Video Motion Anomaly Detection for Military Applications

R J Evans, R G Porges
Roke Manor Research Limited
Roke Manor, Romsey, Hampshire SO51 0ZN

Abstract

The requirements of military operations, amid ongoing civilian activity, are challenging traditional methods of threat detection. This paper provides an update on previously reported work aiming to extend techniques emerging from developments in civil CCTV for military applications. In particular we consider anomaly detection based on the analysis of visual feature motion extended to moving and or/rotating sensor platforms. We summarise earlier work and outline processing algorithms, including some recent improvements, processor requirements and some experimental results.

Keywords: CCTV, Intelligent CCTV, Event Detection, Anomaly Detection

Introduction

This paper provides a view of the EMRS DTC's project on the Analysis of Visual Feature Motion for event and threat detection. This is being carried out by Roke Manor Research Ltd (RMRL). This work aims to transfer techniques emerging in civil applications of CCTV - such as for security and traffic monitoring - to the military arena. More specifically we aim to build on the ideas behind VMAD [1, 2] – a Video Motion Anomaly Detector developed by RMRL.

Aims and Objectives

Military operations are increasingly required where there is terrorist activity within civilian populations and where enemy personnel are not easily distinguished *per se*. In these situations requirements of military and civil security systems converge and the transfer of technology from civil to military use becomes attractive. VMAD is a promising starting point because of its learning capabilities, which reduce the expertise and effort required for system installation.

Our project objective has been apply the VMAD concept to non-static platforms, with a view to applying it in a wider range of situations. Two distinct cases of interest were identified at an early stage. The simpler case is an *emplaced* camera, typically one which might be set up at a command or observation post. It might be hand-held or set on a pan-tilt mount, but its location is fixed. The more challenging case, though probably of longer term interest is a *moving* sensor, typically one which is mounted on a UAV which loiters or revisits the same area of ground in its surveillance.

Progress with the emplaced camera has been good and the recent work has concentrated on this with a view to setting up an experimental implementation. However we have not yet managed to demonstrate the higher levels of anomaly detection that we originally intended. In the moving camera case we have identified a promising approach to the basic processing step (distinguishing between the apparent motion of moving and stationary objects) and recently we have verified this approach with data from an airborne sensor.

At the time of writing we have submitted proposals to pursue this further. The remainder of this paper will concentrate on the emplaced camera.

Previous Work

RMRL's VMAD analyses the motion of independently tracked point features. This contrasts with other approaches to CCTV-based event detection, the main ones being

- Simple "motion" or "activity" detectors: These actually respond to changes in image brightness. Though simple to implement they do not work well if lighting is variable.
- "Segmentation-based"/"object-based" [e.g. 3, 4]: These often use application-specific processing with knowledge of scene geometry, so they need to be carefully set up. "Trajectory analysis" of objects can be rewarding but sophisticated processing is required. Hogg [5], for example, applies active shape models.
- "Frame-based" methods, for example used by Davis & Bobick [6]: Here there is no attempt to segment specific objects. Instead each video frame is analysed as a whole in terms of an image velocity flow pattern.

In VMAD a feature extractor [7] is combined with a 2D tracking system to form a Video Motion Processor. This front-end feeds an anomaly detection system, which can learn the normal behaviour of tracked features and respond when abnormal events occur - see figure 1.

The front-end processor reduces image data to a stream of tokens describable by a four-dimensional vector and a four-dimensional histogram is used to capture a summary of the behaviour observed during training. Once trained, possible anomalies can be detected when track data falls into low-occupancy cells. As feature extraction and track processing is not perfect, some false alarms may be generated by spurious tracks, but these can be suppressed by additional processing logic.

The histogram method has the advantage that a fading memory can be readily implemented if required. The potentially large storage required for a four-dimensional array can be reduced by judicious variation in cell sizes.

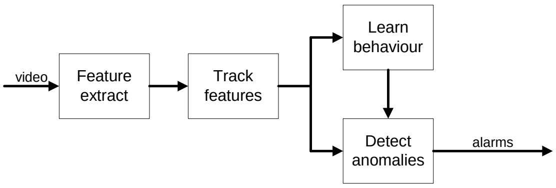

Figure 1 *VMAD block diagram*

These algorithms have been extended for the emplaced camera case, mainly by modification to the tracking algorithms and inclusion of routines to estimate the current camera pointing direction so that activity can be mapped onto a global coordinate system. In studying the emplaced camera case we have also identified valuable enhancements to the underlying VMAD system. The main changes were

- The use of polar coordinates for velocity histograms in place of logarithmically spaced bins in Cartesian coordinates. The earlier scheme made the system more likely to alarm on tracks moving in some directions rather than others.
- Delayed histogram updates: In the original VMAD coding, abnormal tracks were self limiting because data was checked against the histograms and used to update the histogram in the same video frame time.
- Improvements to tracking algorithms as problems were found with clusters of stationary intermittent features.

