
A Comparison of Video and Infrared Based Tracking of Pedestrian Movements

J. Kerridge, R. Kukla, A. Willis, A. Armitage, D. Binnie, and L. Lei

Napier University
Edinburgh Scotland – UK
{initial.surname}@napier.ac.uk

Summery. We provide a comparison of the relative merits of video and infrared based methods for collecting pedestrian movements from the real world and also from experimental environments. We describe the underlying technological basis of both methods and the tools we have developed to help in collection and analysis of the data. The desire to collect such data is driven by the need of modellers and simulation packages to use base data that is founded in valid empirical evidence, rather than some form of inspired supposition, as is the case with many of the current systems. In addition to the collection of speeds we are also interested in understanding and quantifying the ranges of distances people deviate from a straight-on path when confronted by some obstruction in front of them.

1 Introduction

There are many groups who are building pedestrian simulators for a variety of applications, notably for example Exodus, Aseri, Evi, plus others described at PED 2003 [2]. In the main these groups are using little or no quantitative data concerning the movement patterns exhibited by pedestrians in their simulations. Some of them even make suppositions about the normal walking speed of pedestrians. The PEDFLOW project is perhaps unique in that it attempted to build a model of pedestrian movement based upon data about pedestrian movements that had been obtained from observation of the real world [7]. This data was extracted from CCTV video footage from various UK cities using a method that is described in the next section. The PERMEATE project [1] investigated the use of low-cost infrared detectors to capture pedestrian movement data. This technology provides the opportunity to collect pedestrian movement data in real-time. This is described in section 3. In section 4 we compare the two technologies from a number of different points of view.

2 Video Based Data Capture

Video analysis in the widest sense has been used to study the behaviour of pedestrians for a long time, as the video itself provides a permanent record that can be subjected to repeated, possibly different, kinds of analyses. However, before the advances in computer technology that allowed digital image processing, it was a laborious task observing pedestrians in reality and hence limited to counting people in an area, or crossing a datum line or measuring the time it takes for an individual to cover the distance between two datum lines.

In PEDFLOW we developed a methodology that uses digitised video footage and semiautomatic image tracking to greatly improve the accuracy and variety of data that can be extracted from video. It allows precise instant measures of position and time (and hence speed), but most importantly a way to identify points of interest (such as the start of a deviation) as a change in an instant measurement. By having well-defined criteria that characterise such a change, the extracted data becomes independent of the human factor and therefore repeatable. This data combined with information about the scenario where it was collected and data about the observed pedestrians such as, age and gender, can then be used as input to a pedestrian model.

2.1 Filming and Digitising

The first step is to obtain raw video data. Ideally this footage should be recorded with the camera pointing down vertically (90 degrees) as described, for example, by Hoogendoorn [3]. However, since this is difficult to achieve, an angle up to 45 degrees is still acceptable. Depending on the selected survey location, video footage of pedestrians was collected using either a standard hand-held digital camcorder (miniDV) mounted on a tripod, or by CCTV cameras operated by collaborators within local City Councils. We also contracted private survey companies to film areas that cannot be accessed using either of these methods. In all cases it is imperative that the filmed pedestrians are unaware of the fact that they are under observation, so their behaviour is natural.

If a digital camcorder was used the data can be transferred into the computer directly by means of a firewire connection. It is then stripped of its audio component and recompressed using an MPEG4 codec to save space. Also it is converted to an avi file as required by the tracking software. The method results in very good quality footage of 720x576 pixel resolution (CIF) and 25 fps frame rate (PAL standard). The conditions are not as good for the (more easily obtainable) analogue video material. In order to reliably capture it without dropped frames (which would result in unreliable time measurements) and taking into account the generally poorer image quality, the footage was digitised with a quarter PAL resolution of 352x288 pixels (QCIF) at 25fps.

The large and unwieldy files are then split into short clips according to pre-defined, objective sampling criteria that characterise a scenario of interest. If, for example, we were interested in investigating the deviation distance associated with an item of street furniture (such as a lamp post), all instances in which a pedestrian walks within a certain distance of the object would be clipped for inclusion in the sample. The ‘clipping’ reduces the amount of footage that will need to be tracked and it also simplifies the workflow. The clips are time stamped so that the global time distribution is available.

