Height Measurement as a Session-based Biometric for People Matching Across Disjoint Camera Views

C. Madden and M. Piccardi

Faculty of Information Technology, University of Technology Sydney

Email: cmadden@it.uts.edu.au, massimo@it.uts.edu.au

Abstract

A surveillance session can be defined as the period between an individual's entrance into a surveillance system, and their exit. Utilising this framework enables personal features, such as clothing and height estimates, to become invariant for the duration of the surveillance session. These can then be used as session-based biometrics for matching an individual uniquely as they move throughout the surveillance system, which may include significantly disjoint cameras. By utilising a hierarchy of biometrics to match individuals one can improve the speed of the overall matching. This eliminates unlikely matches quickly by using biometrics that are fast to calculate first. This paper proposes height estimation as an example of such a session-based biometric that would be used earlier in the hierarchy. It demonstrates how statistical height analysis can be quickly calculated as one of the session-based biometric components used to match the tracks of the same individual through image sequences taken by two disjoint camera views in the same surveillance session.

Keywords: video surveillance, disjoint camera views, height detection, session-based biometric

1 Introduction

Computer vision-based object tracking can be based upon shape, motion, and appearance features [1]. People tracking is a more complicated subset of the tracking problem as humans are deformable objects, and can also be subjected to self occlusions. Human environments also tend to be less structured and more crowded than other environments, such as streets, and hence are more prone to mutual occlusions. This can make shape and appearance features difficult to exploit. Surveillance systems are also likely to consist of many camera views that are disjointed, often significantly, making object motion cues unavailable for parts of the path. Such scenarios are common because of the cost of acquiring enough equipment to provide full coverage, especially with high enough resolution to measure accurate biometric information.

This paper proposes an algorithm that uses 'session-based biometrics' for tracking single individuals during a surveillance session having disjoint camera views. Conventional biometrics used for identity assessment need to be invariant for the full duration of their use. Instead, our system is not interested in using conventional biometrics for identity assessment. We intend to use biometrics and other features to match a person who has been previously tracked within the surveillance system with their subsequent movements in the system. The notion of a surveillance session is introduced here as the segment of time from when an individual enters the building's surveillance system, moves around inside the surveillance area and then exits. Such a surveillance session will typically be a portion of one day as people enter to perform their work before leaving. It is important to note that the surveillance area does not need to be fully covered by surveillance cameras. The cameras need only to be placed at strategic points of interest. This notion of a surveillance session means that the current identified tracks only need to be compared with the tracks of people who are within the building or parts of it. This provides a practical limit to the amount of comparisons to be made without imposing an arbitrary limit on the number of people or tracks within the system.

A second major advantage of the surveillance session is the increased stability in some transient personal features that can now be used for track matching. Human biometrics are features that do not change for an individual over time and hence can be used for identification. Typical features used here may be face recognition [2], fingerprints [3], iris matching [4], and gait recognition [5], [6], [7]. A normal surveillance system, with low resolution cameras, will have difficulty in recognising these features as a person enters a cameras field of view. Other features, such as motion models are also commonly used, and are effective where cameras overlap [8], or are close together [9]. When the cameras are significantly separated, it becomes unlikely that a person will continue moving in the same direction at the same speed. This renders motion cues ineffective. Other features such as appearance, height, and build are likely to remain stable over the surveillance session. Whilst these features may not provide a personal identification of an individual, they can provide a set of stable session-based biometrics that can be used by low resolution cameras for matching an individual's motion as tracked around the building. Multiple biometric features are required as many features that can be detected at low resolution may not be unique for each individual. For example two people may be of similar height, but are unlikely to also be wearing the same clothes, or have the same build. Moreover, a multiple-feature approach promises to be more robust to errors occurring in the feature extraction phase.

