Proceeding of SPIE : Intelligent Robots and Computer Vision XXVIII : Algorithms and Techniques, 24-25 January 2011, San Francisco, California, US. Pages 78780B-1-13

Real-time People and Vehicle Detection from UAV Imagery

Anna Gąszczak[*a], Toby P.Breckon[a], Jiwan Han[a]
[a]Cranfield University, School of Engineering, United Kingdom

ABSTRACT

A generic and robust approach for the real-time detection of people and vehicles from an Unmanned Aerial Vehicle (UAV) is an important goal within the framework of fully autonomous UAV deployment for aerial reconnaissance and surveillance. Here we present an approach for the automatic detection of vehicles based on using multiple trained cascaded Haar classifiers with secondary confirmation in thermal imagery. Additionally we present a related approach for people detection in thermal imagery based on a similar cascaded classification technique combining additional multivariate Gaussian shape matching. 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. Performance of the detector is *optimized to reduce the overall false positive rate* by aiming at the detection of each object of interest (vehicle/person) at least once in the environment (i.e. per search patter flight path) rather than every object in each image frame. Currently the detection rate for people is ~70% and cars ~80% although the *overall episodic object detection rate* for each flight pattern exceeds 90%.
Keywords: aerial vehicle detection, aerial people detection, UAV image analysis, aerial imagery, thermal, infrared images, FLIR, UAS

1. INTRODUCTION

People and vehicle detection from aerial platforms has become an important aspect of deployment for autonomous Unmanned Aerial Vehicles (UAV) systems in rescue or surveillance missions. This type of system has increasing popularity in civilian and military applications due to its ability to operate over large, difficult terrain. Search and rescue missions in remote areas where access is hampered by mountains, vast land areas without road networks or areas afflicted by natural disaster (earthquake, flood) may be aided and improved by autonomous UAV systems. In such situations, autonomous location reporting for detected objects of interest (e.g. vehicles, people) can eliminate the need for manual image analysis of the live UAV video imagery. Here we describe an approach for the real-time detection of people and vehicles from a UAV thermal/optical imagery applied to several test surveillance scenarios.

The analysis of images obtained from small aerial platform poses many challenges due to rapid platform motion, image instability and the relatively small size of the object of interest signatures within the resulting imagery. Depending on the flight altitude and camera orientation, appearance of the objects within the observed environment changes dramatically which makes automatic object detection a challenging task. Whilst using full motion video can provide spatio-temporal premise which can improve performance of the detection algorithm, the task becomes more difficult when high update temporal information is unavailable (Figure 1). In the proposed system images are sampled at 1Hz (1 fps) in order to maximize available image quality for detection within the limitations of available transmission bandwidth (e.g. within a complex urban environment) and therefore we focus our approach on robust object detection within each individual image frame. As you can see in Figure 1 temporarily consecutive image pairs present drastically different views of the same scene.

Relatively larger objects such as vehicles can be detected in optical images whilst locating people presents a relatively more difficult problem due to the small target size, low contrast of the target with the background or presence of clutter (Figure 1, 8-11). To overcome this limitation, and that of potential optical camouflage, thermal imagery is employed since the human thermal signature is somewhat more difficult to camouflage within the infrared spectrum. Even within thermal imagery the detection of human trace is still a challenging problem because the variability of such thermal signatures due to meteorological conditions, environmental thermography and locale.

[*] anna.gaszczak@cranfield.ac.uk

(a)

(b)

Figure 1. Three consecutive optical (a) and thermal (b) image pairs from our experimental UAV platform (1Hz).

The key challenge is the robust detection of static/dynamic people and vehicles in various environments, in both cluttered and uncluttered aerial views, independent of atmospheric conditions or locale. Here we propose the use of a novel adaptation to the established real-time detection approach of [1] for the detection of people in thermal imagery with secondary shape analysis of the thermal signatures. In addition we extend the previous work on vehicle detection in optical imagery [2] by using thermal traces of the vehicles in corresponding thermal imagery for secondary confirmation.

In addition to the raw detection of people and vehicles the intention is also to return a confidence indicator with each detection instance in order to facilitate the prioritization of detected objects to the UAV operator for review. This information may be also supplied as an input to a secondary multi-platform control system facilitating the deployment of other autonomous vehicles for further investigation. Aspects of the work that we will describe incorporate the fusion of multiple classifiers, multiple classification approaches and multiple sensor imagery input in order to facilitate the derivation of a quantifiable confidence indictor associated with each people or vehicle detection instance.

2. RELATED WORK

Most of recent work on object detection from UAV utilizes full motion video either for tracking and detection of moving objects [3] [4] [5] or to enhance the detection performance of stationary objects by the temporal information obtained from video feed [6]. In [4] the detection algorithm is primarily driven by the isolation of a moving component and then classifying the type of movement to assign object to particular class (i.e. human/vehicle). The approach presented in [6] detects stationary and moving humans in thermal imagery although the detection is only valid if a person is detected in a number of consecutive frames. By contrast, the vehicle detection approach presented in [7] does not use temporal information from video but uses high resolution orthonormal images and an on-line boosted classifier. The orthonormal images in general reduce the viewpoint variance of the vehicles and simplify the appearance of vehicles to rectangular shapes. Our task however is significantly different in terms of varying vehicle perspective views (e.g. see Figure 4b, 8b).

