# REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188



| 1. REPORT DATE *(DD-MM-YYYY)* | 2. REPORT TYPE | 3. DATES COVERED *(From – To)* |
|---|---|---|
| 08-05-2007 | Final Report | 1 November 2005 - 13-Jul-07 |

**4. TITLE AND SUBTITLE**

Object Segmentation and Aerial Tracking of Vehicles

**5a. CONTRACT NUMBER**
FA8655-06-1-3023

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Dr. Noel O'Connor

**5d. PROJECT NUMBER**

**5d. TASK NUMBER**

**5e. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Dublin City University
Glasnevin
Dublin 9
Ireland

**8. PERFORMING ORGANIZATION REPORT NUMBER**

N/A

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

EOARD
PSC 821 BOX 14
FPO AE 09421-0014

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
Grant 06-3023

**12. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution is unlimited.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

This report details our system for vehicle detection and tracking in multi-spectral UAV video data. The work builds upon the tracking framework previously developed, using the spatiogram bank tracker as a core component of the overall system. Camera motion is compensated by automatically performing image alignment and computing the homographic warping that describes the movement of the ground-plane in relation to the camera. By aligning subsequent frames, moving vehicles can be detected using image differencing, adaptive thresholding and selection of appropriately sized regions. Banks of spatiogram trackers are then used to perform object tracking and to fuse data from both the visible spectrum and thermal infrared video. Results are shown indicating the success of this approach on UAV data from the DTO VACE dataset collection. Finally, we note the challenges posed by this difficult application and propose future directions for this work in tackling these issues.

**15. SUBJECT TERMS**
video segmentation, tracking, spatiogram, image processing, EOARD

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18, NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | BARRETT A. FLAKE |
| UNCLAS | UNCLAS | UNCLAS | UL | 10 | **19b. TELEPHONE NUMBER** *(Include area code)* +44 (0)20 7514 4285 |

# Vehicle Tracking in UAV Video using Multi-Spectral Spatiogram Models

Final Technical Report for EOARD

N. E. O'Connor, A. F. Smeaton, C. Ó Conaire, P. Kehoe,

April 27, 2007

## 1. Introduction

This report follows directly from our two previous interim reports that detailed our experiments that developed and evaluated fusion frameworks for multi-feature object tracking. For context, this report first briefly summarises the two previous interim reports. However, the main focus of this report is a technical description of our vehicular tracking system and its performance on Unmanned Aerial Vehicle (UAV) video data. Specifically, we target the simultaneous use of both visible spectrum and thermal infrared video for tracking.

### 1.1 Summary of Previous Work

In our first interim report of March 22, 2006, [1] a number of fusion schemes for combining information from standard CCTV and thermal infrared spectrum video were explored in the context of tracking surveillance objects. We experimented on numerous real world multimodal surveillance sequences in challenging environments. The result of evaluating the performance of these fusion schemes, led to the development of our fusion framework for multi-spectral object tracking using banks of individual spatiogram trackers.

This general-purpose fusion framework for multi-spectral object tracking was described in our second interim report of February 28, 2007 [2]. In this report, the technical details of the spatiogram bank tracker were described and the justification behind its formulation was given. We showed how this followed from the fusion scheme evaluation of the first report and how this was better suited to multi-feature tracking than the commonly-used histogram-based tracking. Tracking performance was shown to be superior to other standard tracking methods for a variety of testing sequences.

### 1.2 Overview of this report

This report details our system for vehicle detection and tracking in multi-spectral UAV video data. The work builds upon the tracking framework previously developed, using the spatiogram bank tracker as a core component of the overall system. Camera motion is compensated by automatically performing image alignment and computing the

homographic warping that describes the movement of the ground-plane in relation to the camera. By aligning subsequent frames, moving vehicles can be detected using image differencing, adaptive thresholding and selection of appropriately sized regions. Banks of spatiogram trackers are then used to perform object tracking and to fuse data from both the visible spectrum and thermal infrared video. Results are shown indicating the success of this approach on UAV data from the DTO VACE dataset collection. Finally, we note the challenges posed by this difficult application and propose future directions for this work in tackling these issues.