2.2 Tracking

Tracking is the semi-automatic extraction of pedestrian co-ordinates from every individual video frame over a period of time (the length of the clip). We use a commercial package from Mikromak [10] that was originally developed for use in the medical environment but proved to be flexible enough for our purpose. It has the advantage that it can transform the tracked pixel

position to real-world co-ordinates directly, provided it is calibrated correctly.

Calibration of the software requires the exact position of 20 objects/points that are visible in the frame. The corners of paving slabs are suitable, but we also used road markings, street furniture and even chalk markers depending on the situation. We developed a method to efficiently record these positions by measuring the distance of each point to two fixed co-ordinates on a (virtual) base line. Using a spreadsheet, these distances are then converted into co-ordinates using trigonometric functions. In the software package the same 20 points are identified in the video frame and associated with their real-world co-ordinates. Technical data of the recording equipment (such as CCD size and resolution) also needs to be specified. The program then uses an interactive, iterative process to build a transformation model between the two co-ordinate systems.

Image matching does the actual tracking: an area marked in Frame A is matched to a similar area in Frame B within a specified area surrounding the original position of that patch. The new position and pattern is recorded and used as the basis for the matching in frame C and so on. Selecting the head and upper body was found to be most effective – the position on the ground is obtained by adding an offset. This works so well that human intervention is only required if lighting conditions change (shadow) or the person is occluded. For movements towards or away from the camera the offset needs to be corrected as well. The software can track several people simultaneously.

The precision is so great that the software also records the sideways movement of the body caused by the shifting of weight between the two feet while making a step, which results in a wavy line superimposed on the underlying trajectory (see Fig. 1 where the pedestrian is walking in the y-direction to emphasize the sideways movement). This higher frequency can be filtered out using the 'smooth' function of the tracking software, which averages between neighbouring points. This has proved to be more effective than the polygon approximation, which is also available.

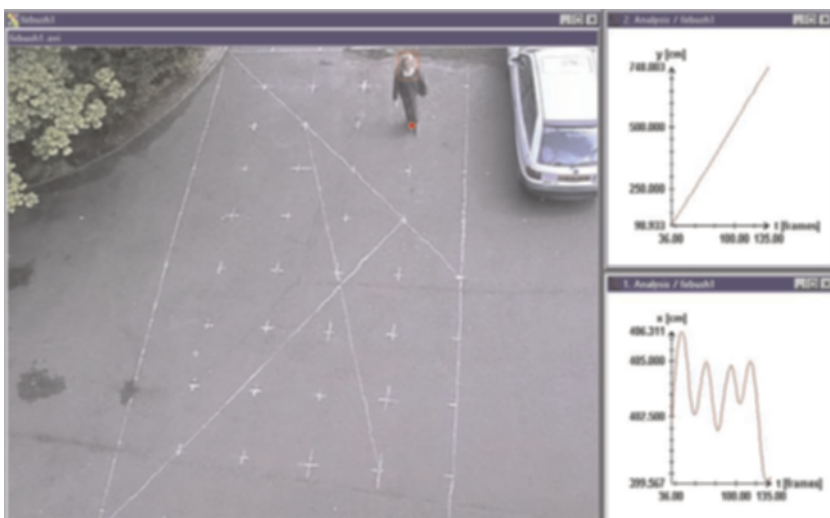


Fig. 1.

The result of the tracking process is a comma-separated file for every tracked person. A file of the same structure is also generated to capture the static mark up of the scenario such as building edges and street furniture as polygons

2.3 Data Exploration and Extraction

Having the raw numerical data is only the starting point of the analysis. Relevant configurations need to be identified with regards to time (frame) and place (co-ordinates) and the information recorded.

To help with this task, a software tool has been developed (Fig. 2). It will load all the data files related to one clip and draw the trajectories in an x-y co-ordinate system. A slider at the bottom of the window is used to adjust the time. The program displays the position of the pedestrian(s) at the selected time as a circle. By moving the slider, one can explore the data dynamically and find areas of interest such as two pedestrians who are about to avoid each other.