The related approach of [8] uses infrared imagery to detect stationary and moving vehicles by extracting local blob-like descriptors of vehicles and then exploiting the repetitive pattern of multiple vehicles on the roads-ways (usually forming

a linear pattern) to aid in successful detection. The model in [8] assumes stable thermal signatures of the cars which is not valid in general. Vehicle thermal traces vary based on environment (sunlight to vehicle heat transfer) and recent utilization characteristics (hot or cold engine) (Figure 4a). In our previous work [2] optical imagery was the primarily source used for the vehicle detection. Here we present an extended approach which in addition utilizes the thermal information as a secondary confirmation and confidence indicator discriminating imagery source.

Person detection from aerial imagery is a challenging problem mainly due to the small size of a human body observed from an aerial platform. In general person detection and particularly pedestrian detection from ground platforms has received considerable attention within literature and proven to be very successful [9]. These approaches generally make use of large object size within an image together with texture and shape information. By contrast, from aerial imagery people appear as small objects comprising few pixels only and lack number of features normally present when dealing with ground based surveillance scenarios and pedestrian detection. However, the recent work of Oreifej [10] presents use of Histogram of Oriented Gradients (HOG) [9] for people detection and identification from airborne optical images. Although reported results are promising, the scenarios are presented in simple and uncluttered environments which limit the general application of this technique. Suffering from similar limitations is a detection approach in [11] based on the geometric constraints of the orientation of shadow cast by a person in the scene with respect to the metadata (global position, time) and the relationship between average person height and the size of its corresponding shadow. The shortcomings of the approaches using optical imagery may be overcome by use of thermal imagery where the human body has a more stable signature. In [6] Haar features with an AdaBoost classifier are employed similarly to our approach although the use of 25 fps video allows the collecting of classified person statistics and the pruning of false positives (minimum observations threshold). The detection approach presented in [5] is also based on assumption that potential detection region has to be present over a number of consecutive video frames. First the track of a potential human is build by extracting points of interest and matching them in corresponding frames, from which the potential person signature associated with this track is classified using standard template matching.

We propose a very different approach to both people and vehicle detection due to the constraints of a single frame images (1Hz) and the associated lack of temporal information and metadata. Our proposed people detection approach uses thermal imagery and performs independent classification on every image frame. It uses a combined set of cascaded Haar classifiers [1] with additional multivariate Gaussian shape matching for secondary confirmation. By contrast, our proposed vehicle detection relies primarily on optical imagery and is again based on using multiple cascaded Haar classifiers trained for different vehicle orientations which are then combined to produce a coherent set of detections [4]. Thermal vehicle signatures are then used for secondary confirmation if detectable within corresponding thermal imagery.

3. DETECTION APPROACH

3.1 Imagery characteristics

The system used for capture comprises of a small fixed wing UAV platform with optical and thermal cameras mounted onboard. The UAV operates at an altitude of around 60m with camera pointing to the ground plane with a 45 degree angle which gives an approximate distance of 180 m to the target in a straight line. Images are sampled at 1Hz (1 fps) in order to maximize available image quality for detection within the limitations of available transmission bandwidth. These images are transmitted to a ground station via radio link where automated people and vehicle detection is performed in real-time on a standard Intel 4-core CPU workstation (2.33 GHz).

3.2 Overview

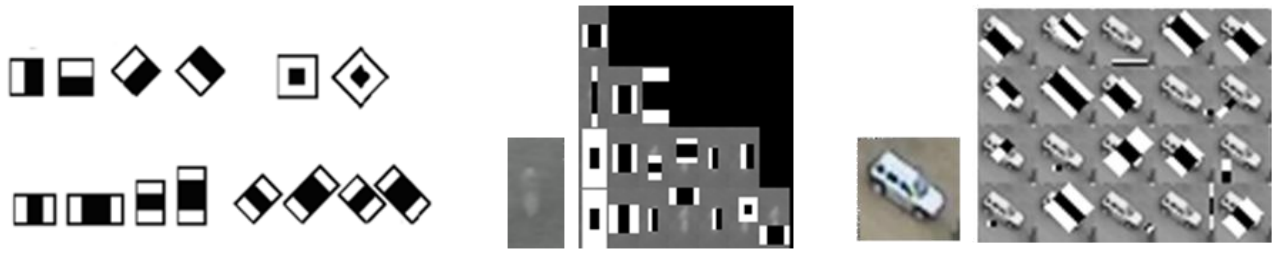
A pair of temporarily and geometrically synchronized optical and thermal images is used for the people and vehicle detection routines. First these images are processed independently and then related information from both image sources is combined. For vehicle detection the optical image is processed with several independent cascaded Haar classifiers trained for detecting vehicles in different orientations and light/dark colours. Next the overlapping detections from each of these independent classifiers is merged to produce a coherent set of detections. Once the vehicle is detected within optical image, we search the geometrically corresponding area in the thermal image for a presence of a thermal signature (varying from the environment) and if present the confidence level of this detection reported to the end user is increased.

In a similar manner, for people detection the thermal image is processed with several cascaded Haar classifiers whilst we simultaneously extract and process contours within image. Fused concurrent detections obtained by the sets of classifiers form the final result. The confidence level of this detection is further boosted if a contour of potential person matches the Gaussian shape model and is present within the corresponding detecting region. The contour shape analysis is introduced as a secondary confirmation step and is not a primarily source of detection due to the problems of cluttered environments where contour clusters occur frequently.

3.3 Cascaded Haar classifier

The primary means of people and vehicle detection makes use of multiple cascaded Haar classifiers [1] [12]. Cascaded Haar classifiers were firstly proposed by [1] and later improved by [12] with a primary application domain for the detection of faces [13]. 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) to a given type of feature at a given orientation within the image as shown in Figure 2a.

