Computer vision for security applications

Kingsley Sage
Police Scientific Development Branch
Langhurst House
Langhurstwood Road,
Horsham
RH12 4WX
United Kingdom

Stewart Young
Sira Technology Centre
South Hill
Chislehurst
BR7 5EH
United Kingdom[1]

Abstract — In an age which bears witness to a proliferation of Closed Circuit Television (CCTV) cameras for security and surveillance monitoring, the use of image processing and computer vision techniques which were provided as top end bespoke solutions can now be realised using desktop PC processing. Commercial Video Motion Detection (VMD) and Intelligent Scene Monitoring (ISM) systems are becoming increasingly sophisticated, aided, in no small way, by a technology transfer from previously exclusively military research sectors.

Image processing is traditionally concerned with pre-processing operations such as Fourier filtering, edge detection and morphological operations. Computer vision extends the image processing paradigm to include understanding of scene content, tracking and object classification. Examples of computer vision applications include Automatic Number Plate Recognition (ANPR), people and vehicle tracking, crowd analysis and model based vision. Often image processing and computer vision techniques are developed with highly specific applications in mind and the goal of a more global understanding computer vision system remains, at least for now, outside the bounds of present technology.

This paper will review some of the most recent developments in computer vision and image processing for challenging outdoor perimeter security applications. It also describes the efforts of development teams to integrate some of these advanced ideas into coherent prototype development systems.

INTRODUCTION

Image processing first began to have an impact on security technology in the early 1980s with the introduction of Video Motion Detection (VMD) systems which claimed to revolutionise Perimeter Intruder Detection Systems (PIDS). Poor performance (in particular, high false alarm rates) of early operational installations led to research effort directed towards producing more sophisticated solutions. Concepts from artificial intelligence, such as neural networks and expert systems, have been incorporated into systems which are more computer vision orientated.

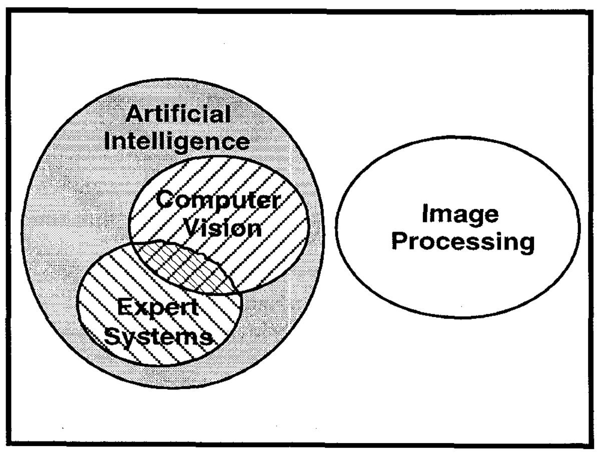

Figure 1: The relationship between image processing and computer vision

[1] In conjunction with the Department of Physics and Astronomy, University College London, London, United Kingdom under the Sira/University College of London Postgraduate Training Partnership.

0-7803-4535-5/98/$10.00 ©1998 IEEE

This paper considers novel developments in 3 areas:

- improved imaging of intruders;
- automatic alarm verification by video means; and
- automatic threat assessment in real time.

HISTORICAL BACKGROUND

The earliest intruder detection image processing algorithms employed frame differencing, a technique which still forms the cornerstone of many of today's more complex systems. The idea is to pick out all those scene areas which are different from one frame of video to the next. This (typically first order) differential function picks out frame to frame movement. When a movement metric exceeds a preset threshold, an alarm is signalled. This simple technique (see figures 2 & 3) provides poor discrimination between intruder movement and camera shake (a particular problem), rain and structural changes to the scene content brought about by sunlight and shadow effects. Later refinements included Fourier and median filtering to counter rain (low pass filter to limit high frequency effects; also reduces strength of scene edges) and lighting gradient effects (high pass filter to remove low frequency and dc transients making the whole scene appear rather flat).

Figure 2: Intruder approaching a fence

But there are problems to this simplistic approach which are that:

- it provides no concept of image understanding;
- there is little or no exploitation of domain knowledge;
- there is no concept of what the target is and how it moves; and
- it is susceptible to camera movement and weather effects.

It is still not COMPUTER VISION.