Example Static System

An interesting illustration of the basic VMAD system is provided by processing data from a camera set up for experimental purposes to monitor a section of the Roke Manor's grounds.

Selected alarm images are shown in Figure 2. In this experiment, the system picked up events of an expected nature, such as vehicles on the track just outside the perimeter fence and pedestrians. The blue borders in the figure indicate true alarms. False alarm sources included birds, insects (attracted by the illumination used at night) and some un-attributed events. The system also detected a man with a gun – hired for pest control!

Figure 2 *Selected alarm images from static trial*

The Emplaced Camera

Algorithms

Algorithms for the emplaced camera scenario have remained largely as previously reported [8], so only an outline is presented here. A key step in the processing chain is the calculation of frame-to-frame camera rotation. We have used for this purpose the *de-lurch* algorithm originally developed in the EMRS DTC 3D Computer Vision project [9].

The de-lurch algorithm has two main stages. First a very high confidence subset of stationary features matched between frames is selected. This is done by considering the angles subtended by pairs of features. If the features are correctly matched and the features are stationary, then the subtend angles should not change from frame to frame. We consider all possible pairs of features and discard as unreliable those matched features that are frequently involved in pairs whose subtended angle changes appreciably. In the second stage of the algorithm, the camera motion is measured by a least-squares fit to the retained feature matches.

With appropriate scaling of track parameters, the revised video motion processor can be interfaced to the existing anomaly detection code. Simple changes to the track maintenance rules, with different coasting and track deletion rules for stationary and moving features, ensure that when the camera moves and stationary features move out of view, they are still retained for future reference.

Processing Requirements

We have now rationalised the implementation and improved the efficiency of the code so that its runs in real-time on a standard PC. The software has been benchmarked on a 3.2GHz Pentium 4 processor. When processing pre-recorded data sequences the overall speed is approximately 20% faster than real time (with an image size of 384 x 288 monochrome pixels at 25 frames per second). We have measured the proportion of time spent in different parts of the code.

Extrapolating to an operational system we estimate that the implementation is within the reach of a 1.2GHz Pentium M processor, which is low power and suitable for military use.

Recent Car Park Experiment

The emplaced camera was set up overlooking Roke Manor's main car park. (We had attempted to obtain a site more representative of a military application.) The camera used was an indoor "dome" camera - i.e. an integrated camera and pan-tilt mount - positioned on an indoor window ledge, looking outward. The camera was set to a constant instantaneous horizontal field of view of about 42º. The camera was then set to scan horizontally over a total field of view of about 103º, a complete scan cycle taking 30 seconds.

The car park set-up was run for several days, logging alarms and video clips at the time of alarms, though only the first two days results were analysed in detail. The fading memory half-life was 5 minutes, thus the system should ignore relatively short term activity, such as trees waving in the wind and the like, but still respond to the movement of cars and people.

Figure 3 shows hourly counts of alarms from all sources over two days from start-up. (The first slot is scaled from a count over half an hour.) Night-time data has been omitted for reasons given below. The peaks in this graph reflect the daily pattern of activity, notably lunchtimes and the end of core working hours at 4:30pm on a Thursday and 3:30pm on a Friday.

Figure 4 shows true and false alarms counted per hour on the first day of the experiment. By a *true* alarm we mean an alarm which is associated with the passage of a vehicle, bicycle or person on foot. By a *false* alarm we mean an alarm arising from any other source. Alarm status was determined by inspecting the logged video clips and still frames by eye. This has allowed a fairly accurate assessment of alarms though some events are difficult to interpret and may have been misclassified. From the Thursday results, the false alarm rate appears to be about 10 alarms per hour. The first slot is understandably higher as this is dominated by the initial learning period. The event detection probability has not been assessed. Inspection of the logged video clips provides some additional insights into alarm performance because the clips occasionally record other activity besides that which has triggered the alarm. As a result we believe the detection probability for foreground events (pedestrians on the nearby path or vehicles close to it) is high, possibly 90%, but the detection rate for background events is much lower.

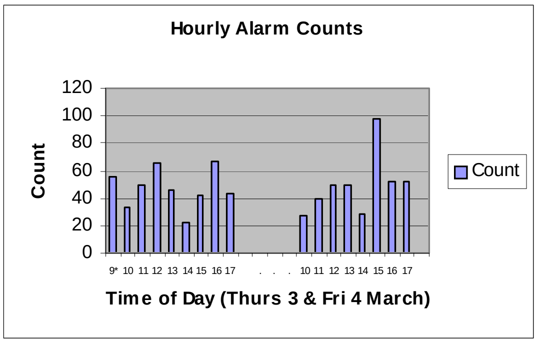

Figure 3 *Hourly alarm counts from the recent car park experiment – daytime only.*

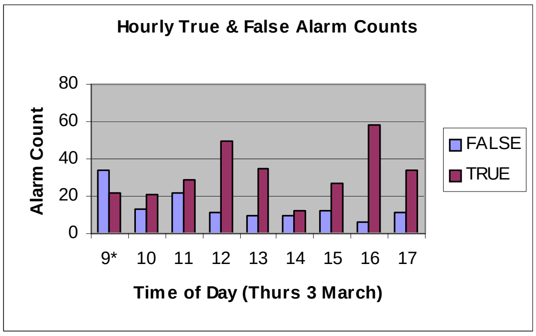

Figure 4 *Hourly true and false alarm counts from the recent car park experiment.*

Figure 5 shows a selection of alarms images, i.e. images captured at the time of an alarm, from the car park experiment.