In practice these features are computed as the sum of differences between differing rectangular subregions at a localised scale which although limited in scope as individual features can be computed extremely efficiently. 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 [13]. A cascaded Haar classifier is trained using AdaBoost over a set of a few hundred positive object and negative object training images. The use of boosting techniques facilitates classifier training to select a maximally discriminant subset of these Haar-like features, from the exhaustive and overcomplete set, to act as a multi-stage cascade [1]. 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 search window approach [1]. 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 windows with a minimal evaluative (and hence computational) requirement. In this way the Haar cascade classifier thus combines successively more complex 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 the detector, provides an underlying robustness to changes in scale and maintains achievable real-time performance [1] [13].



(a) Haar-like features (b) People detection (c) Vehicle detection

Figure 2. Example of Haar-like features (a). Illustration of Haar features identified as stages within cascaded trained classifiers overlaid on exemplar images: (b) thermal people classifier, (c) 45° vehicle classifier.

3.4 People detection

In our approach to people detection a set of several separate cascaded Haar classifiers are trained on samples of human thermal signatures categorized with respect to varying orientation and thermal contrast (with the environment). The training sets were constructed in a top-to-bottom approach with regard to signature orientation with the largest set

containing large variation in human body signature orientation (with large perceived angular offsets from the vertical) to the smallest set containing body signatures largely in an upright vertical position. An example of several of the Haar features identified as stages within the trained cascaded classifier are overlaid on an example human body thermal signature in Figure 2b. As we can see from Figure 2b the identified features for person detection largely relate to differences in the horizontal and vertical directions (as can be expected).

The results obtained from the evaluation of this set of classifiers are then post-processed based on the size constraints of the objects appearing in the image. Each detected candidate subregion (rectangular) is tested against maximum width and height and rejected if it does not fulfill the requirements. These restrictions are based on *a priori* knowledge of approximate UAV altitude and camera field of view. The remaining regions are merged based on their coincidence and minimum overlapping area. This approach is based upon assumption that an object may exist at any of these locations if one or more of the classifiers were triggered at that position. In this way we develop a fuzzy logic style approach to detection - an object may belong to the detection set returned by one or more of the set of classifiers due to subjective nature of the set boundaries with respect to orientation and contrast. Empirically this approach is found to work well.

Secondary detection confirmation is achieved by a multivariate Gaussian shape matching technique. In order to first extract the shape contours we use mean-shift color clustering [14] (Figure 3a) followed by edge detection with Canny edge detector [15] (Figure 3b). We use mean-shift color clustering as a preprocessing stage for determining the contours due to its ability to reduce variation of pixel values and to preserve small entities in the image whilst at the same time merging larger, uniform image regions. For illustration purpose extracted contours are over drawn in color in Figure 3c.

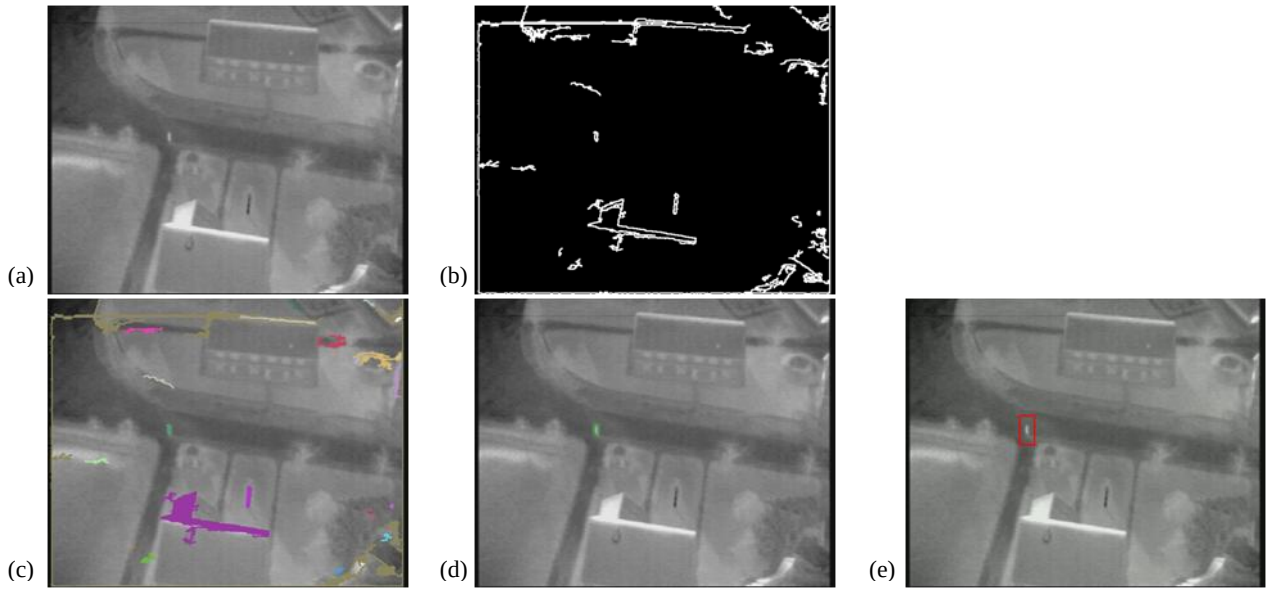


Figure 3. Contour based human body detection processing pipeline (a→e). (a) Mean-shift segmented thermal image, (b) edge image, (c) extracted shape contours, (d) highlighted contour matching the model, (e) Haar classifier detection

In each of the images we consider all of the extracted contours and use Hu invariant moments as shape descriptors [16]. These features are then compared against a statistical model shape descriptors derived from a set of sample human thermal signature contours. By treating contour moments as random variables we can model the distribution of these features in the training set $\mathbf{X}$ as a multivariate Gaussian distribution [17]. This distribution can be constructed by calculating mean feature vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$ as follows (Eq. 1):

$$\boldsymbol{\Sigma} = E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T] \qquad (1)$$

where $\mathbf{X}$ is a matrix of the training set feature vectors and $E[]$ is the expectation operator.