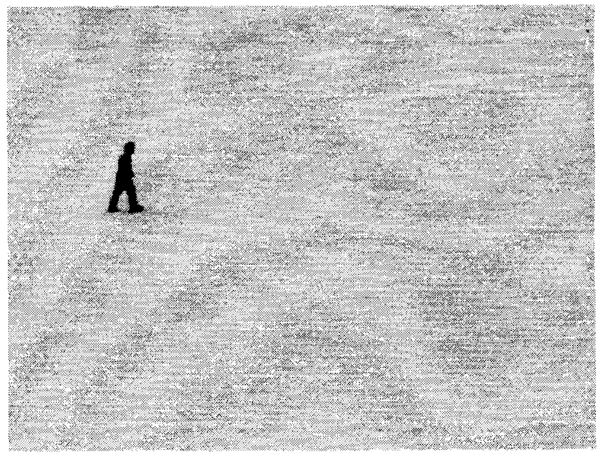

Figure 3: Difference image between figure 2 and a previous reference image

EXPERT SYSTEMS: RULE BASED PARADIGMS

Several research groups turned to expert systems for a next generation approach to video based PIDS. The PSDB FABIUS project [1] used a frame based expert system approach and separated the computer vision task into 3 separate tasks:

- object detection (by image processing means);
- target object recognition (by nearest neighbour and similar deterministic approaches); and
- target object tracking (by frame to frame correspondence of feature vectors).

A frame based expert system written in PROLOG was then used to classify target objects and their behaviour into overall classes, depending on their feature vectors and how they changed through a frame sequence. This approach allows domain knowledge to be incorporated in a meaningful manner, e.g.:

IF rabbit_object AND moves_at rabbit_speed THEN classify RABBIT

While this approach yields a better performance, it has the disadvantage that it requires extensive amounts of explicit domain knowledge. For example, information is required on what different types of wildlife may be present in the scene for each and every application. This amount of information becomes very large especially if probabilistic models (e.g. of weather effects) are incorporated into the framework.

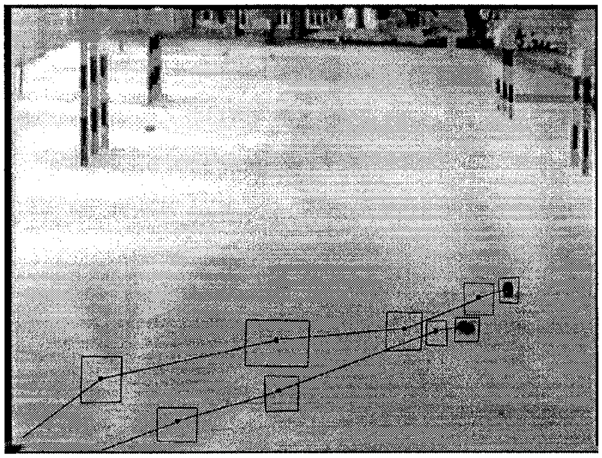

Figure 4: Analysis of 2 rabbit tracks

THE USE OF ARTIFICIAL NEURAL NETWORKS

One approach to a classification problem where the input/output relationship is highly non-linear, simply unknown, or where explicit knowledge carries too heavy an overhead is to use a neural network classifier.

There are 2 ways in which neural classifiers can be incorporated into video based PIDS:

- to perform low level processing of pixel values directly; and
- to perform classification on the outputs of image processing operators which process the raw pixel data first.

The low level approach has the advantages that the networks can learn highly non-linear statistical models for localised scene activity. This allows a system to be built which is capable of learning about complex, highly localised scene variations such as moving foliage, shadows and insect activity. An example of such an approach can be found at [2].

The second "classic" approach has the advantage of greater predictability and operation transparency and access to interim scene representation primitives. It is easier to build predictive models of system behaviour and performance [3].

The neural network approach overall has the advantages that:

- it can be highly self organising as well as used in supervised training modes; and
- it is highly adaptive, which makes it more suited to processing imagery derived from outdoor scenes.

Experimental work by PSDB shows that, despite these advantages, the performance of neural classifiers still remains at best one order of magnitude worse that other forms of PIDS sensor technology. Researchers have turned their attention to new approaches to try and match the performance of video based PIDS with other forms of sensors (such as those which are linear microphonic cable based).