# 2. System Overview

The system can be viewed at a high level as a single block which simply takes input of frames from the visual and infrared spectrum videos, and after processing returns the locations of tracked vehicles in each frame. The system is composed of a number of subsystems, depicted in the Figure 1, along with intermediary inputs and outputs. Inputs and outputs are represented as boxes, while circles represent subsystems. The various stages of processing involved are outlined in more detail below.

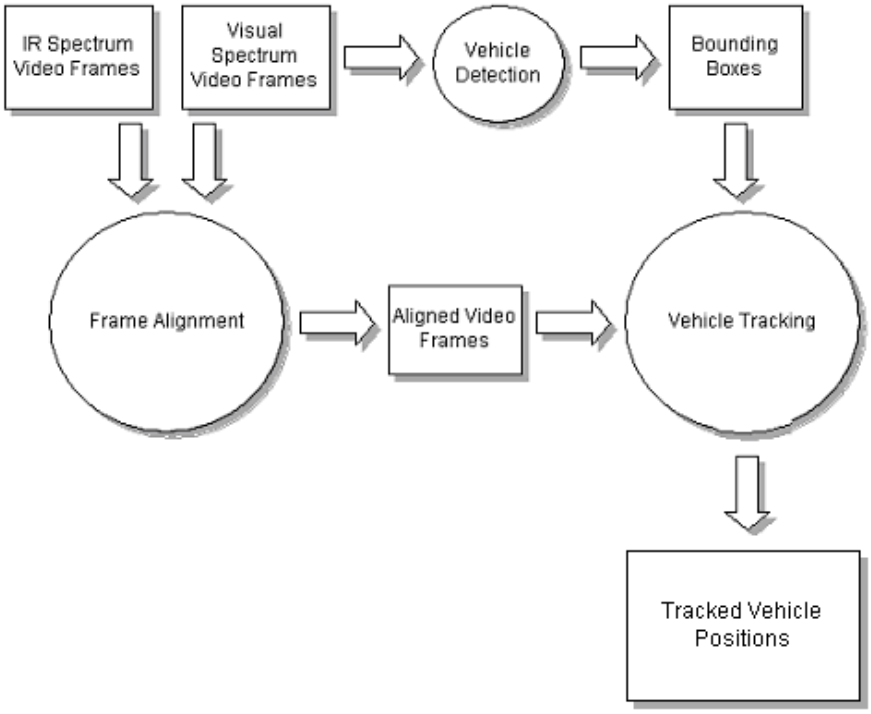



Figure 1: High Level System Overview

## 2.1 Frame Alignment

Where infrared video is available, it is often inconveniently misaligned with its visible spectrum counterpart. In order to use these two sources of data successfully in the

tracking stage, they need to be reasonably well aligned. There are two types of alignment to consider – temporal and spatial. It is assumed that both cameras capture frames at the same rate, so that, once aligned, the temporal displacement does not drift. This was found to be a reasonable assumption in the video data we used. Alignment such as this can also be implemented in hardware, as we did in our previous work using camera gen-lock [5]. The second aspect, spatial alignment, is necessary due to potential differences caused by slightly different camera locations on the UAV itself and differing fields of view as a result. As a consequence, the IR images must be aligned to the visible images in order to provide suitable input for the tracker.

Once the video frames are temporally aligned, the frames are spatially aligned by scaling the resolution of the IR frame to match the visible frame (the IR frame is typically smaller). A homography is then computed in a least-squared-error approach using manually selected corresponding points between the infrared and visible images. This homography is used to warp the IR image to the visible image. A single homography is sufficient for a single sequence, since the cameras are in a fixed position relative to each other. This is illustrated in figure 2.