In order to determine if a contour belongs to the distribution we calculate distance of a given shape descriptor $\mathbf{x}$ from the statistical model embodied in mean feature vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. In this case we use Mahalanobis distance which accounts for the covariance in the data so that the similarity measure is not biased in any variable dimension. The Mahalanobis distance is defined as follows:

$$D(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})} \qquad (2)$$

where $\boldsymbol{\Sigma}^{-1}$ is the inverted covariance matrix and $\mathbf{x}$ is a shape feature vector.

If the calculated Mahalanobis distance is below a certain threshold and the contour interior is brighter than the local image background the shape is qualified as matching the predefined model and the confidence level indication of this detection is adjusted. Contours matching the trained moments model are highlighted in Figure 3d with the green bounding box. This information is used to improve the confidence level of the initial detections obtained by cascaded Haar classifier (Figure 3e).

3.5 Car detection

Analogous to the person detection approach the primary detector used for vehicle detection is also the cascaded Haar classifier [1]. The detection approach relies primarily on optical imagery as vehicle thermal signatures are unstable under varying conditions (Figure 4a). As we can see in Figure 4a example vehicles taken under a range of environmental conditions and locale have varying thermal signatures.

For our application to aerial vehicle detection four separate cascaded Haar classifiers where trained based on sample vehicle images categorized into one of four positional orientations to the horizontal - the resulting set of clockwise angular offsets {0°, 45°, 90°, 135°}. A large training set was empirically sub-divided based on the perceived nearest angular offset of the vehicle wheelbase with vehicles assumed to be symmetric about the axis perpendicular to the angle of the wheelbase. As we can see from Figure 2c the detection of vehicles relies largely on feature differences within the 45° orientation of the vehicle training set in use for this particular classifier (as expected).

As in case of people detection, the set of regions detected by Haar classifier is verified based on the size constraints. The spatially coincident detections are merged with respect to their intersection area and a similar fuzzy style approach is used – a detected vehicle can belong to the detection set returned by one or more orientation specific classifiers. The confidence level of the vehicle detection is increased proportionally to the number of classifiers triggered. For example, if a particular vehicle is detected by the 45° and 90° orientation specific classifiers then it is more probable that it is true positive classification and such the associated confidence level of the detection is increased. In addition the confidence level of the detection may also be increased by incorporating cues of vehicle presence within the corresponding thermal imagery.

(a) (b)

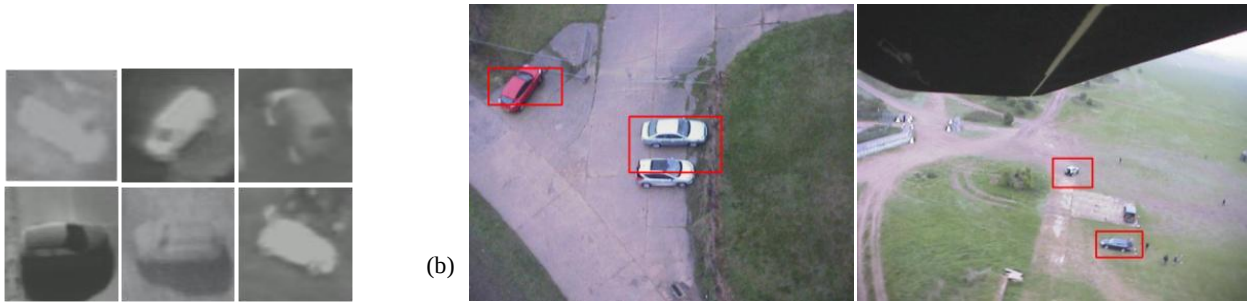


Figure 4. (a) Thermal imagery - signatures of cars; (b) Example detections of vehicles in optical imagery.

3.5.1 Thermal signature confirmation

In some cases vehicles can exhibit a strong thermal signature due to the sunlight to vehicle heat transfer or engine/interior temperature. We can incorporate this information to increase the confidence level of the overall vehicle detection. This is achieved by generating a list of thermal hot spots present in the thermal imagery that meet the requirements of the maximum pixel value and contrast. Maximum pixel intensity is set to be above a certain threshold based of the percentage of the maximum pixel value identified within the image and the spread of the pixel value distribution. A set of identified high intensity pixels is used as seed locations for watershed clustering [18].

From a set of identified maximum value pixels hot spot regions are grown – each connected component of high value pixels is used as a seed point for a watershed clustering algorithm [18]. Pixels with similar values in a fixed range from a seed point are aggregated until a gradient of image intensity exceeds a threshold. This process provides a set of contours (Figure 5a) which are then post-processed with a rule based constraints with respect to the foreground to background contrast and the intensity level of foreground (Figure 5b, 5c). Figure 5b shows a binary mask of the thermal signatures that meet the contrast and maximum pixel value constraints. Bounding boxes of the final set of verified thermal signature are presented in Figure 5c.

Figure 5. Thermal imagery whith identified high intensity and contrast thermal signatures.

