TRACKING PEDESTRIANS USING VISIBLE & INFRARED SYSTEMS

Dr. Alistair Armitage, Dr. David Binnie, Timothy Chamberlain, Mattias Nilsson and Dr. Malcolm Rutter

Faculty of Engineering & Computing

Napier University

Abstract

This paper describes research on several aspects of pedestrian detection and tracking being carried out at Napier University. A description is given of measurements of pedestrian trajectories made with low-cost, low resolution thermal imagers, produced by Irisys (Infrared Integrated Systems Ltd). These detectors provide accurate measurements over an area of approximately 10 square metres. We report on work that attempts to correlate the output of multiple detectors, so that tracking can be performed over wider areas, and in real-time (previously, recordings have been analysed off-line). Some new ideas on the use of these detectors for measuring pedestrian gait and signature analysis are discussed. Work on the use of higher-resolution infrared detectors is also described, and image processing algorithms for segmenting infrared images and recognising pedestrians are presented.

Introduction

The requirement for the detection and tracking of people and other mobile objects in indoor and open environments is becoming increasingly important, primarily for safety and security reasons, but also for urban planning Kerridge [2001], Kukla [2003] . Video surveillance is commonplace in UK city centres, shopping precincts and transport termini. Street scenes and active public areas are monitored by networks of cameras which produce large volumes of real-time video data on a daily basis. In transport video surveillance technology is used to measure traffic flow, detect road accidents and vehicle congestion. Increasingly CCTV systems are used to monitor pedestrian traffic such as congestion in public spaces, consumer behaviour in shopping malls and routine maintenance tasks at industrial facilities.

Currently observation and interpretation of scene events in real time, or recorded video, is a labour intensive human task. Continuous 24-hour monitoring of surveillance video is not viable due to the cost and impracticalities of mass data storage and analysis. Yet to be alert to a safety or security problem, real-time observation and interpretation is required if such a security breach is to be prevented or acted upon. Mounting video cameras is relatively cheap, but finding human resource to monitor the output is expensive and error prone. Although surveillance cameras are already prevalent in banks, stores, and parking lots, video data currently is used only "after the fact" as a forensic tool, thus losing its primary benefit as an active, real-time medium. There is thus an immediate need for automated surveillance systems in commercial, law enforcement and military applications. The most advanced systems have video motion detection capability. These detect all frame-to-frame changes in a video sequence but do not attempt to classify objects or interpret activity.

Over the last ten years there has been a number of excellent research examples of computer based scene interpretation, yet most examples fail when applied to real life applications. Reading[, Haritaoglu[2000] More research is required in this area if computers are genuinely to be used to find people and interpret position and behaviour in an unstructured scene. In the past, computational barriers have limited the complexity of real-time video processing applications. As a consequence, most systems were either too slow to be practical, or succeeded by restricting themselves to very controlled situations. More recently, faster computers have enabled researchers to consider more complex, robust models for real-time analysis of streaming data.

These new methods allow researchers to begin modelling real world processes under varying conditions.

The key property of a video surveillance and monitoring system is that it should be robust to whatever objects are in the visual field and whatever ambient changes occur. It should be capable of dealing with shadows, lighting changes, effects of moving elements of the scene (e.g. swaying trees), slow-moving objects, and objects being introduced or removed from the scene. Traditional approaches based on backgrounding methods typically fail in these general situations. The problems with daytime outdoor video surveillance are that the primary light source is 93 million miles away, the illumination changes slowly throughout the day and the illumination is prone to strong local variations in intensity due to weather conditions. Further, direct and indirect light can be affected by atmospheric effects and ground level objects. The key point is that for people detection, video cameras detect reflected light which in uncontrolled situations, is subject to many temporal and spatial variations in intensity, colour and dynamic range.

A more recent approach to people detection is to use infrared cameras: Fang [2003A, 2003B]. Infrared (thermal) cameras detect energy in the 7-14 μm waveband where the human body has its radiation peak. In this application, the thermal camera detects emitted not reflected radiation and is insensitive to the changes in the intensity of visible light. Furthermore, infrared radiation has better transmission properties in foggy and wet conditions. The thermal camera is an excellent tool for the detection and tracking of people. This is well known and portable hand-held thermal cameras are extensively used by search and rescue and security organisations. There are two reason why it is rare for thermal cameras to be used in static mounted surveillance applications: 1) Cost. - Thermal cameras are approximately one hundred times more expensive than visible wavelength video cameras and have around a quarter of the spatial resolution. 2) Lack of scene information. - a thermal picture of an outdoor scene will usually indicate the presence of a person or animal but the background detail or context information is much less than that of a visible wavelength camera. The identification of a person, exactly where they are, and what they are doing is not necessarily clear from a thermal image.