Choosing the personal biometric features to be used is a crucial component of designing an accurate real-time surveillance system. The design is also evolving over time as advances in both computational power, and affordable camera technology impact upon the system design. Many potential features for the session-based surveillance system are currently being researched, such as gait analysis [5], [6], [7] and colour appearance modelling [10], but this paper will focus upon the use of height estimation as its example. Height estimation was investigated because though many people may be similar in height, it can provide a fast biometric to be used as a complement to the other features. Height estimation differs from gait analysis because it tries to filter the gait effects to provide a single measurement, rather than provide an analysis of the changes in the person as they move.

This paper is composed of four sections. Section 2 discusses the use of height estimation as a session biometric. Section 3 reports on the experimental results for tracking between disjoint track matching both within a single camera, and across two disjoint cameras. Section 4 concludes with a summary of the results of the current research.

2 Height Estimation as a Session-based Biometric

Height estimation is discussed here as a biometric to highlight the use of a feature that becomes stable in the session-based framework for video surveillance. Traditional methods of height estimation require an individual to stand up straight near a vertical surface with markings that can be used for measurement. This can produce accurate height estimations, but places restrictions on the individual for the accuracy to be achieved. As this would be impractical in a video surveillance setting because of the behaviour restrictions, alternative methods are required.

Criminisi *et al.* [11] demonstrated their single view metrology with an example of height estimation. This was taken from a single image of a person standing and seems to provide a reliable measurement, however it still requires a person standing up straight in the image. Video surveillance within buildings does not often achieve such images, but is predominated with people walking. Thus a useful height estimation from such motion could be a session-based biometric for matching tracks of people. This situation however introduces the gait effects of how the person is walking into any possible height estimations. Though the analysis of gait as a biometric has been studied for many uses including human classification, and human motion analysis [6], the authors are unaware of publications that have used it for height estimation in a real-time multiple object system. Many methods including Green and Guan's articulated human models [7] may improve height accuracy, but seem to be computationally expensive, especially for tracking multiple objects.

The method we have used determines the location of the top of the head in the image, and its projected location on the ground plane to estimate the height of the individual. Figure 1 shows the image feature points, after compensating for lens distortion, extracted from a sequence where a single person walks across Camera one's field of view.. The pixel coordinates are given in a top left coordinate system. This clearly shows the periodic motion of each heel throughout the gait. The heel positions when stable represent the foot being on the ground. By assuming that the head position on the ground is directly below the head top position, and lies in the midpoint between the heel locations, then an estimate of the ground point directly below the head can be determined. Note that currently this information is being manually determined, but analysis of the foot and heel locations could lead to effective automation of identifying both the ground and head locations.

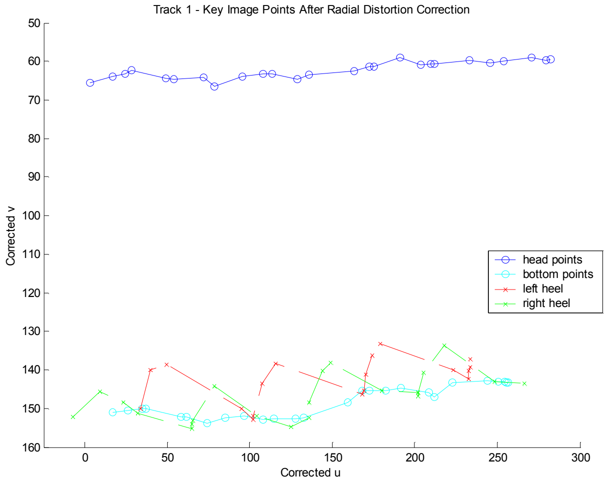

Figure 1: Image positions of key object features

With a two camera view of the scene a simple 3D position estimation of the head would be possible [12]; however this situation is uncommon in building surveillance systems. The cameras used for such a system are generally fixed cameras, which allow for camera calibration that will hold over the surveillance session. This process involves the matching of known points in the world coordinate system to image plane points, which defines the camera calibration matrix defined by (1) [13].