The combined use of these detected thermal signatures and the earlier optical imagery detections (Section 3.5) together as a fused sensing capability requires thermal and optical images to be temporally and spatially synchronized. The latter is performed using an image based homography.

3.5.2 Homography estimation

Detected thermal signatures (Section 3.5.1) are assigned to the corresponding optical vehicle detections by calculating the projection of the thermal image plane onto the optical image plane. This is carried out by using a two dimensional homography based technique which is a projective transformation that maps one image plane to the other. Here we consider the problem of determining the homography that maps points in the thermal image to the corresponding points in the optical image. Equation 3 illustrates the relationship between points in the two corresponding images.

$$p_{optical} = Hp_{thermal} \qquad (3)$$

where H is a 3x3 homography matrix and point p in the image denoted as vector p = [x, y, 1]'

Figure 6. Thermal (a) and optical (b) images of the same scene fused using homography (c).

We calculate homography matrix from pairs of corresponding feature points visible in both images (e.g. corners of the runway in Figure 6a and Figure 6b). In Figure 6c we see the example of thermal image overlaid onto the optical image according to the homography matrix calculated from the pairs of feature point correspondences between Figure 6a and Figure 6b.

(a) (b) (c)

Figure 7. (a) Identified thermal signatures; (b) Vehicle detections; (c) Fused detections from thermal and optical images.

Figure 7 shows that the perspective transform using the homography obtained from the two reference images is not in general fully accurate on another set of frames. This is attributable to the practical aspects of both variations in platform position and temporal image synchronization. Additionally we do not account for the correction of the intrinsic camera parameters as this is left as an area for future work.

Essentially the homography calibration gives us an approximate thermal to optical mapping which we can then successfully use for fusing the optical and thermal detection information. To take into account of these errors, instead of considering only the area inside the optical detection window we consider also the general proximity of this window within the projected thermal image (dashed line rectangles in Figure 7c). This is done using the extended bounding rectangle of the detection into the thermal image. If a thermal signature is present in the extended area the confidence level of the detection is increased. We can see in Figure 7c identified thermal signatures from the thermal imagery (Figure 7a) fused with the Haar classifier detections from the optical imagery (Figure 7b).

4. RESULTS

The UAV platform used for this work was a fixed wing platform with a 1m wingspan and operating weight of 1.8kg. This platform operates at an altitude of approximately 60m based on either hand or catapult launched and can carry a dual optical/thermal camera payload. Control is via an autonomous waypoint following autopilot. For this work the platform was equipped with a standard CCD optical camera and uncooled thermal sensor both positioned at 45° angle to the horizontal in the direction of flight.

Our approach facilitates the real-time detection of both static and moving vehicles invariant to orientation, colour, type and configuration (Figure 4, Figure 8). The results presented also show the successful detection of people under varying conditions (e.g. atmospheric conditions, UV/IR exposure, background temperature) in both isolated rural and cluttered urban environments with minimal false positive detection (Figure 9, Figure 10). Performance of the detector is *optimized to reduce the overall false positive rate* by aiming at the detection of each object of interest (vehicle/person) at least once in the environment (i.e. per episode/UAV search flight) rather than every object in each image frame. Currently the per-image frame detection rate for people is ~70% and vehicles ~80% although the *overall episodic object detection rate* for each flight pattern or search mission exceeds 90%. In addition there is a mechanism for grading the detections based on associated confidence level. This technique allows further filtering and reduction of false positive detections as required.

In Figure 8 we see the automated detection of people within the thermal imagery and the corresponding detection of vehicles within the optical imagery in the same environment. It should be noted in Figure 8 that this was taken under direct sunlight and hot thermal conditions. Although person thermal signatures have a low contrast, our approach still

successfully identifies people in the thermal imagery. In addition thermal signatures of the vehicles are present (Figure 8a) therefore the confidence level of the vehicle detection in the optical imagery (Figure 8b) may be increased using this additional fused sensor information.

Figure 8. People detection results in thermal imagery (a) with optical images of the same scene with vehicle detections (b) (summer conditions).

An interesting case is presented in Figure 9. Despite the fact that people are fully occluded by the trees and are barely visible within the optical image (Figure 9b) the targets are successfully detected in corresponding thermal images (Figure 9a). As you can see in the thermal imagery (Figure 9a) all of the people present are not detected within each individual image frame (Figure 9a, 9b) although are successfully detected over both images with every human target in the environment being identified. This is a good example of the per frame versus episodic detection performance of the approach. In addition a small fragment of a telegraph pole within environment is falsely identified as a human thermal signature due to its similarity in shape. Instances of such false positives are limited from episodic stand point.

Examples of detections in cluttered/urban scenes are showed in Figure 10-12. In Figure 10 the thermal signatures of several objects present are strong with high contrast and thus difficult to classify for a human operator. Several of these are successfully detected as people present within the environment with a minimal number of false positives (Figure 10a). These false detections occur as a result of an infrared reflecting from the metal containers within the environment. In Figure 10a, 10b the orientation and intensity of the thermal traces make them viable candidates for human thermal signatures and as such are correctly classified with minimal false positive impact.