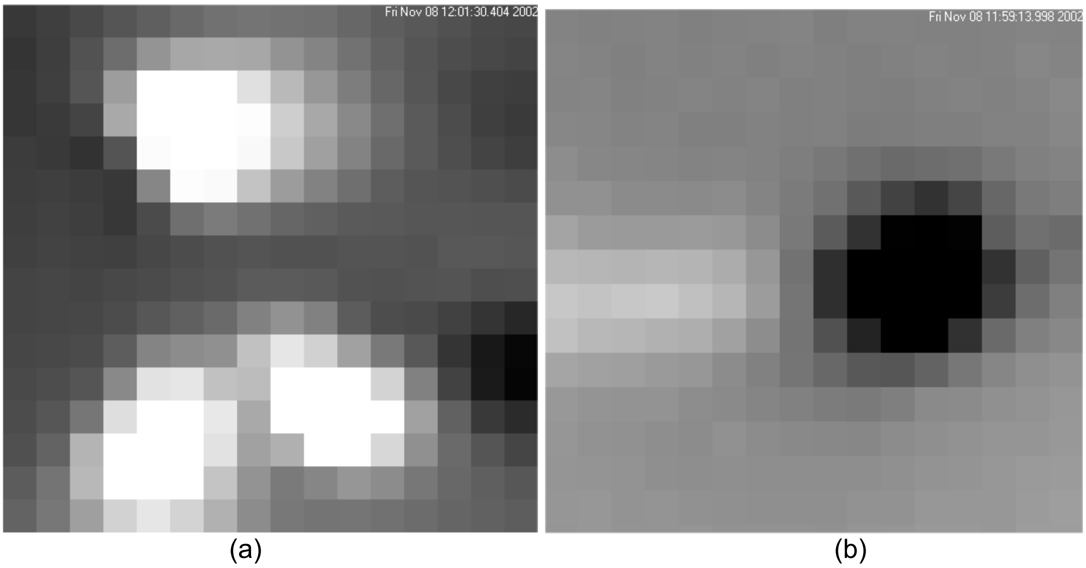
The processing of thermal image sequences for scene interpretation is relatively new. Several groups have proposed system for on vehicle pedestrian detection [MIT]. It is reported that techniques for the detection and tracking of people from thermal images are distinctly different from those employed in visible video sequences and that the range of radiation intensities from human body is limited in range, but clothing can totally prevent the detection of body heat. Recently, due to technical advances in materials and electronics, lower cost thermal imagers have become commercially available. Previous research by our group has contributed to these developments Weller [1999]. These primarily are designed for static monitoring of hazardous environments like industrial plant. These imagers have a lower spatial resolution than the high-end cameras. They are good people detectors, but provide poor context information.

We have developed two types of bi-spectral pedestrian and tracking systems one which uses a very low cost infra red sensor (16 x 16) array and integrated signal processing specifically developed to detect people manufactured by IRISYS and the second which uses a higher resolution FLIR thermal camera (256 x 256) to segment the image into regions that are likely to be people and a region which is background.

People Detection and Tracking using Irisys low-resolution infrared detectors

The detectors used in this study have been developed by Irisys, Stogdale [2003], as an attempt to provide a relatively low cost infrared array detection device (the detectors described on the company web site are actually a later model than the ones used here). The detector is based on an array of pyroelectric ceramic detection elements. The array is a square format, with 16 rows and 16 columns of pixels. This is a relatively small number of elements, resulting in a low-resolution image (figure 1). However, the cost of the detector is of the order of 1/20th of that of a traditional IR array.

At this cost, these low-resolution infrared video cameras provide an attractive potential for identifying and tracking movement. Their potential applications include observations for retail and marketing, behavioural research and security



(a) (b)

Figure 1. In image (a), three pedestrians are moving from right to left. This is the normal situation, where pedestrians are hotter than the background. In (b), a pedestrian has just entered an indoor environment and is moving left to right. Taken in Edinburgh in November, the outer layers of the pedestrian are colder than ambient, and they show as a dark mass. Note the faint warm 'wake'. This is caused by the detector being sensitive to changes in temperature. In this case the change from cold pedestrian to warm ambient produces a temperature increase, i.e. an apparent warm wake.

