3D Computer Vision Techniques Applied to Infra-red Imagery

Carl Stennett, Chris Harris, Richard Evans
Roke Manor Research

Roke Manor, Romsey, Hampshire SO51 0ZNLA

Abstract

Thermal infra-red cameras have the ability to operate at night and in smoke, two significant advantages compared with visual-band cameras when used in military applications. Much vision processing research is performed for convenience with visual band cameras and the question of whether the developed techniques carry over into the infra-red (IR) domain is either not addressed or assumed to be straightforward. Here we consider the carryover into the IR of a 3D computer vision system (DROID) which is based on the detection and tracking of image features. We also briefly consider the carryover of an event detection system (VMAD) which is also based on feature extraction and tracking. We summarise the findings of an initial study based on readily available IR imagery, and describe recent work based on the experimental processing of project-specific recorded data. We compare the characteristics of IR and visual band imagery with respect to these applications and discuss their advantages and disadvantages.

Introduction

Roke Manor Research (Roke) has been engaged in the development of two vision processing systems, both based on the detection and tracking of point features. These are DROID, a 3D vision system, and VMAD, an event detection system.

As is convenient for research work, these systems have been developed in the context of visual band cameras as these are readily available at low cost and with high specifications. One of the major limitations of visual band cameras is that they do not work at night without active illumination, greatly diminishing their military benefit. Thermal IR cameras however provide 24 hour operation and operate in smoke, so the applicability of the processing techniques is greatly enhanced if they can be carried over from the visual band into the IR.

Here we report work investigating the carryover of feature-based techniques and we have taken DROID and VMAD as representative applications.

DROID and VMAD

DROID is a feature-based structure-from-motion system. Briefly DROID detects image features viewed by a camera moving through a scene and tracks their apparent motion. By analysis of the feature motion DROID is able to measure in 3D both the positions of the tracked features and the path taken by the camera platform. The features in question are extracted using the Harris corner operator and feature tracking is performed directly in 3D. An early but detailed description of DROID is given by Harris [1]. More recently DROID has been enhanced in a number of ways in work supported by the EMRS DTC [2]. DROID is being exploited for autonomous system applications in work [3] supported by the SEAS DTC.

VMAD is an event detection system which is also based on extraction and tracking of features [4]. As in DROID, Harris corners are used but tracking is performed in 2D, i.e. in image coordinates. Briefly VMAD accumulates statistics about the normal distribution of moving features in both position and velocity. Features which do not behave consistently with the model of normal activity are flagged as anomalies. VMAD was originally developed for use with static cameras, and has been extended to include applications based on moving cameras in EMRS DTC projects [5].

Initial Study

Before embarking on an experimental study we completed an initial assessment based on available imagery and literature (e.g. on the web). This led us to focus attention on uncooled IR cameras, partly because of interest in low cost solutions in general, but also because we were interested in comparing visual and IR systems on a broadly like-for-like basis in terms of cost and application; the use of cooled cameras would significantly increase the overall vision system cost. We also identified a preference for LWIR (Long Wave Infrared 8-14 μm) systems because of the reduced impact of sunlight transients in this band.

We also identified issues likely to impact the success of the feature-based techniques when applied to IR data. Most notable was the apparent blandness of some IR scenes. In addition the reduced sensitivity of IR systems (relative to visual-band systems) affects exposure times and image depth of field, causing increased risk of image blur because of motion and lack of focus.

Experimental Data

Data Collection

We collected data from two uncooled LWIR cameras. These were a commercially available thermographic camera and a military imaging camera in the form of a Driver's Vision Enhancer (DVE). The thermographic camera was largely used for preliminary work with most final results obtained with the DVE. Key specifications of these cameras are shown in the following table.

	Thermographic	**DVE**
Device type	Thermographic sensor	Imaging sensor
Band	8-13.5µm LWIR	8-14µm LWIR
Field of view	45°	45°
Resolution	640 x 480 pixels	320 x 240 pixels
Frame Rate	30Hz	25Hz
Focus	Variable (<1m to infinity)	Set to infinity
Sensitivity/NETD	60mK at 30°C	n/a

Figure 1: Key camera parameters

Datasets were recorded with both a static camera for use with VMAD and a moving camera for DROID. For convenience the moving camera data was recorded from the open boot of a car driven around the Roke Manor site; recorded data was then time-reversed for processing. Due to the relatively long exposure times (about 1/30sec in both cases) the speed of the moving platform was limited to about 3m/s to avoid motion blur. For the DVE, with its fixed focus, imagery in the near foreground was slightly out of focus. For the thermographic camera we chose a compromise focus setting with the middle distance in focus.