The program supports the following additional features to aid the exploration; a) online speed visualization b) online direction visualization and c) visualization of local maxima in direction change. The speed distribution over the time period is displayed at the bottom of the tool with a line indicating the current frame/time. It can be used to quickly identify maxima and minima of walking speed. The current walking direction is drawn as a line from the centre of the circle for every pedestrian in the tool. The black dots help to identify the points at which the person changed direction. Since this is usually not instantaneous, the tool defines one point of maximum change as being the representative. These points of greatest change are calculated by

taking the second derivative of the direction with parameters that can be adjusted by further sliders.

Having found a configuration of interest, the program greatly simplifies the collection of associated measures. It uses two markers, which are set by mouse clicking on the area. The software immediately updates frame, x co-ordinate and y co-ordinate. It also calculates and displays the distance between the markers and (if they are on different frames) the time-difference between them. If the markers are associated with the same person, average speed is presented as well. At the press of a button this information, together with the filename, a label and a comment is stored in a file. After analysing several clips, the collected data can easily be imported into a spreadsheet or statistical package for further

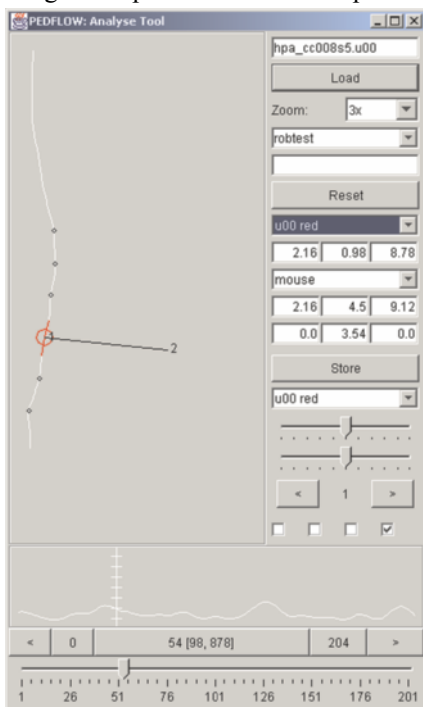


Fig. 2.

analysis [8], which is outside the scope of this paper.

3 Infrared Based Data Capture

The Irisys detector [5,6,9] is primarily designed for counting people moving back and forth across a datum line specified by the user. In this mode a single detector can be used, for instance, to count movements along a corridor or through a doorway. Multiple detectors can be aligned to create a larger counting line. In this mode detectors have been used successfully: to count pedestrian movements in and out of supermarkets, in station concourses, and over the Millennium footbridge in London during trials after the bridge was modified. We have been able to extract more information from the detectors, in particular to record pedestrian trajectories [4].

The detector uses a 16 x 16 array of pyroelectric ceramic detectors to measure changes in temperature. The detectors we have been using only measure temperature differences, and rely on the pedestrian being at a different temperature from the background. This has the advantage that the background disappears from the image, leaving pedestrians as clear targets. In Figure 3 the white areas indicate a person, with the darker areas indicating the background.

The normal mode of using the detector is to mount it at a height of three to four metres vertically above the region of interest. At this height, the detector covers a ground area three to four metres square. The detector has a processor on-board that undertakes image pre-processing. Likely targets have an ellipse fitted to them, and the center of the ellipse is then calculated and communicated down a communications link to the data-gathering computer. The X and Y positions are recorded approximately three times per second. These

values are presented as a floating-point number in the range 0 – 16 for each axis. One unit, being the width of a pixel, corresponds approximately to 20 centimeters on the ground, when mounted 3 meters above the ground and 25 centimeters when mounted at 4 meters. However the centering process effectively averages data from a number of pixels, the position is actually recorded with sub-pixel accuracy; hence we can locate pedestrians to an