THE DIRECTION OF CURRENT RESEARCH

So as to pursue improvements in video based PIDS technology, PSDB decided to pursue research in 3 areas:

- to improve system effectiveness by co-ordinating non-video sensors with video sensors: in particular, to use the video systems as a means of providing automated verification. This approach (the AMETHYST project) is reported by [4];
- to improve the sophistication and performance of underlying image processing techniques; and
- to adopt a systems approach using a number of different forms of video processing and combine them into a unified interpretation. Such a system could incorporate movement analysis, optic flow measurement, predictive weather model and human gait analysis and other features.

SCATTER PLOT ANALYSIS

As previously discussed, most first generation video based PIDS used image differencing. This first order technique, in some form or another, forms the basis of much image pre-processing to this day. Increasingly sophisticated variants on this approach rest on two principles: threshold selection (for determining when sufficient activity is taking place); and the generation of reference images (differencing between current and reference frames as opposed to strict frame to frame differencing). Strategies for threshold selection include methods based on: the standard deviation of a pixel's value over a sequence [5], the median of the absolute deviation [6] and hystereysis thresholds [7]. For background generation, the techniques include using median filters [8], least median of squares estimate [9] and an adaptive smoothness filter [10].

Many of these techniques are designed to cope with a particular sub-set of image processing problems. What is really required for an effective video PIDS is for a reasonably robust process which will expose image areas where coherent movement activity is taking place while suppressing unwanted incoherent detail (such as high frequency rain effects) and evolving structural changes (such as those caused by global lighting changes). Such an efficient representation must also be realised in a formalism which can be realistically implemented on a system with a level of processing power suitable for the security market (typically, a desktop PC).

One such novel approach by Young [11] uses an alternative representation for intensity information using a scatter plot. This is a two dimensional diagram, with the horizontal (x) axis being intensity in image I1, and the vertical (y) axis being intensity in another image I2 (NB: $x,y \in \{0,255\}$). Let the two sets of N pixels, one from I1 and the other from the corresponding area in I2 be labelled v1 and v2. Every pair of spatially corresponding intensity values creates a single scatter point on the plot:

Scatter Point [I] = (v1[i], v2[i])

Spatial information is not visually explicit in this representation but it is implied by the labelling of the points. Differences between the images may be detected by points which are displaced from the line y = x. If the two images are identical, all the plotted points will lie on this line. If the two images are nominally identical, differing only in electronic noise, the scattered points will spread out around y = x.

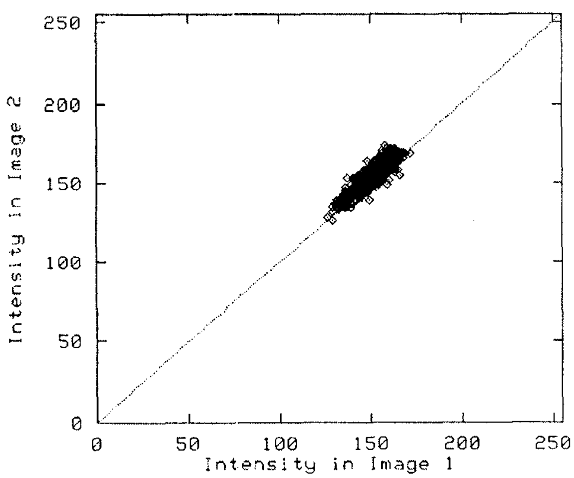

Figure 5

Figure 5 shows the scatter plot for an image of a sterile zone in daylight; by contrast figure 6 shows the scatter plot for the same scene with an intruder wearing dark clothing.

However, this method alone remains limited by the rigid spatial coupling between pixels. Consider dividing the image up into small areas, each containing N pixels. Once the image has been divided into these small areas, the task of change detection is redefined. Rather than simply asking whether or not an event has occurred anywhere in the scene, the question is asked about each small area. The output therefore retains some spatial structure, that of the relationships between the areas, and further processing is required to interpret the results and obtain a binary decision (alarm/no alarm).