$$\begin{bmatrix} su_i \\ sv_i \\ s \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad \textbf{(1)}$$

Expanding this for u and v gives:

$$u = \frac{p_{11}X + p_{12}Y + p_{13}Z + p_{14}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}} \quad \textbf{(2)}$$

$$v = \frac{p_{21}X + p_{22}Y + p_{23}Z + p_{24}}{p_{31}X + p_{32}Y + p_{33}Z + p_{34}} \quad \textbf{(3)}$$

This calibration can include the estimation of the ground plane surface that the individuals are walking along. A pixel defines a ray in 3D space according to the intersection of (2) and (3) [12]. A 3D ground point in the world coordinate frame can be defined as the intersection of this ray and the ground plane. In practice a 3x3 homography can be determined to map pixel locations onto this ground plane [12]. Choosing the world coordinates so the ground plane is defined by the X and Y axis, with Z pointing vertically, simplifies (1) as $p_{13} = p_{23} = p_{23} = 0$ and $Z_i = 0$.

This simplifies (1) to a 3x3 matrix, which can be inverted to allow for simple mapping of the heads projected ground plane position, or indeed any other ground plane position, from the image plane into real world coordinates on the ground plane. Using the position of this point in the world coordinate system, the height of the individual can be estimated as the location where the ray defined by the top of the head in the image passes above the ground plane position. This is found by rearranging (2) to get (4), and using the ground plane X and Y location values. Similarly (3) could be rearranged to determine the height as a function of v, however (4) was found to be a more stable estimate of the height.

$$Z = \frac{(P_{11} - uP_{31})X + (P_{12} - uP_{32})Y + P_{14} - uP_{34}}{uP_{33} - P_{13}} \quad \textbf{(4)}$$

This resultant height over a series of frames will give a range of heights as the individual walks through the camera's field of view. From here simple statistics can determine the average sequence height and standard deviation. This statistical data can then be compared to other object tracks to determine matches.

The computational complexity of session-based biometrics is an important consideration. Identifying a unique person may be based on many features, and there may be many people in any camera view. Therefore it is important to note that some session-based biometrics are going to be faster to calculate, whilst others may provide a greater degree of discrimination between individual people. For this reason a hierarchical approach can be used to improve the speed of the overall matching. Those features which are faster to calculate and have a reasonable degree of discrimination, such as our height estimation method, can be called lightweight biometrics. These should be calculated earlier in the biometric hierarchy. This allows for the fast dismissal of many of the very different matches within the surveillance session. Heavyweight biometrics, such as colour comparisons [10] and model based gait analysis [7], could then be used to provide greater discrimination between those individuals that are considered to be potential matches after the height comparison. This reduces the usage of the relatively slow biometric comparisons between very different individuals.

3 Experimental Results and Analysis

The implementation of the height estimation was performed in Matlab® using the Camera Calibration Toolbox developed at California Institute of Technology. This provided the framework for determining the camera calibration using measured points in the real world coordinate system, whose position in the image was then manually estimated. The toolbox then calculated both the intrinsic and extrinsic parameters for each camera used.

Two cameras were used to verify the system. Camera one was situated in an open area where people tend to walk across the field of view from left to right. There are also stairs situated in the middle of the image, and a set of lifts out of view to the bottom of the camera. An image of the area is shown as Figure 2 below.