Preprocessing

We found that both cameras exhibited a low level noise pattern. As this varied over periods of several minutes we will refer to this as a fixed pattern noise (FPN). For imaging purposes this was barely perceptible (as the eye and brain can perform the necessary integrations) but we found it interfered with the performance of the feature extraction algorithms.

In the case of the moving camera we were able to reduce the FPN level by a combination of spatial and temporal filtering involving robust estimation techniques. A typical "before and after" result is shown in Figure 2. We believe an improved FPN reduction would be possible, particularly given more knowledge of the noise generating process. For example knowledge of the time constant of the pattern would enable better selection of noise reduction parameters, and we have not exploited the row-column structure that is apparent in the noise pattern.

Figure 2: Typical IR images before and after FPN reduction (images are shown hi-pass filtered to enhance the noise pattern)

Comparison of IR & Visual Data

Richness of features

To the eye, IR images do appear blander than visual band ones, but it is difficult to make an objective comparison of blandness. Even with comparable images it is possible to vary feature extraction threshold to give any number of detected features. Instead we made a comparison by processing data from a moving camera, setting low feature extraction thresholds and applying feature tracking algorithms. We assume that features that form reliable tracks are genuinely features in the viewed scene rather than the result of transient noise patterns.

To obtain data for this experiment we mounted a visual band camera side-by-side with the DVE camera. As the visual camera had a higher resolution and slightly different field of view we cropped and resampled the visual data to match the field of view and resolution of the DVE. This enabled us to compare three sets of data:-

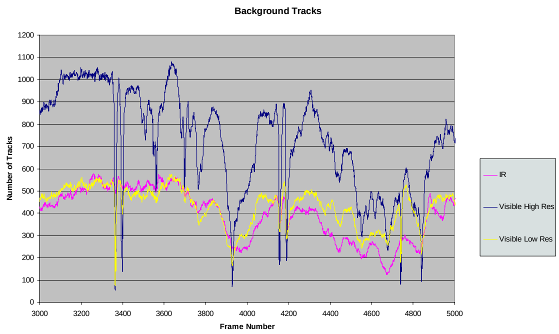
- The visual data at its original resolution, cropped to the field of view of the DVE. The resulting resolution was 584 x 405 pixels.
- The visual data cropped and resampled to the field of view and resolution of the DVE, i.e. 320 x 240 pixels.
- The DVE IR data, 320 x 240 pixels.

In each case we counted the number of tracks as a function of frame number, counting tracks for the image background and foreground separately. As the horizon was typically halfway down the picture we defined the top half of the image as background and the bottom half as foreground. Results are shown in Figure 3 for a typical sequence.

For image backgrounds we find the numbers of features generated by the IR and low resolution visual imagery are very similar. The high resolution visual data provides the largest number of features. The occasional glitches in the traces appear to coincide with times when the vehicle is turning sharply, a situation which the particular tracking algorithm used for this work finds hard to cope with and where image motion blur may be most pronounced.

For image foregrounds we again see the high resolution visual data provides the highest number of reliable features, though the margin is not as great as for image backgrounds. Similar numbers of features are provided by the low and high resolution visual data, except at the largest peak occurring here. The IR data generally provides fewer features, though there are usually about 25 or more foreground IR features. Occasionally the number of IR features drops to about 3 or less, and from experience with DROID we would expect processing to be problematic on such occasions.

Inspection of the images by eye revealed that the moments when all three traces are at similar levels coincide with a large number of fallen leaves being seen on the ground.