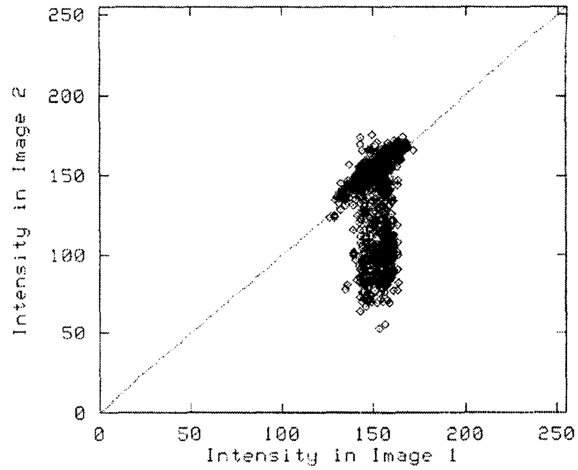

Figure 6

However, there is no need to pair pixels together according to the spatial ordering of the image. If we index the individual elements of v1 and v2 with respect to their values according to:

$v[i+1] \geq v[i] \ \forall \ i$

then the result is two so-called rank ordered vectors. Again, these two vectors may be compared using the scatter plot method:

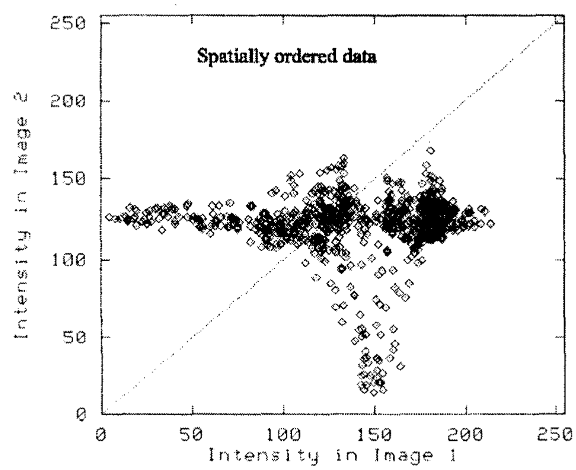

Figure 7: Data prior to rank ordering

It can be seen that rank ordering can have a significant effect upon the scatter plot. If the two images are nominally identical, the scattered points will, as before, lie close to the line y = x. In fact, the ordering will typically reduce the spreading of the points because of electronic noise. Another effect due to the rank ordering is that the vector index parameterises a specific path P, across the x-y plane.

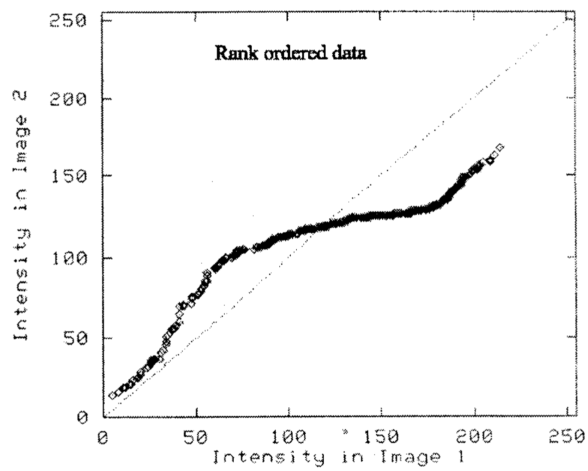

Figure 8: Rank ordered data

The shape and orientation of this path contains information about the relative numbers of pixels in each area having particular intensities and so, in some way, compares the histograms of the two areas.

How can we analyse the shape of this path ? We are interested in a number if properties of the distribution: the straightness of the line (if there are no changes then the path is a straight line); the orientation of the data (changes in illumination may simply change the slope of the path); and how evenly the spread the data is (how distorted from a straight line is the distribution). A statistical analysis of the data can give a parametric description.

A SYSTEMS APPROACH

An alternative approach to improving video based PIDS performance overall is to evaluate multiple sources of less reliable information together to produce a result with a greater confidence than any of its parts in isolation (the concept of data fusion). PSDB is currently undertaking programmes of work on exactly this so as to provide powerful video based automatic threat assessment. Examples of the generic threat that might be analysed include:

- suspicious activity of people in car parks (movement between vehicles may indicate the presence of thieves);
- suspicious activity of vehicles in proximity to banks (may be indicative of an impending armed robbery); and
- suspicious patterns of movement of people near safety critical equipment (possibly indicative of an impending act of vandalism).