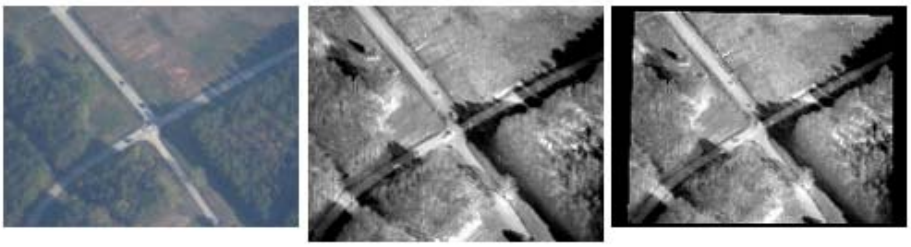



Figure 2: Example frame alignment: (a) visible frame (b) temporally aligned infrared frame, (c) spatially aligned infrared frame

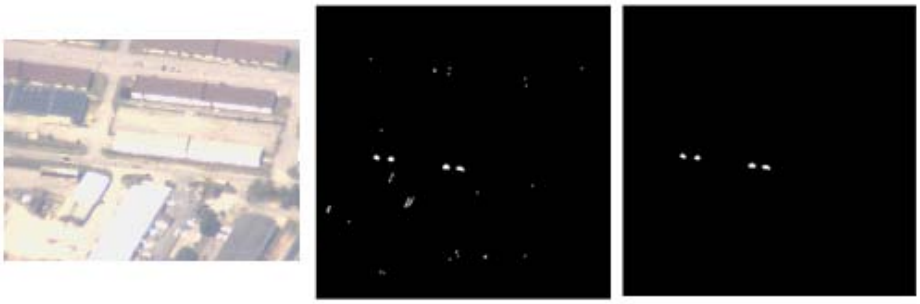
## 2.2 Vehicle Detection

For initialisation, our tracker requires external input in the form of bounding boxes encompassing the regions to be tracked (i.e. the vehicles) in the first frame of video. To achieve this, a vehicle detection phase takes as input the first frame of visible footage. The detection phase assumes the vehicles are moving, and highlights them by detecting the dominant camera motion and subtracting this to highlight those objects moving independently of the camera. It does this by computing a homography between a set of points in the first frame and the set of points *n* frames later. Typically we use a time step of 20 frames. We warp the frame using the computed homography and then subtract this from the original frame. The result is an image which highlights objects moving independently of the ground plane (i.e. vehicles) as pairs of bright ‘blobs’ – see figure 3.

(a) (b)

Figure 3: Intermediary output of detection: (a) visible frame, (b) corresponding vehicles highlighted as pairs of bright ‘blobs’

This result can sometimes be quite noisy, and thus some further processing is required before we can clearly identify relevant ‘blobs’. The image is binarised using the dynamic thresholding method of Otsu [4], and then small regions caused by noise are removed using size-based filtering – see Figure 4, for example. We calculate the spatial centres of the remaining blobs, and based on distances between those centres, pairs of blobs related to a single vehicle are identified and a bounding box is constructed.



(a) (b) (c)

Figure 4: Noise removal (a) visual spectrum image, (b) black & white highlighted ‘blobs’ with a little noise, (c) after noise removal

## 2.3 Vehicle Tracking

These initial bounding boxes, corresponding to the top left *x* and *y* co-ordinates of the box, and its *width* and *height* are used to initialise the tracker – in subsequent frames it uses the bounding boxes from the previous frame. The tracker uses the visible spectrum and thermal infrared image of the current frame to create a multi-band image using the RGB components of the visible image and the IR image as a fourth channel. Using this multi-spectral image, the object model for each vehicle is created by extracting a bank of spatiograms at the corresponding bounding box location.

In subsequent frames, tracking continues by finding the best matching candidate in a search window centred on the previous location of the vehicle. Matching is determined

using a newly proposed spatiogram similarity measure that was shown to be superior to the originally proposed measure. This new measure was described and evaluated in our recent publication [3].

# 3. Testing

In order to evaluate the system's performance we used video from the DTO VACE Unmanned Autonomous Vehicle (UAV) data corpus. There were a number of issues to be considered in using this collection. First, the amount of video that does not feature any moving vehicles at all is non-negligible, and we removed such data from further consideration. Thus we looked for data which had visible and corresponding IR video that was reasonably aligned.

Videos were fed to the system as sets of jpeg frames extracted from the videos. The tracker returned its results as a new set of frames taken from the visible videos, with vehicles highlighted. Examples of the results can be seen in the tracked sequences in figure 5. These sequences illustrate some of the benefits of the proposed tracker.