Measuring Trajectories

The Irisys detectors contain an on-board processor that matches an ellipse to the detected target. This is then used to provide position and velocity information. This assumes that the detector is mounted in a standard configuration, at a height of 3-4 metres, looking vertically down on a scene covering 3-4 metres square. Although the detectors have a relatively low resolution, the use of an algorithm to fit ellipses round targets means that the average centroid of the target is located with sub-pixel accuracy. With a field of view of 3 metres, each pixel spans approximately 19 cms, and the location of the target is known to an accuracy of the order of 10 cms. A typical recording is illustrated in figure 2.

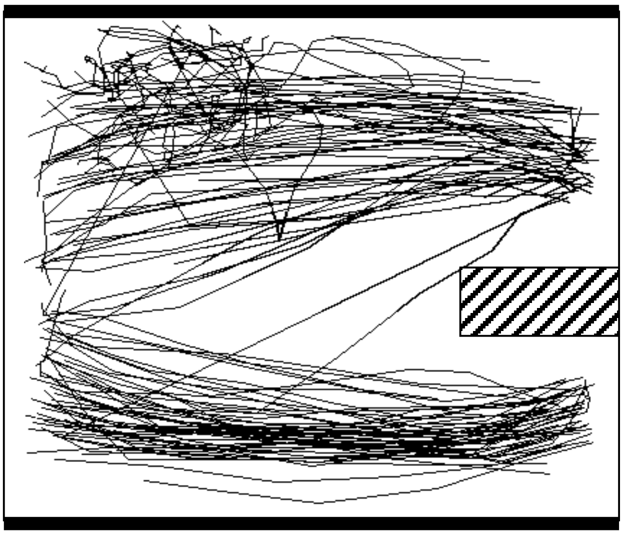


Figure 2. Flow of pedestrian around an obstruction in a corridor. Note the appearance of a 'loitering' pedestrian in the upper left quadrant.

Some limitations of the Irisys detectors were reported in Armitage [2003A]. In particular, problems were caused by the low frame rate that led to errors when tracking pedestrians as the moved between the fields of view of different detectors. A typical experiment (reported in more detail in Armitage [2003B]) involved one or two groups of pedestrians moving along a corridor round an obstruction. Three infrared detectors were used to record the experiment (as well as an ordinary video camera, used for calibration). The infrared detectors were arranged along the centre of a corridor of width 2.5 metres, and length 10 metres (approximately). The aim of the experiment was to see if trajectories could be accurately continued from one detector to the next. To be able to track correctly, the pedestrians have to be tracked as they move from the first to the second, then to the third detectors, in other words across two boundaries. The results are summarised in table 1.

Move	**Pedestrian Movements**	**Time secs**	**Flow Rate People/m/min**	**Average Speed (m/sec)**	**Fruin LoS**	**% Total Matches**
1	14	41	15.29	1.46	A	93
2	14	19	32.99	1.23	C	93
	14	22	28.49	1.33	C	79
3	14	18	34.83	1.36	B	93
4	28	20	62.69	1.31	D	79

Table 1: Summary of results for constricted flow round an obstacle.

As the experiment was done under controlled conditions, it was possible to relate the flow rates to levels of service as described in Fruin [1971]. At low levels of service (A, B), the algorithm used to track between detectors performed well. However, as the flow rate increased, the proportion of accurate matches decreased to an unacceptable level. The main problem was that the detectors had a low frame rate (3 frames per second), and at high levels of service, several people can cross between detectors in the inter-frame gaps, resulting in some people not being correctly tracked. The Mark 2 detectors (described later), have a much higher frame rate, and are far less prone to these sorts of errors.

In another experiment, Detectors were mounted under an archway leading to the main entrance of one of the University's campuses. Data was recorded for the first week (Monday morning to Monday morning) of the 2003-2004 session. Approximately 17,500 people entered the campus during this time: figure 3. The pattern of daily flow can be clearly seen, as well as the reduction in numbers at weekends.