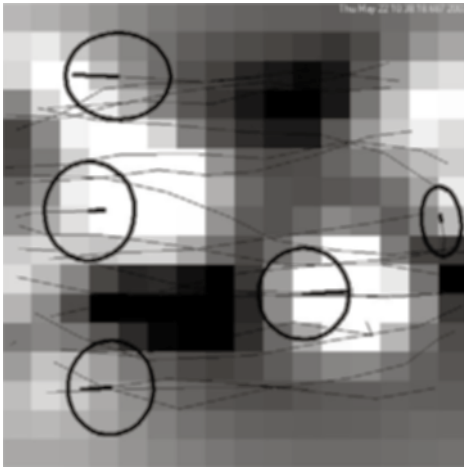


Fig. 3.

accuracy of 2.5 cm or better. Fig. 3 shows the ellipses that have been matched and the trajectories of these targets and other previous targets. The line within the ellipse shows the direction of travel. Due to the time it takes to process the image the fitted ellipse always lags behind the actual image. Fig. 3 also demonstrates that it is impossible to identify an individual from the image.

3.1 Multiple Detector Systems

The main problem with multiple detector systems is tracking the trajectory from one field of view to the next. We cannot guarantee the relative alignment of the detectors and thus the data analysis systems have to be able to overcome such a situation and cannot rely on a perfect alignment of the fields of view of adjacent detectors to a predefined specification. For each detector a set of data is captured in the central computer of the form shown in Fig. 4. This represents a single person moving through the field of view of a detector. A line starting with C gives a time stamp for the data reading with an accuracy of milliseconds, which is generated by the data gathering computer. A line starting with T gives data about the ellipse that has been fitted to a person by the detector's internal processor. If there were more than one person in the field of view then there would be as many T-type records for each C-type record. If no targets are detected a C-type record will be immediately followed by another. Thus the person in Fig. 4 starts from a position of [0.416, 4.789] at time Thu May 22 10:36:41 531 2003 and leaves the field of view at location [13.064, 3.179] at time Thu May 22 10:36:43 000 2003, if we ignore data points with S = 1. A simple calculation, assuming each pixel is 0.2m and we ignore movement in the y direction, which is just over 1 pixel yields a walking speed of 1.72 m/sec, from $(13.064 - 0.416) \cdot 0.2 / (43.000 - 41.531)$

Timestamp	Target Data				
	ID	S	X	Y	Mode
C Thu May 22 10:36:40.906 2003					
C Thu May 22 10:36:41.218 2003	T	2	1	0.304	5.089 257
C Thu May 22 10:36:41.531 2003	T	2	0	0.416	4.789 261
C Thu May 22 10:36:41.859 2003	T	2	0	0.776	3.303 261
C Thu May 22 10:36:42.171 2003	T	2	0	4.256	2.860 261
C Thu May 22 10:36:42.375 2003	T	2	0	7.636	3.020 261
C Thu May 22 10:36:42.687 2003	T	2	0	10.831	2.749 261
C Thu May 22 10:36:43.000 2003	T	2	0	13.064	3.179 261.
where					
ID - Target Identifier	S - Status; 1 - invalid; 0 - valid				
X - x co-ordinate	Y - y co-ordinate				
Mode - Further data when S = 1					

Fig. 4.

The detector's internal processor operates as follows. The identification number is unique to a detector. The sequence of identification numbers is reused from the initial value once the detector has been unable to detect any targets in the field of view for some time. This means that the lower identification numbers are reused many times during an observation, hence we have to undertake additional processing between the streams of data if we are to track people across multiple detectors.

3.2 Algorithmic Approach

For the purposes of explanation we shall consider the scenario shown in Fig. 5. Consider a corridor surveyed by three detectors such that people can enter

through the left-hand edge of Detector 0 and then exit from the right-hand edge of Detector 2 and vice-versa. We have built a system in Java in which each detector is managed by its own process. Thus as a person is detected as having left the field of view, of say, Detector 0 their trajectory details are passed to Detector 1. Once the person has been detected as having arrived in Detector 1 then the process dealing with Detector 1 can search incoming trajectories from Detector 0 to find the best match. In this manner we can build up the complete trajectory of a person as they move from an entry to an exit edge. The system is capable of processing the incoming data and consequent saving of trajectory information for each person in real time.