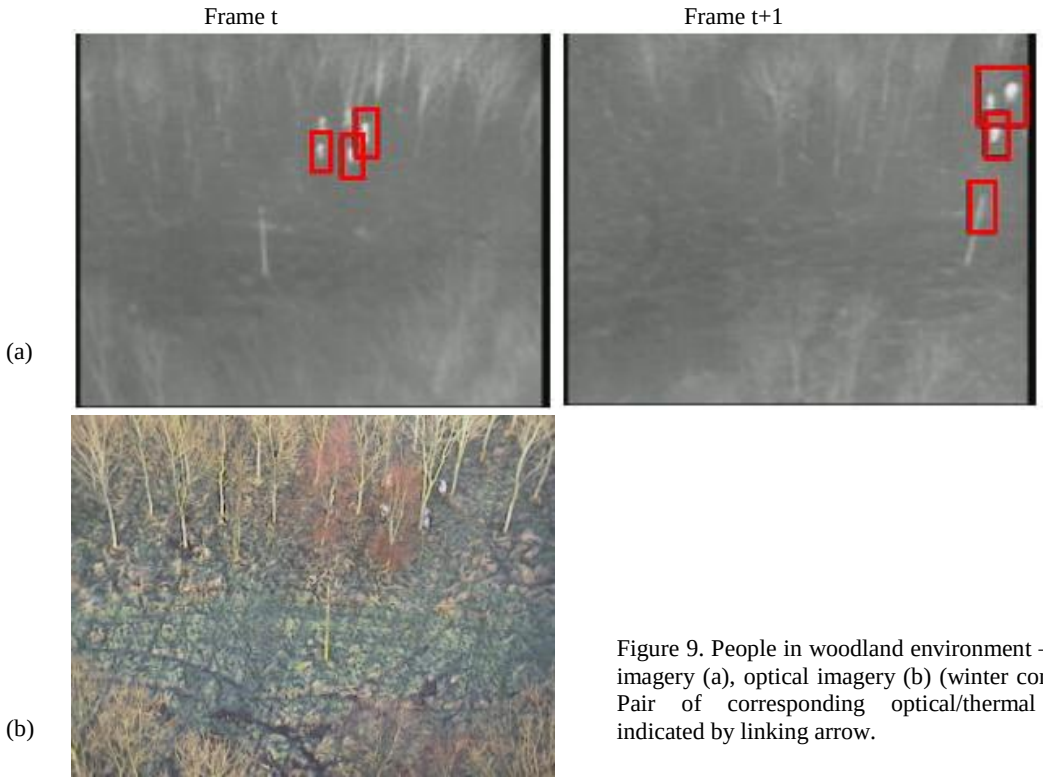


Figure 9. People in woodland environment – thermal imagery (a), optical imagery (b) (winter conditions). Pair of corresponding optical/thermal images indicated by linking arrow.

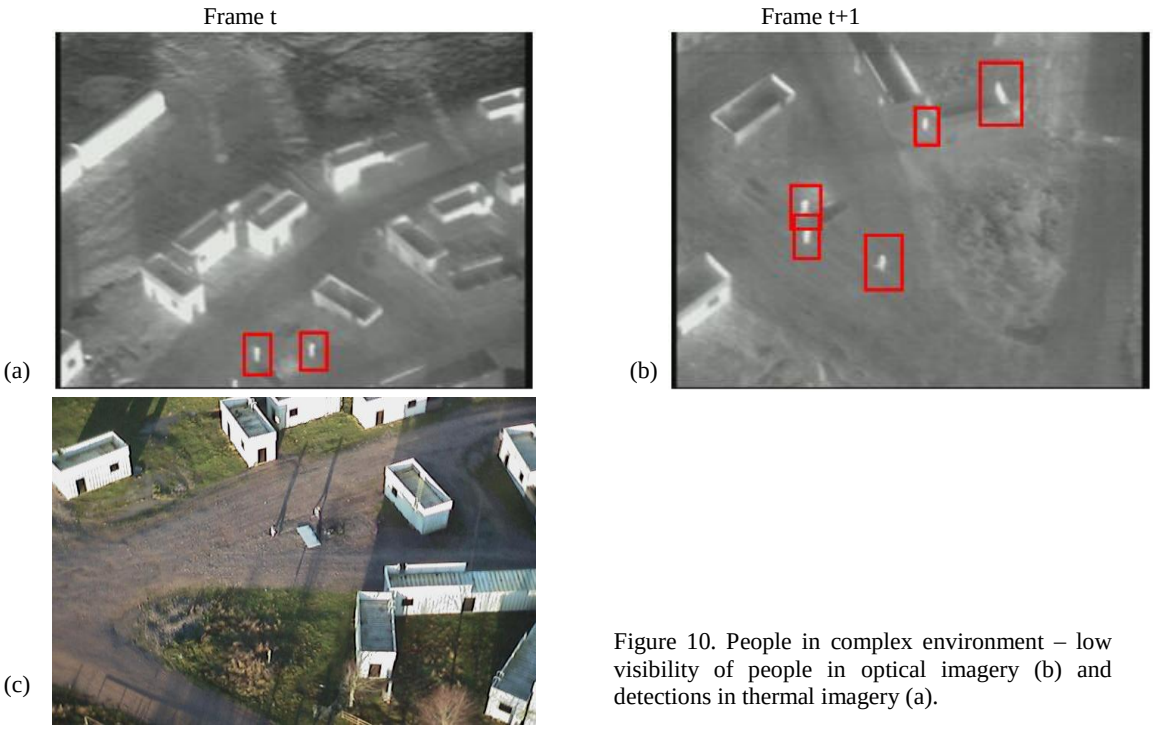


Figure 10. People in complex environment – low visibility of people in optical imagery (b) and detections in thermal imagery (a).