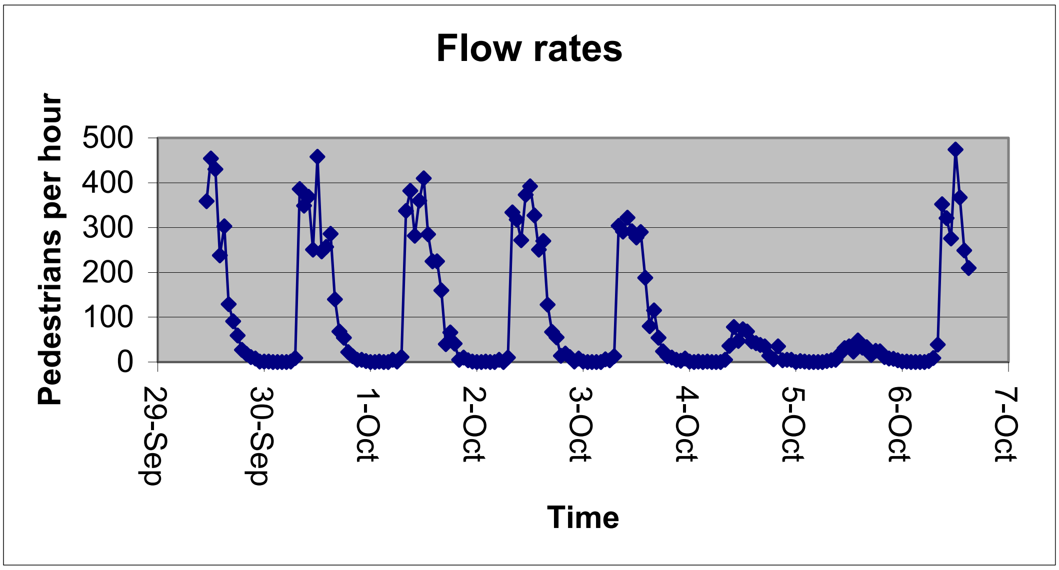


Fig 3: Pedestrian flow into the Merchiston campus for the first week of the 2003-2004 academic year.

Development of the software

Early measurements were done by recording to disk, and then analysing the recordings off-line. This is satisfactory for many applications, but is not adequate for applications (particularly related to security) that require real-time display. The software that tracks pedestrians across multiple detectors has therefore been re-written. In order to handle multiple detectors, perhaps connected to multiple PCs, the software uses the Java Communicating Sequential Processes (JCSP) model. This has enabled us to extend the experiment to areas covered by square arrays of detectors, and not just linear arrays. This has been made possible, in part, by the superior data available from the second generation Irisys detectors.

The IRISYS MK2 people counter

Current investigations involve the IRISYS MK2 people counter, which has a higher frame rate than the earlier counters. These have two output modes, which may be used as alternatives or in parallel. The first is straight infrared. This provides a view of the 256 pixels with some artefacts, such as the appearance of a cold wake following a moving person. The second output mode is derived from a signal-processing microprocessor within the device and places an ellipse around each identified target.

The signal-processing algorithm is accurate most of the time, but is occasionally deceived. This pre-processing option provides a welcome front end for many applications. However, a need for deeper analysis may necessitate a return to the raw camera output mode.

Current work includes modelling the anomalies and artefacts in the raw sensor data.

- The entry of a warm body into the field of view causes a background increase in all pixel values.
- There is a dip in pixel value shortly before a person moves into that pixel.
- A stationary person's image fades over 10 seconds and is eventually lost.
- There is an artefact that shows a cold trail behind a moving person.

Further experiments involved the extraction of stereo data from two ceiling-mounted detectors. A measure of height variation was calculated by subtracting the estimated positions of the target as relayed by the sensors:

$$h = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Where x_1, y_1 and x_2 y_2 are the estimated positions of the target on sensor matrices 1 and 2.

The variable h is a measure relating to the height of the target's head.

The resulting data information, although primitive, allowed differentiation between walkers and non-walkers. In future work, "non-walker" could equate, for example, to "cyclist".

In theory, accurate gait information could in future allow the identification and tracking of known individuals. We have not realised that potential yet. Limitations include our uncertainty about the in-built signal-processing algorithm, which so usefully identified people, and the limited resolution of the sensors themselves.

Other investigations include the use of IRISYS Infrared Systems in combination with conventional visual wavelength video systems in order to enhance target identification and background subtraction. A target that shows on an IR display is highly likely to be mobile. Further, conventional sources of target/background ambiguity in visual wavelength displays (trees in wind, televisions) do not appear as targets at IR wavelengths.

People Detection and Tracking using combined Thermal and Visible imaging.

Whereas the system described above operates indoors with a low resolution infrared sensor array looking down on the scene, the higher resolution bi-spectral imaging system can be used outdoors in a variety of situations. The objective of this work is to significantly reduce the processing time and power required for pedestrian detection and tracking in the visible by fusing the thermal image data with a visible image of the same scene such that real-time detection and tracking is possible using an embedded system. An operational flow diagram for a typical image processing system for people detection is shown in Figure 4 below.