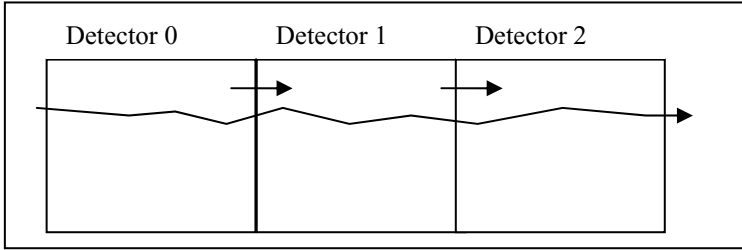


Fig. 5.

Figure 6 shows part of the printed output from the system for one person as they moved from Detector 0 to Detector 2. Each person is given an identifier derived from the edge of the detector by which they entered the observed area together with a unique sequence number. The Entry and Exit points with corresponding times are stored so as to make it easier to calculate average speed through the observed area. The remainder of the person record comprises the sequence of points through which the person moved in passing through the observed area. The location data comprises; [Detector Number, Identifier Tag, X-value, Y-value, Status followed by the time the person was at that point]. By observation it can be seen that in Detector 0 (Locations 0 to 6) the person was given the identifier tag 12, then in Detector 1 (Locations 7 to 11) the tag 2 and finally in Detector 2 (Locations 12 to 19) the tag 10. This highlights the difficulty in forming the complete trajectory.

A relatively simple, though not perfect, heuristic has been used to determine a match between an incoming partial trajectory and one that is being generated in an adjacent detector. It deals with the cases where detector fields of view both overlap and are disjoint. The amount of under and overlap is always kept as small as possible when setting up the detectors. Consider a person moving from Detector 0 to Detector 1 of Fig. 5; then the exit time from Detector 0 must be within 1.5 seconds of the entry time to Detector 1. However depending on whether there is underlap or overlap we may have started to build the trajectory in Detector 1 before the data from Detector 0 arrives. This time differences also takes account of the fact that edge effects in detector image processing may mean there is some time before a valid target is returned in Detector 1. If we assume that the base edges of the detectors are more or less in line, that is the edges of the detectors abut almost perfectly in the y-direction, then we also require that the y-pixel value at the point of exit from Detector 0 is within 2 pixels of that observed on entry to Detector 1. This is probably reasonable for people walking down a corridor

as in this case but is less likely to be true if we are observing a larger rectangular area where people may move with more freedom.

```
[PersonId, 0, 16, 11]
Entry point PointData - [ Detector 0, x 0.833, y 9.925]
Exit point PointData - [ Detector 2, x 14.901, y 11.065]
Entry time [Time, 10:37:28.220 ]
Exit time [Time, 10:37:35.480 ]
Location 0, [Location, 0, 12, 0.833, 9.925, 0, [Time, 10:37:28.220 ]]
Location 1, [Location, 0, 12, 2.58, 10.458, 0, [Time, 10:37:28.550 ]]
Location 2, [Location, 0, 12, 4.768, 10.734, 0, [Time, 10:37:28.880 ]]
Location 3, [Location, 0, 12, 7.824, 11.004, 0, [Time, 10:37:29.210 ]]
Location 4, [Location, 0, 12, 10.87, 11.069, 0, [Time, 10:37:29.540 ]]
Location 5, [Location, 0, 12, 13.151, 10.798, 0, [Time, 10:37:29.870 ]]
Location 6, [Location, 0, 12, 15.335, 10.472, 0, [Time, 10:37:30.200 ]]
Location 7, [Location, 1, 2, 4.058, 11.696, 0, [Time, 10:37:31.190 ]]
Location 8, [Location, 1, 2, 7.195, 11.673, 0, [Time, 10:37:31.520 ]]
Location 9, [Location, 1, 2, 10.152, 11.55, 0, [Time, 10:37:31.850 ]]
Location 10, [Location, 1, 2, 12.481, 11.332, 0, [Time, 10:37:32.180 ]]
Location 11, [Location, 1, 2, 14.212, 10.979, 0, [Time, 10:37:32.510 ]]
Location 12, [Location, 2, 10, 0.8, 10.74, 0, [Time, 10:37:33.170 ]]
Location 13, [Location, 2, 10, 2.011, 11.596, 0, [Time, 10:37:33.500 ]]
Location 14, [Location, 2, 10, 4.287, 11.955, 0, [Time, 10:37:33.830 ]]
Location 15, [Location, 2, 10, 7.096, 12.015, 0, [Time, 10:37:34.160 ]]
Location 16, [Location, 2, 10, 9.964, 11.902, 0, [Time, 10:37:34.490 ]]
Location 17, [Location, 2, 10, 11.896, 11.682, 0, [Time, 10:37:34.820 ]]
Location 18, [Location, 2, 10, 13.472, 11.337, 0, [Time, 10:37:35.150 ]]
Location 19, [Location, 2, 10, 14.901, 11.065, 0, [Time, 10:37:35.480 ]]
```