Sequence (a) illustrates tracking of multiple targets of very small size, and also showcases the tracker's handling of shadowed areas. Shadows can present a challenge by significantly altering the distribution of colour in a target area, and in the presence of particularly hard shadows, it is possible for dark vehicles to disappear almost completely. In this sequence, there are instances of shadowing or partial shadowing which the tracker recovers from – at the first corner, and from the trees along the straight stretch of road.

In sequence (b), the difficulty of challenging colour is presented. In this sequence, the video is highly saturated due to strong sunlight, causing the vehicles to appear nearly the same colour as their surroundings i.e. white. In this instance, however, the tracker can still successfully hold its targets.

In sequence (c), we are presented with a similar scenario as in (a), in terms of the presence of shadowing and vehicle size. However this sequence also features a camera that shifts quite significantly in a reasonably short period of time. Although not perfect, the tracker manages to retain focus on 2 out of 3 targets.


Figure 5: The tracker highlights vehicles on the image using a transparent red mask.

The system's flexibility in accommodating multiple features without incurring the exponentially increasing computational cost of other techniques such as histograms is one of its key strengths, and can be leveraged further for improved capability and robustness. Next we will discuss opportunities for future improvement generally.

# 4. Future Work

The system presented here presents a number of opportunities for further work. We will discuss some specific issues in various stages of the system and propose strategies that could be implemented to address them.

### 4.1 Detection

As discussed earlier, the tracker needs to be initialised with bounding boxes. These are obtained from the previous frame by the tracker, but to provide initialisation, bounding boxes need to be provided outside of the tracking subsystem. As described, these are provided by an automatic detection phase. Thus, in the current system, vehicles in the first frame can be detected and subsequently tracked, but this does not accommodate the introduction and tracking of new vehicles beyond the first frame. It would be reasonably straightforward to build upon the current detection system so that new trackers can be initialised as new vehicles are detected; however there is a challenge in balancing the detection phase such that subsequent tracking is not made redundant. Detection of vehicles on a per-frame basis would be tantamount to per-frame tracking of vehicles and tracking in such a manner using our current detection system would be much more computationally expensive than the spatiogram trackers.

We can balance these concerns by perhaps exploiting temporal and spatial coherency in the detection phase by only running detection every *n* frames, and performing per-frame tracking in between using the spatiogram tracker. Using such an approach there may be a small lag between when a vehicle appears and when it begins to be tracked, but human perception would gauge that lag to be small.

Furthermore, the detection phase could be improved to be more reliable. This phase can be sensitive to noise, making reliable detection difficult, but this could be offset by checking multiple frames and seeing where detection results between those frames agree.

### 4.2 Bounding Boxes

The success of the tracker can be sensitive to the fit of a bounding box around a target. Boxes that include the target but are of varying size or location – even if that variation is small – can give significantly different results. Tuning the detection phase to consistently produce a good fit should result in improved results.

### 4.3 Occlusion

The tracker can lose sight of vehicles in the presence of occlusion, typically from trees. The tracker already attempts to compensate via prediction, attempting to guess the location of the vehicle until it emerges from the occluded area. Long periods of occlusion can be problematic, but with short periods successful recovery is often possible as in figure 6, sequence (a).

### 4.4 Appearance Change

In addition to opaque occlusion, the appearance of the vehicles in the image can change due to a number of factors, which can lead to challenges for our tracker. The tracker

currently is sensitive to changing orientation of targets within bounding boxes, for example. Spatiograms preserve some detail of the spatial distribution of colour within the target area, and this distribution changes when the orientation of a vehicle or other detail within the target area changes. This occurs quite typically when, for example, the vehicle turns a corner and the camera's orientation does not similarly change. Another factor which can influence appearance is the existence of environmental shadows. Vehicles passing under dark shadows can be lost or confused with other vehicles. For example, in the sequence presented in figure 6 (b), the tracker loses a target as it enters a shadowed area. As with occlusion, it attempts to predict its location while it waits for the target to re-emerge, but in this instance the tracker confuses another vehicle for the shadowed vehicle and starts tracking it instead. Interestingly, shadows appear very strongly in the infrared video due to the temperature differences between areas that are shaded from direct sunlight and those that are not.