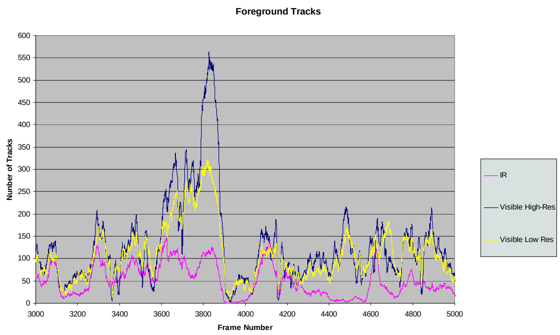


Figure 3: Plots of trackable featuresfeatures in imaage background (top) and foreground (bottom) regions

Other Comparisons

Our analysis above objectively shows visual band data to be less bland than equivalent IR data. From our initial study and the other areas of the more recent work we have identified a number of other advantages of visual band data over IR.

Overall, visual band cameras offer higher specifications at lower cost. An obvious aspect of higher specification is the higher spatial resolution available with current visual band sensors. The higher sensitivity of visual band sensors has several knock-on affects.

- Much reduced noise & sensor artefacts. This improves the detectability of low contrast features
- Shorter exposure times are possible. This reduces the risk of motion blur for sensors mounted on a moving platform.
- Smaller aperture lenses with greater depth of field. This makes it easier for a camera to accommodate a spread of operating ranges such as would be required in a vehicle mounted system. A further advantage of a good depth of field is that calibration targets of more manageable size can be used for applications like DROID where a geometric calibration of lenses is required.

Conversely we have concluded visual band data has several disadvantages compared with IR data in addition to the well advertised inability to operate at night and in smoke. In the visual band, sunlight is responsible for the following troublesome effects, all of which are essentially missing from LWIR data:-

- Sun glint and flare overloading the sensor and leading to other related artefacts such as streaks across the image – this can occur when the camera directly views the sun, indirectly views it reflected from windows etc. or sees internal reflections in the camera lens.
- Moving shadows – typical examples are the shadows of tree branches blowing in the wind or the shadows of moving people or vehicles.

These can be problematic in vision systems as they can be misinterpreted as moving objects in their own right.

- Deep shadows – these are most problematic at times of low sun angle when an image can include both brightly illuminated regions and shadow regions simultaneously. This is a particularly stressful situation for AGC (Automatic Gain Control) typically used on visual band cameras.
- Changes in ambient illumination caused by the sun emerging from behind or being obscured by a cloud.

While algorithms have been developed to counter these artefacts in the visual domain, they remain problematic. Overall we believe that the advantages of IR over visual band data will have a very significant effect on the ability to exploit vision processing systems.

3D Vision Processing of IR Data

The application of DROID to IR image sequence data is illustrated in the following figures. Figure 4 shows an example frame with tracked features and their estimated range from the camera is indicated by colour coding: red closest to the camera, then yellow, green, with blue furthest away. The colour codings are consistent with an interpretation of this scene by eye. Figure 5 shows the vehicle path, again extracted by DROID processing of the IR data. The track shown accords well with the track actually taken. The loop in the top left is formed by the passage between the rows of cars.

In processing this data we estimated the lens calibration by trial and error. (We normally use a calibration target, but this would have been problematic for IR data as mentioned above.) This was possible as DROID processing of monocular data is comparatively forgiving of calibration errors, though had we used stereo data or required a metrical optimisation of the 3D results then a more precise calibration would have been needed.