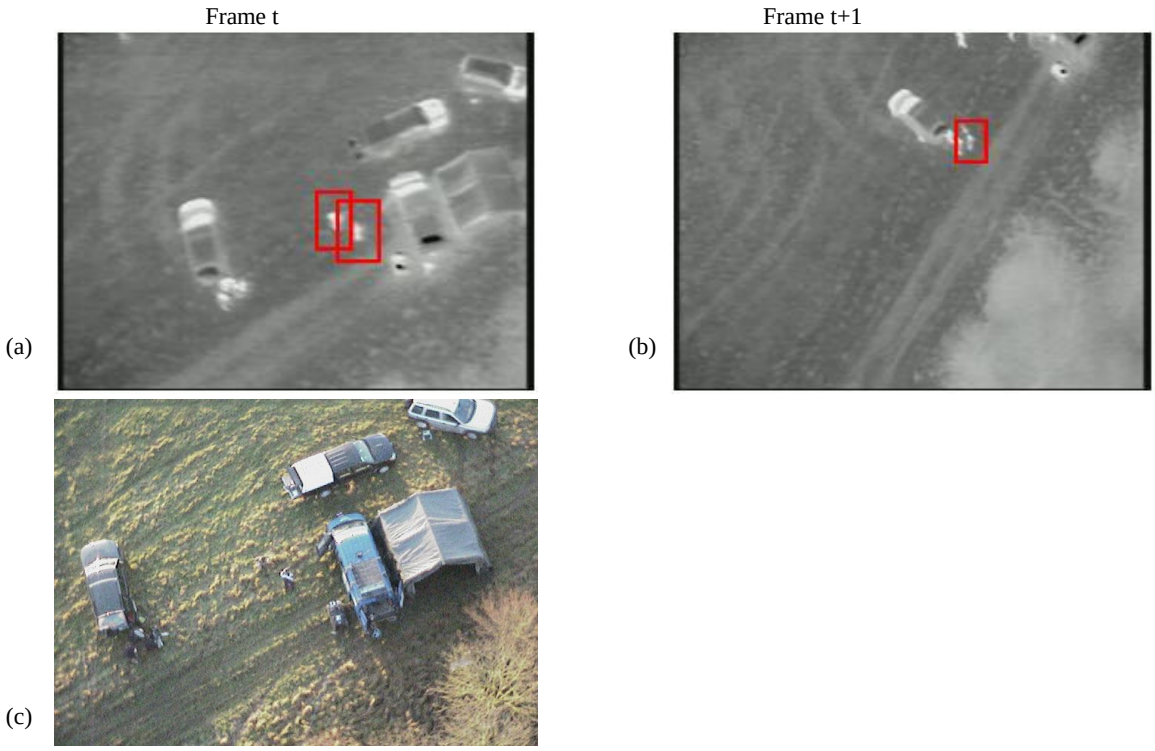


Figure 11. People in rural environment – optical imagery (a, b) and thermal imagery (c).

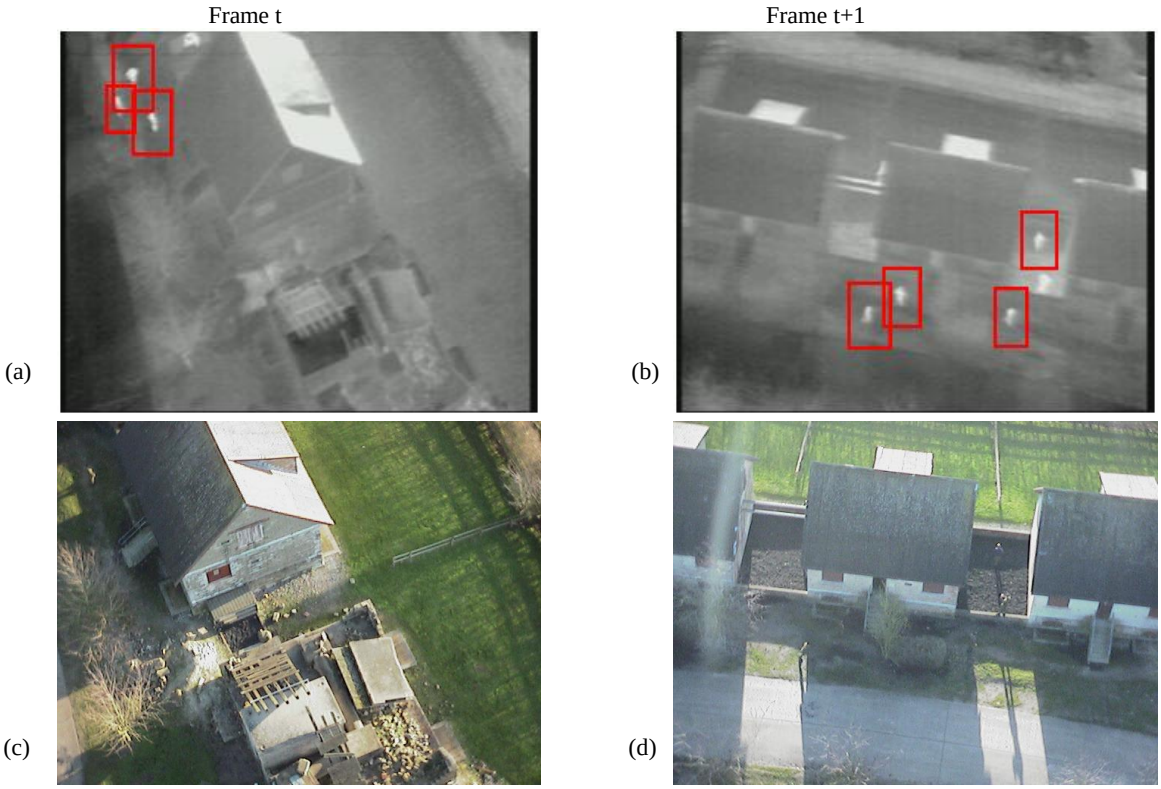


Figure 12. Urban environment - thermal imagery (a, b) with corresponding optical images of the same scene (c, d).