An effective threat assessment system might combine:

- Automatic Number Plate Recognition (ANPR);
- the ability to detect specific types of vehicle (e.g. is that the normal cash delivery type of vehicle ?); and
- crowd density analysis (based on different types of optic flow analysis).

This paper cannot do justice to all of these technologies. Promising work worthy of note includes that of Hogg et al [12]. This work points the way to how an integrated system might work. It combines two model based approaches:

- using active shape models to track people as non-rigid objects; and
- using geometric 3D models to tracks cars as rigid objects.

These systems have been integrated to provide a powerful model based paradigm for image understanding highly relevant to high security applications. Applications investigate thus far include measuring patterns of pedestrian movement which are unusual and producing natural language descriptions of scene interactions. It is not difficult to imagine security supplications that could benefit from an integrated systems approach such as this.

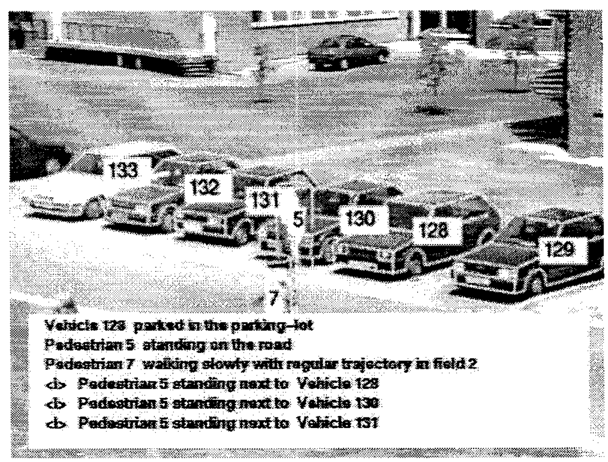

Figure 9: Model based surveillance system[2]

CONCLUSIONS

It is self evident that there is much more information present in a video stream than can be meaningfully processed by current technology. The difference is analytical prowess between even the most advanced standards in computer vision and an average security guard bear witness to this. That said, increasing pressures to increase guard effectiveness over long periods and reduce

[2] (courtesy of Leeds and Reading Universities: developed under grant from the EPSRC)

the cost of providing guarding services remain the driving forces behind research into advanced computer vision techniques. Such research will benefit from both continued work in basic areas as well as integrated vision systems, designed around the operational needs of any particular site. PSDB hope to be able to report on progress of their work in threat assessment using integrated systems in future conferences.

[1] Ellis, T J et al, "Model based vision for automatic alarm interpretation", *in International Carnahan Conference on Security Technology*, 1990 pp. 62-67

[2] Stubbington, B and Keenan, P, "Intelligent Scene Monitoring", *in International Carnahan Conference on Security Technology*, 1995

[3] Sage, K et al, "Estimating performance limits for an intelligent scene monitoring system (ISM) as a perimeter intrusion detection system (PIDS)", *in International Carnahan Conference on Security Technology*, 1994

[4] Horner. M, "AMETHYST: An Enhanced Detection System Intelligently Combining Video Detection and Non-Video Detection Systems ", *in International Carnahan Conference on Security Technology*, 1995 pp. 59-66.

[5] Atherton, T and Kerbyson, D, "Reducing false alarm rates in surveillance imaging using significance testing", in IEE Colloquium on image processing for security applications (1997/074), 1997, IEE London

[6] Rosin, P and Ellis, T, "Image difference and threshold strategies", in British Machine Vision Conference, 1995

[7] Canny, J, "A computational approach to edge detection", in IEEE Transactions PAMI, 1986 pp 679-698

[8] Rosin, P and Ellis, T, "Detecting and classifying intruders in image sequences", in British Machine Vision Conference, 1991

[9] Yang, Y H, "The background primal sketch: An approach for tracking moving objects", Machine Vision and Applications, 1992, pp 17-34

[10] Long, W and Yang, Y H, "Stationery background generation: An alternative to the difference of two images", Pattern Recognition, 1990, pp 1351-1359

[11] Young, S, "Video based intruder detection", MPhil report, University College London and Sira Technology Centre, Chislehurst Kent, UK, 1997 (unpublished)

[12] Hogg, D et al: "An integrated traffic and pedestrian model based vision system" , in British Machine Vision Conference, 1997, pp. 380-389