The red boxes in the images show the approximate position of the activity which generated the alarm. Some arise from genuine movements in the scene, e.g. movement of vegetation caused by sudden gusts beyond the levels learnt by the anomaly detector. Other false alarms appear to be the result of tracking errors in particularly busy areas of the field of view.

At night our experimental set-up proved to be seriously lacking. The camera AGC was inadequate (we had no special lighting) and auto-focus failed at times; there were internal reflections from the window and raindrops outside distorted the optics. (Fortunately it did not rain during the first two days of the trial.) As a result of these defects the system did not function overnight.

These deficiencies meant that the camera pointing-angles were not maintained overnight. As previously described, the system needs to maintain consistent estimates of camera pointing-angles, in order to tell what is normal or abnormal activity at any time for the current field of view. Consequently we tested the system in the lab for several days to check pointing angle consistency. Even so in the trial we noted sporadic errors in calculating the camera pointing-angle during the day. These errors took the form of step changes in the estimates of camera azimuth/elevation. There were no “slippages” on the Thursday but there were three on the Friday. With the learning time half-life set at 5 minutes, much less than the time between pointing angles slippages, the impact of these errors on event detection should be small. Had the learning time been larger than the interval between pointing angle slippages, then the impact would have been substantial because the system would never complete a learning period before a slippage occurred. Unfortunately ambitions to learn patterns of activity over periods of several hours are thwarted with the present level of reliability.

Figure 5 *A selection of alarm images from the recent car park trial.*

Earlier in the project we performed a trial of just 2.5 hours duration, monitoring a smaller car park [8]. Figure 6 shows a two alarm images. The false alarm rate was about 5-10 alarms/hour and in that short experiment all visible pedestrian and vehicle activity was detected. It should be noted that though nominally similar, the scene in the earlier experiment was much less "busy" than that of the more recent experiment and almost all the visible activity of interest takes place in the foreground of the scene. This may account for the superior alarm detection performance observed earlier.

Figure 6 ***A selection of alarm images from the 1st car park trial.***

Conclusions

We have developed a system for video motion anomaly detection, online to an emplaced camera – without other instrumentation - and these algorithms appear to be amenable to a military implementation. Key concepts have been demonstrated. Exercised in a fairly busy scene, however, performance has been limited partly by occasional failures in feature tracking, but more so by the current reliability of algorithms to calculate camera pointing angles.

At the present state of development, this weakness limits the available training periods and the complexity of normal activity that can be learnt. Assuming these weaknesses can be overcome, greater care will be needed in setting up equipment for trials over extended periods.

Acknowledgements

The work reported in this paper was funded by the Electro-Magnetic & Remote Sensing Defence Technology Centre established by the UK Ministry of Defence (Project reference EMRS/DTC/3/64).

The authors acknowledge the assistance given by other RMRL staff including Ed Sparks, Carl Stennett and Esin Turkbeyler.

References

1. 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.
2. GB Patent Number GB2392033 "Video motion anomaly detector" filed 18 November 2002, granted 20 October 04.
3. Cernium Inc., see www.cernium.com
4. Foresti G L, and Roli F, 2000, "Learning and Classification of Suspicious Events for Advanced Visual-Based Surveillance", in Ed Regazzoni et al, "Multimedia video-based Surveillance Systems: requirements, issues and solutions", Kluwer Academic Publishers
5. N. Johnson and D. Hogg, 1995 "Learning the distribution of object trajectories for event recognition", Proc. British Machine Vision Conference, 583-592
6. J. Davis and A. Bobick, 1997 "Representation and recognition of human movement using temporal templates", Proc. IEEE Conf. Computer Vision and Pattern Recognition.
7. Harris C and Stephens M, "A Combined Corner and Edge Detector", Proc. 4th Alvey Vision Conference.
8. R J Evans, E Turkbeyler, E P Sparks, " Analysis of Visual Feature Motion for Event and Threat Detection", 2nd Annual Technical Conference of the EMRS DTC, Edinburgh
9. R J Evans, C G Harris, *et al* "3D Computer Vision for Autonomous Remote Sensing Platforms", 2nd Annual Technical Conference of the EMRS DTC, Edinburgh.