Institution of Engineering and Technology, 2012.
26. A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36–58, 2001.
27. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. Int. Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I–511–I–518, 2001.
28. P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int. Journal of Computer Vision*, vol. 57, pp. 137–154, May 2004.
29. R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proceedings. International Conference on Image Processing*, vol. 1, pp. I–900–I–903, IEEE, 2002.
30. N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, pp. 886–893.
31. B. Besbes, A. Rogozan, and A. Bensrhair, "Pedestrian recognition based on hierarchical codebook of SURF features in visible and infrared images," in *Proc. Intelligent Vehicles Symp.*, pp. 156–161, IEEE, June 2010.
32. Q. Zhu, M. Yeh, K. Cheng, and S. Avidan, "Fast Human Detection Using a Cascade of Histograms of Oriented Gradients," in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, pp. 1491–1498, IEEE, 2006.
33. S. Vural, Y. Mae, H. Uvet, and T. Arai, "Multi-view fast object detection by using extended haar filters in uncontrolled environments," *Pattern Recognition Letters*, vol. 33, pp. 126–133, Jan. 2012.
34. C. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
35. C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Wiley-Blackwell, 2010. ISBN-13: 978-0470844731.
36. J. Han, T. Breckon, D. Randell, and G. Landini, "The application of support vector machine classification to detect cell nuclei for automated microscopy," *Machine Vision and Applications*, vol. 23, no. 1, pp. 15–24, 2012.
37. J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 679–698, Nov. 1986.
38. Y.-S. Lee, H.-S. Koo, and C.-S. Jeong, "A straight line detection using principal component analysis," *Pattern Recognition Letters*, vol. 27, pp. 1744–1754, Oct. 2006.
39. M. Magnabosco and T. Breckon, "Cross-spectral visual Simultaneous Localization And Mapping (SLAM) with sensor handover," *Robotics and Autonomous Systems*, vol. 63, pp. 195–208, February 2013.
40. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd ed., 2004.
41. R. Craig, J. Mindell, and V. Hirani, "Health survey for England," *Obesity and Other Risk Factors in Children. The Information Centre*, vol. 2, 2006.
42. K. A. Stonex, "Review of Vehicle Dimensions and Performance Characteristics," *Highway Research Board Proceedings*, vol. 39, 1960.
43. J. Happian-Smith, *An introduction to modern vehicle design*. Elsevier, 2001.
44. B. Lau, C. Sprunk, and W. Burgard, "Efficient grid-based spatial representations for robot navigation in dynamic environments," *Robotics and Autonomous Systems*, p. to appear, Aug. 2012.
45. F. Poiesi, R. Mazzon, and A. Cavallaro, "Multi-target tracking on confidence maps: An application to people tracking," *Computer Vision and Image Understanding*, vol. 117, pp. 1257–1272, Oct. 2013.
46. S. B. Lazarus, A. Tsourdos, P. Silson, B. White, and R. Zibikowski, "Unmanned aerial vehicle navigation and mapping," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 222, pp. 531–548, June 2008.
47. I. Katramados and T. Breckon, "Real-time visual saliency by division of gaussians," in *Proc. International Conference on Image Processing*, pp. 1741–1744, IEEE, September 2011.
48. J. Sokalski, T. Breckon, and I. Cowling, "Automatic salient object detection in UAV imagery," in *Proc. 25th International Conference on Unmanned Air Vehicle Systems*, pp. 11.1–11.12, April 2010.
49. J. Han, A. Gaszczak, R. Maciol, S. Barnes, and T. Breckon, "Human pose classification within the context of near-ir imagery tracking," in *Proc. SPIE Security and Defence: Optics and Photonics for Counterterrorism, Crime Fighting and Defence*, SPIE, September 2013. to appear.
50. P. Pinggera, T. Breckon, and H. Bischof, "On cross-spectral stereo matching using dense gradient features," in *Proc. British Machine Vision Conference*, pp. 526.1–526.12, September 2012.
51. I. Tang and T. Breckon, "Automatic road environment classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 476–484, June 2011.

52. F. Mroz and T. Breckon, "An empirical comparison of real-time dense stereo approaches for use in the automotive environment," *EURASIP Journal on Image and Video Processing*, vol. 2012, no. 13, pp. 1–19, 2012.
53. O. Hamilton, T. Breckon, X. Bai, and S. Kamata, "A foreground object based quantitative assessment of dense stereo approaches for use in automotive environments," in *Proc. International Conference on Image Processing*, IEEE, September 2013. to appear.
54. R. Chereau and T. Breckon, "Robust motion filtering as an enabler to video stabilization for a tele-operated mobile robot," in *Proc. SPIE Security and Defence: Electro-Optical Remote Sensing*, SPIE, September 2013. to appear.