Fig. 6.

Once the pixel locations have been transformed into actual offsets in the observed area we have data from that shown in Fig. 6 in the correct format for analysis by the tool developed for the video based data capture method. Hence we can process the data to determine deviation points and distances.

4 Comparison of Methods

Video based measurement has the obvious advantage that a human can immediately interpret the image, extracting information about the scene such as the layout, type of any obstruction, the weather and possibly the gender and any mobility impairment a person may have. It utilizes consumer-based products and thus is cheap to buy. Its main disadvantages are that processing the data is expensive either in terms of human time required to process the images (typically, 1 hour of video requires 1 week to analyze) or automatic processing requires very expensive image processing equipment. It may also cause data protection and privacy issues.

Conversely, the infrared detection equipment uses non-consumer technology and thus is more expensive but does allow the capture of trajectory information in real-time due to the incorporation of an internal processor in the detector. It is not possible to obtain mobility, gender and other aspects directly but the method does not have the data protection and privacy pitfalls of video images.

The best compromise is probably a combination of both techniques whereby trajectories are captured using infrared technology and the scene is observed by a web-cam, which allows an observer to mark points of interest in the data stream as it is collected.

Acknowledgements

The research has been supported by a grant from the UK Department of Transport in the LINK programme Future Integrated Transport with the project PERMEATE (GR/N33706). It has also been supported by a research grant from the UK Engineering and Physical Science Research Council for the project PEDFLOW (GR/M59792) for the video based data collection.

References

1. Armitage, A., Binnie, T.D., Kerridge, J.M., & Lei, L., (2003), "Measuring Pedestrian Trajectories with Low Cost Infrared Detectors: Preliminary Results", Pedestrian Evacuation and Dynamics – 2003, ER Galea (ed) University of Greenwich, London, UK.
2. Galea, E.R. (ed), "Pedestrian Evacuation and Dynamics 2003", University of Greenwich, London UK.
3. Hoogendoorn, S.P., Daamen, W. and Bovy, P.H.L., "Extracting Microscopic Pedestrian Characteristics from Video data", Transportation Research Board 2003 Annual Meeting, CD-ROM, Paper No 477, National Academies, Washington USA, 2003.
4. Kerridge, J.M., et al, Monitoring the Movement of Pedestrians Using Low-cost Infrared Detectors: Initial Findings, to appear in Transportation Research Board 2004 Annual Meeting, Washington USA, 2004.
5. Mansi, M. V., Porter, S. G. Galloway, J. L. & Sumpter N., "Very low cost infrared array based detection and imaging systems" (SPIE) Aerosense 2001, Orlando, Florida USA, 17-19 April 2001.
6. 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.
7. 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, Schreckenberg, M. and Sharma, S.D. (eds), Springer, Berlin, pp 181-186.
8. Willis, A., Gjersoe, N., Harvard, C. Kerridge, J.M., and Kukla, R., "Human Movement Behaviour in Urban Spaces", submitted for publication Environment and Planning B.
9. www.irisys.co.uk (accessed 9/7/2003).
10. <http://www.mikromak.com/> (accessed 27/10/2003).