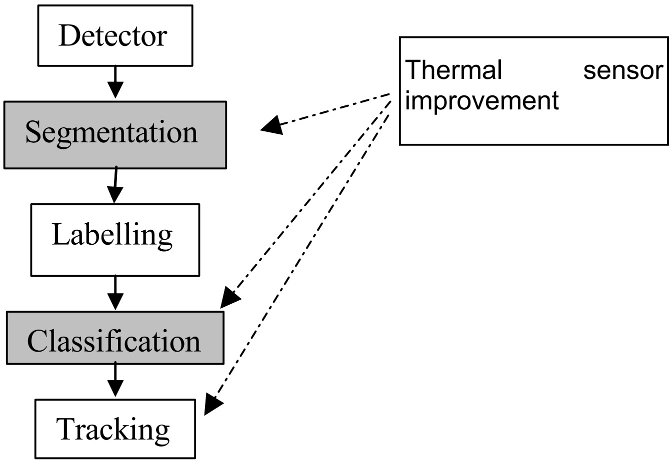


Figure 4: Shows the main processes involved pedestrian detection by image analysis.

Our work has shown that fusing infrared data with the visible image can reduce the computational power and memory operations required for the key stages of segmentation and of object classification.

Segmentation

In all pedestrian detection systems based on imaging, the first fundamental problem encountered is the extraction of the image regions corresponding to a person or persons. Previous attempts at segmenting people from a background have taken one of two approaches. Commonly background subtraction is used where the issue is deciding what is and what is not background. Most methods use intensity, texture and contrast properties in the image over a period of time to construct a background model. The background model is updated by averaging frames over a period of time to account for slow changes in illumination. The background model is the subtracted on a pixel-by-pixel basis from the current image. After noise removal, the remaining pixels are then held to be foreground or regions of interest. The second type of approach is based upon image motion, making the assumption that the background is stationary, or at most slowly varying, and that the person is moving. If the person stops moving for a period of time he/she will be interpreted as background. This method also requires constant or slowly varying geometry, reflectance, and illumination. Our bi-spectral approach adopts a combination of these two methods. The intensity levels of a pedestrian in the thermal image are significantly higher than most backgrounds. A simple adaptive threshold algorithm serves to isolate hot spots in the thermal image. Areas that are not hot are mapped onto the visible image and constitute our initial background image. Even in the thermal image the hot areas vary on a longer timescale. As the background slowly heats up and cools down during the day, the pixel values change. Thermal detectors also have a slow dc drift over time, and there will be new objects added to the scene, such as parked cars that produce a temperature difference between the background image. We use three support maps for the background updating process. Not all hot objects are pedestrians and at this stage we eliminate small clusters of hot objects that are unlikely to be people and classify the area as background. Figure 5 below shows a thermal image of a street scene in which moving hot spots are detected by our active threshold algorithm and small thermal hotspots are removed. Static hot spots such as the parked car are not detected.

Fig 5: Detection of moving hotspots.

Classification

Prior to classification we have groups of hot objects, some of which may be people. We use a straightforward shape analysis in the thermal representation, which is robust to range and occlusion. Vertical histogram projection is used to get the shape of an object, we find that pedestrians in an urban environment have a shape, similar to a normal Gaussian curve, which differs from cars, busses, and other hot objects. This means that we get a simple template model to work with. To reduce the noisy edges given in figure 5, a reduction of the spatial resolution is done by resampling the object down to a level that depends on the size of object found and the size of the frame. The width and height values give the Gaussian constants for the curve width σ, and the curve height.