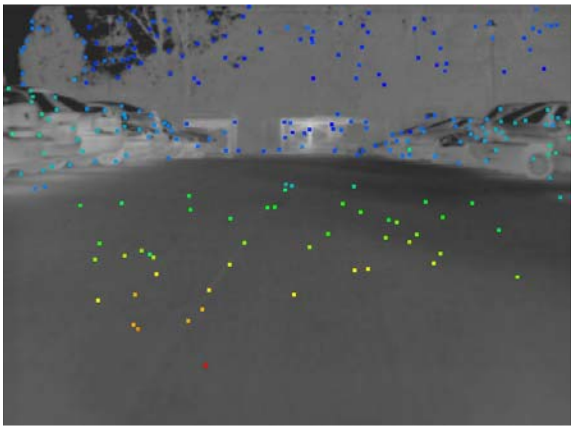


Figure 4: 3D range data generated by DROID processing of an IR sequence

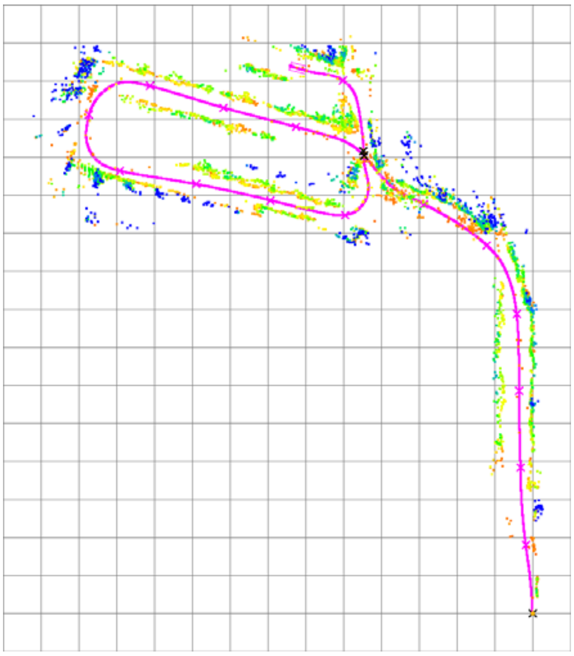


Figure 5: Vehicle path data generated by DROID processing of an IR sequence

We have also used a technique developed in our SEAS DTC to resolve the *speed-scale* ambiguity which needs to be overcome when processing monocular data. Briefly, we have exploited knowledge of the position of the camera on the vehicle to determine the scale of the extracted 3D information.

Figure 6 provides further examples of the 3D feature data at other places along the route. Again the colour codings are consistent with an interpretation by eye.

In practice we have found DROID processing of our IR recordings to be very promising. In places the shortfall of features extracted and tracked in image foregrounds is sufficient to make for poorly defined scene structure. In extreme cases a lack of features in image foregrounds can cause the translational components of vehicle motion to be ill-defined, leading to a catastrophic failure of the system. In practice such extremes have not been observed.

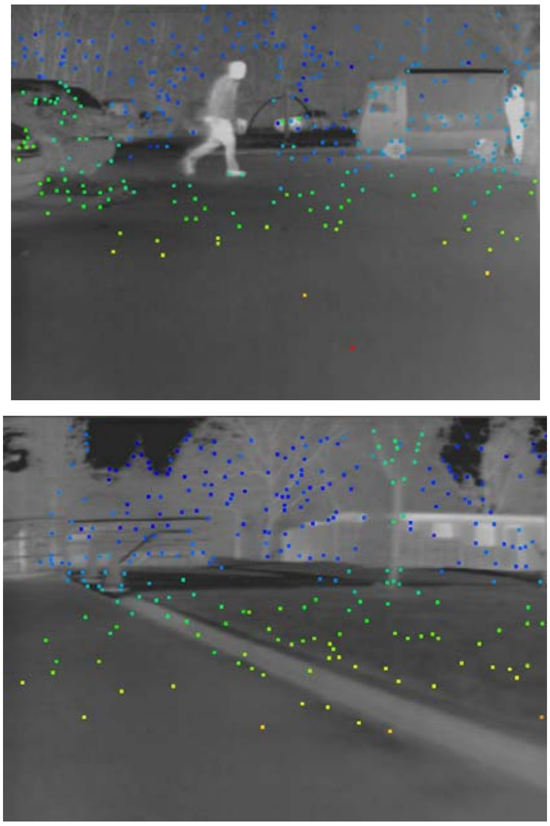


Figure 6: Further examples of DROID 3D feature output

Event Detection Processing of IR data

Figure 7 shows examples of VMAD output for an indoor and an outdoor scene. In both of these situations VMAD appeared to work well and functioned generally as it does with visual band data, though we did not perform a quantitative comparison. Moving objects were readily detected. Though it was not possible to apply the FPN reduction algorithms for VMAD because VMAD uses a static camera, the features generated by moving objects are generally higher contrast and so easier to detect.