Detections within rural environment are presented in Figure 11. Although there are no objects obstructing the field of view, there is a group of people close to the vehicle which is not identified in the first image in Figure 11a. This occurs due to the straight downward perspective so that full vertical thermal signature of a person is not present. Nevertheless, the same group of people is correctly detected in the next consecutive image frame Figure 11b therefore accurately detecting all targets within the environment.

Figure 12 shows people in a complex urban environment. Thermal signatures are clearly visible in the thermal imagery (Figure 12a, 12b) despite low target visibility in optical imagery because of the background camouflage (Figure 12c, 12d). Notably one of the five people present in the thermal imagery is not classified as a person due to the lower contrast of the thermal signature (Figure 12a).

Figure 8-10 show the successful detection of vehicles and people from UAV aerial imagery using the proposed approaches under a variety of environmental conditions.

5. CONCLUSIONS AND FURTHER WORK

A novel methodology of people and vehicle detection from UAV imagery has been proposed. Moreover a concept of the confidence indicator associated with each detection instance in order to facilitate the prioritization of detected objects has been introduced. The methodology has been tested over a range of outdoor scenarios, in different rural and urban environments and under variety of weather conditions. The obtained results appear promising as we have been able to detect more than 90% of vehicle and people instances within these scenarios with occasional false positive detections.

Future work in this area will include the integration of the metadata obtained from the UAV platform (e.g. position) with the detection approach to enable more effective pruning of false positive detections. In addition the consideration of full motion video (within the constraints of available bandwidth) would give an advantage of incorporating temporal information for a greater robustness of detection.

6. BIBLIOGRAPHY

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 511-518 (2001)

[2] T.P. Breckon, S.E. Barnes, M.L. Eichner, and K. Wahren, "Autonomous Real-time Vehicle Detection from a Medium-Level UAV," Proc. 24th International Unmanned Air Vehicle Systems, 29.1-29.9 (2009)

[3] C. Huang, Y. Wu, J. Kao, M. Shih, and C. Chou, "A Hybrid Moving Object Detection Method for Aerial Images," Advances in Multimedia Information Processing, 357-368 (2010)

[4] J. Xiao, C. Yang, F. Han, and H. Cheng, "Vehicle and Person Tracking in Aerial Videos," Multimodal Technologies for Perception of Humans, 203-214 (2008)

[5] A. Miller, P. Babenko, M. Hu, and M. Shah, "Person Tracking in UAV Video," Multimodal Technologies for Perception of Humans, 215-220 (2008)

[6] P. Rudol and P. Doherty, "Human Body Detection and Geolocalization for UAV Search and Rescue Missions Using Color and Thermal Imagery," Proc. IEEE Aerospace Conf, 1-8 (2008)

[7] H. Grabner, T.T. Nguyen, B. Grubner, and H. Bischof,"On-line boosting-based car detection from aerial images", ISPRS Journal of Photogrammetry and Remote Sensing 63(3), 382-396 (2008)

[8] S. Hinz and U. Stilla,"Car detection in aerial thermal images by local and global evidence accumulation", Pattern Recognition Letters 27(4), 308-315 (2006)

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proc. Conf. Computer Vision and Pattern Recognition, 886-893 (2005)

[10] O. Oreifej, R. Mehran, and M. Shah, "Human identity recognition in aerial images," Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 709-716 (2010)

[11] V. Reilly and M. Solmaz, "Geometric Constraints for Human Detection in Aerial Imagery," Proc. European

Conference on Computer Vision, 252-265 (2010)

[12] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," Proc. International Conf. Image Processing (2002)

[13] P. Viola and M. Jones,"Robust Real-Time Face Detection", International Journal Computer Vision 57(2), 137-154 (2004)

[14] D. Comaniciu and P. Meer, "Mean Shift Analysis and Applications," Proc. International Conference on Computer Vision, p1197 (1999)

[15] J. Canny,"A computational approach to edge detection", IEEE Trans. Pattern Anal. Mach. Intell. 8(6), 679-698 (1986)

[16] M.K. Hu,"Visual Pattern Recognition by Moment Invariants", IRE Transactions on Information Theory 8, 179-187 (1962)

[17] C. M. Bishop, [Pattern Recognition and Machine Learning (Information Science and Statistics)], Springer, Berlin, 78-85 (2007).

[18] F. Meyer, "Color image segmentation," Proc. Int Image Processing and its Applications Conf., 303-306 (1992)

https://dspace.org/

Real-time people and vehicle detection from UAV imagery

Gaszczak, Anna; Breckon, Toby P.; Han, Jiwan

Röning, J.; Casasent, D. P.; Hall, E. L.

2011-01-24T00:00:00Z

http://dx.doi.org/10.1117/12.876663

Anna Gszczak, Toby P.Breckon and Jiwan Han. Real-time people and vehicle detection from UAV imagery. Proceeding of SPIE : Intelligent Robots and Computer Vision XXVIII : Algorithms and Techniques, 24-25 January 2011, San Francisco, California, US. Pages 78780B-1-13

Downloaded from a DSpace repository. https://dspace.org/