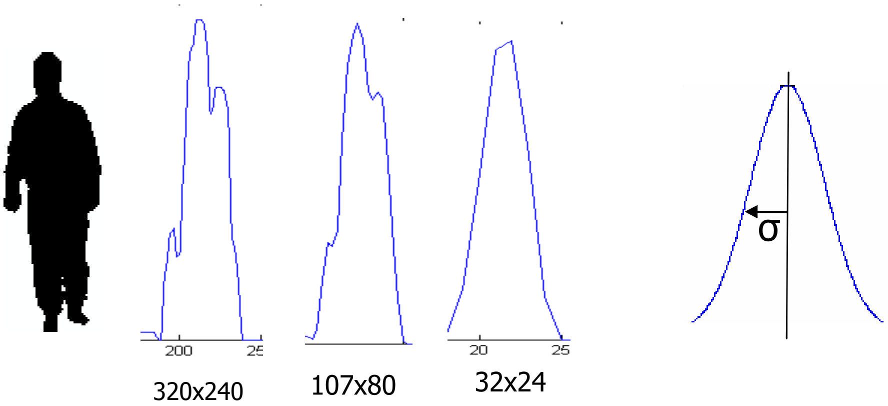


Fig 6: histogram projection at reduced resolutions

Figure 7 below shows a thermal image frame of an active street scene where the pedestrians have been detected by our software. The algorithm includes noise removal and cluster shape analysis to eliminate vehicles and other hot spots. Note that the car and its hot exhaust are now not identified.

Fig 7: Successful classification of pedestrians

Conclusion

Accurately tracking the movement of pedestrians in crowded and cluttered urban scenes is not a trivial exercise. For detailed studies, where pedestrian movements need to be correlated with factors such as the pedestrian's gender or age, we are still limited to manually analysing video recordings.

In this paper, we have looked at a number of possible methods for automating this procedure. Video image processing, low- and high- resolution thermal imaging, and combinations of visible and thermal imaging have been investigated, with encouraging results. In particular, we can use low-cost, low-resolution thermal detectors to monitor pedestrians over areas of a few tens of square metres. Also, we can use high-resolution thermal imagers to guide the process of segmentation in order to isolate pedestrians within complicated visible images.

Each technology has its own strengths and weaknesses. For instance, the low-resolution thermal imagers have simple, easily interpreted images. However, the spatial coverage of these imagers is limited. Consequently, there is still no universal solution to the problem of accurately tracking pedestrians.

Armitage, A., Binnie, T. D., Kerridge J. & Lei, L., (2003A) “Measuring pedestrian trajectories using a pyroelectric differential infrared detector”, Proc. of the twelfth conference on Sensors and their Applications, Limerick, Ireland, ISBN 0750309784

Armitage, A., Binnie, T.D., Kerridge, J.M., & Lei,L., (2003B), “Measuring Pedestrian Trajectories with Low Cost Infrared Detectors: Preliminary Results”, Pedestrian Evacuation and Dynamics – 2003, Galea, E.R. (ed), University of Greenwich, London, UK.

Fang, Y., Yamada, K., Ninomiya, Y., Horn, B. and Masaki, I., (2003A), “A Shape-Independent-Method for Pedestrian Detection with Far Infrared-images.”, Submitted to Special Issue on "In-Vehicle Computer Vision Systems" of the IEEE Transactions on Vehicular Technology.

Fang, Y., Yamada, K., Ninomiya, Y., Horn, B. and Masaki, I., (2003B), “Comparison between Infrared-image-based and Visible-image-based Approaches for Pedestrian Detection.” IEEE Intelligent Vehicles Symposium IV2003, pp.505-510.

Fruin, J.J. (1971), Pedestrian Planning and Design, Metropolitan Association of Urban Designers and Environmental Planners, New York.

Haritaoglu, I,. Harwood, D. & Davis, L. S., (2000) “W4: Real-Time Surveillance of People and Their Activities”, PAMI Vol. 22(8), pp. 809-830

Kerridge, J., Hine, J. and Wigan, M. (2001). Agent based modelling of pedestrian movements:. Environment and Planning B: Planning and Design, 28, 327-341.

Kukla, R., Kerridge, J. and Willis, A. (2003). Application of Context Mediated Behaviour to a multi-agent pedestrian flow model (PEDFLOW). Transportation Research Board Annual Meeting, Washington, 82nd Annual Meeting, .

Reading, I., “Pedestrian detection by computer vision”, (1999), PhD thesis, Napier University, Edinburgh.

Stogdale, N., Hollock, S., Johnson, N. & Sumpter, N. (2003), “Array based infrared detection: an enabling technology for people counting, sensing, tracking and intelligent detection”, SPIE, USE 3, 5071-94

Weller, H., Setiadi, D., Armitage, A. and Binnie, D. (1999). *Monolithic CMOS pyroelectric focal plane array* September. Cardiff, GB: by IoP Publishing, Bristol. ISBN 0750306629

Willis, A., Kukla, R., Kerridge, J. & Hine, J. (2001), “Laying the foundations: The use of video footage to explore pedestrian dynamics in PEDFLOW”, Proceedings of Pedestrian and Evacuation Dynamics, Shreckenberg,M. et al (eds), Springer, Berlin, pp 181-186.