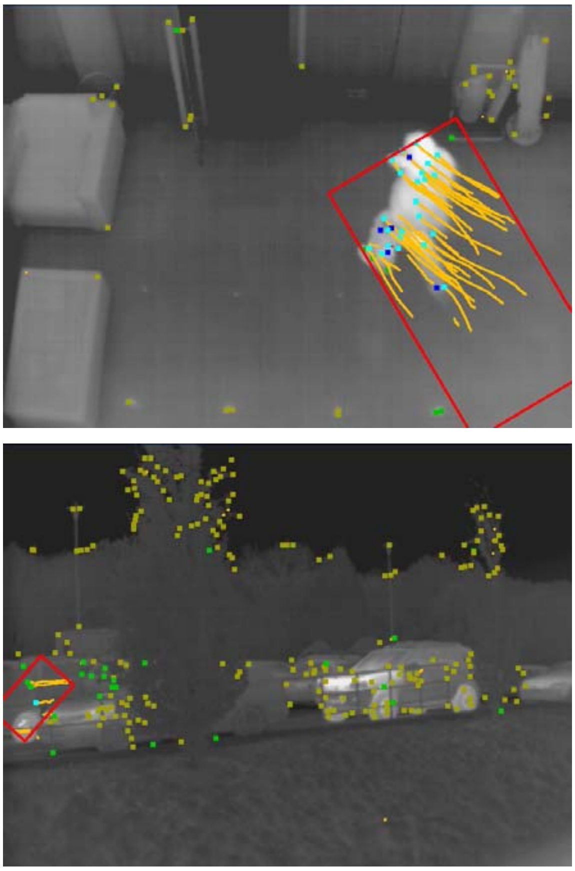


Figure 7: Examples of VMAD output

Supplementary Processing Techniques

In view of the relative lack of foreground features in IR data we have considered approaches to improvements for the DROID application.

Where the lack of features is sufficient to cause less well-defined 3D surfaces in the foreground we have identified two potential approaches

- Use of simple scene interpretation techniques, based on the assumption that visually flat areas are geometrically flat, to generate surfaces from the features that are available.

- Extrapolation from the ground beneath a vehicle to connect with the nearest features of the observed foreground.

Where the lack of features is more extreme, techniques to detect additional features or images flows have been identified, and these would also aid surface definition.

- Application of the existing feature detector at multiple scales.
- Limited use of patch correlation methods in areas of special interest.

Conclusions

Overall we have found a good carryover of feature-based vision processing techniques from the visual domain into the infrared. The use of feature-based structure-from-motion techniques (such as DROID) is generally successful though image foregrounds are found to be less richly featured in the IR, resulting in less well defined 3D structure. The problem has not been found to be so extreme in practice to cause loss of vehicle path data, and approaches to supplementary processing to improve this aspect of performance have been identified if required. The application of feature-based event detection methods (such as VMAD) appears to be straightforward.

Given the more predictable nature of IR data and the absence of transients caused by sunlight effects, we believe it may be easier to exploit the vision processing techniques considered here in the IR rather than the visual band, provided the sensor limitations can be overcome. In particular the lower sensitivity of IR sensors means imagery is more prone to poor focus and motion blur problems than visual band cameras – factors which adversely affect their use for 3D vision processing applications. With this in mind we recommend that future improvements in detector sensitivity be traded for improvements in depth of field and exposure times areas rather improvements in spatial resolution.

As sensor developments can be expected we do not recommend generic vision-processing research at this stage based on IR data – as much of the work would be taken up with in addressing sensor issues. However, with the results obtained here, we recommend future work addressing specific applications of existing sensors as this is likely to achieve a more rapid pull-through of research into exploitation.

Acknowledgements

The work reported in this paper was funded by the Electro-Magnetic Remote Sensing (EMRS) Defence Technology Centre, established by the UK Ministry of Defence and run by a consortium of SELEX Galileo, Thales UK, Roke Manor Research and Filtronic.

.

The authors wish to thank Thales Optronics Ltd for the loan of a Driver's Vision Enhancer and providing associated technical support.

References

1. Chris Harris, "Geometry from Visual Motion" in Active Vision, Eds Blake & Yuille, MIT Press, 1993
2. Chris Harris, R J Evans, P Saddington, "Stable Scene Surfaces from Computer Vision", 4th EMRS DTC Technical Conference – Edinburgh 2007.
3. C Harris "Strategies for Visual Exploration of Buildings", 2nd SEAS DTC Technical Conference, Edinburgh 2007
4. R J Evans, E L Brassington "Video Motion Processing for Event Detection and Other Applications" IEE Annual Conference on Visual Image Engineering, VIE2003, University of Surrey

5. R J Evans, R G Porges "Video Motion Anomaly Detection for Military Applications", 3rd EMRS DTC Technical Conference, Edinburgh 2006