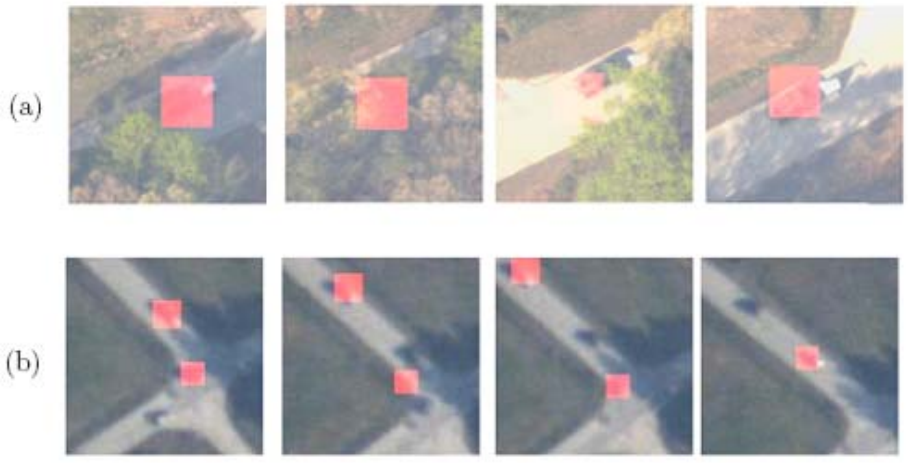



Figure 6: Instances of occlusion: (a) successful recovery from tree occlusion, (b) unsuccessful encounter with environmental shadows, with subsequent confusion of one vehicle for another

For dealing with these last issues of occlusion and appearance change, as well as other challenges like large and fast shifts in camera motion, a general strategy for recovery from 'lost' targets is desirable. It would be possible to implement a system whereby the tracker could detect target loss or tracking failure. It could do so by checking the best similarity score for a target in each frame. If the best score has fallen below a certain threshold, the tracker could assume it has lost the target, and defer to the detection phase in order to relocate it. Another option would be to keep track of the vehicles direction of motion and velocity, and interpolate the bounding boxes using that information until the similarity score recovers. One or both of these approaches in unison could help significantly improve the tracker's robustness and ability to deal with the issues mentioned above.

Leveraging the framework's flexibility in adding new features, it would also be interesting to investigate the addition of edge-based features alongside the visible and

infrared spectrum data. Such an extension could help reduce the tracker's sensitivity to lighting and contrast changes in the video. More generally, though performance was not a guiding concern in this work, optimisation of the processing involved would be desirable in order to improve performance.

# 5. Conclusion

This report describes our work in automatic vehicle tracking from UAV video data using visible spectrum and thermal infrared data. We first summarised our previous work on multi-modal tracking and justified the use of this fusion framework for our vehicle tracking system. Our system for vehicular tracking was described in detail and illustrative results of the detection and tracking performance were given. We concluded by providing many potential directions for future research.

# References

[1] N. E. O'Connor, A. F. Smeaton, and C. Ó Conaire, 'Comparison of Fusion Methods for Visible Spectrum and Infrared Image Data', March 22 2006.

[2] N. E. O'Connor, A. F. Smeaton and C. Ó Conaire, 'A General Purpose Fusion Framework for Multi-spectral Object Tracking', February 28, 2007.

[3] Ciarán Ó Conaire, Noel E. O'Connor and Alan F. Smeaton, 'An Improved Spatiogram Similarity Measure for Robust Object Localisation', IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2007.

[4] N. Otsu, 'A threshold selection method from gray-level histogram,' *IEEE Transactions on System Man Cybernetics*, Vol. SMC-9, No. 1, 1979, pp. 62-66.

[5] C. Ó Conaire, E. Cooke, N. E. O'Connor, N. Murphy and A. F. Smeaton, 'Fusion of Infrared and Visible Spectrum Video for Indoor Surveillance', WIAMIS 2005 - 6th International Workshop on Image Analysis for Multimedia Interactive Services, Montreux, Switzerland, 13-15 April 2005.