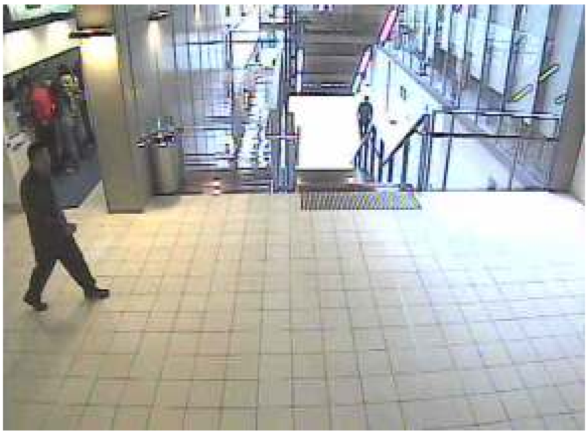

Figure 2: View obtained from Camera one

Camera two is situated in a corridor outside a set of lifts. This is in a separate area within the building where people tend to walk through the camera view from bottom to top, generating a perspective distortion in the size of their appearances. An image of this area is shown as Figure 3 below. These two cameras were chosen as they tend to provide the most difference in image appearances across each set of an individual's tracks. Where the tracks through camera one tend to remain a relatively constant image size, camera two adds the perspective distortion, making the size of people in the image change through the sequence. This comparison provides one of the most challenging set of circumstances for the matching of an individual's height.

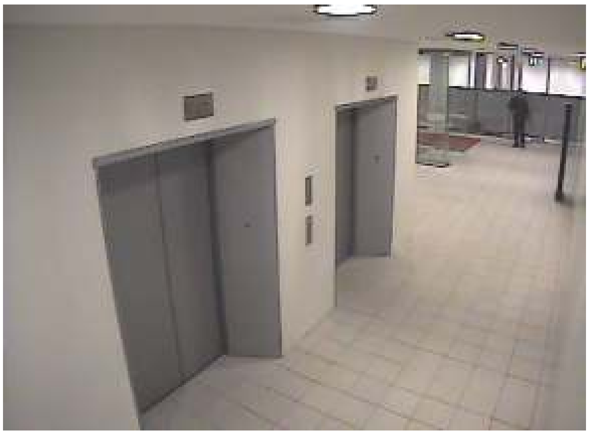

Figure 3: View obtained from Camera two

Five sequences of images were chosen from the recorded footage for analysis as separate tracks of three different individuals. These were processed manually to extract the head point, the estimated projection of that head point vertically down onto the ground plane, and the heel positions. An example of the points that were extracted for the track 1 sequence is shown in Figure 1 in section 2 above. Using the calibration matrices of both cameras, these were converted into height estimations over the sequence where that person can be seen. These were then statistically analysed to extract an average height value, and their variances for comparison.

Track 1 follows individual one as he walks through the view from camera one from the left edge to the right edge. Track 2 follows individual one as he walks through the view from camera one from the right edge to the left edge. Track 3 follows individual one as he walks through the view from camera two from the top through the bottom edge. Track 4 follows individual two as he walks from the bottom of the view from camera one to the stairs at the middle of the view. Track 5 follows individual three as he walks from the top of the view in camera two to the bottom edge.

Figure 4 below shows the ground plane positions of the individual's movement through the different camera views. This information is used as the XY positions for the height estimation. It is easy to see that though tracks one and two are quite similar for individual 1, the third track for the same individual is considerably different as he walks through the view of camera two.

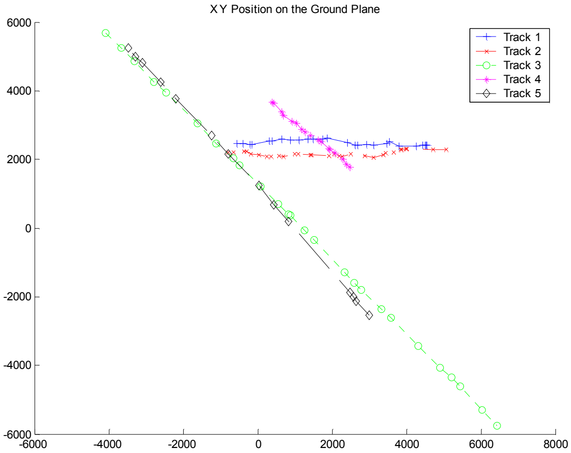

Figure 4: Ground plane coordinates of object tracks

The height estimations for the five tracks are shown in Figure 5 below. They clearly indicate distinct differences between the individuals measured. The statistical analysis for this data is shown in Table 1 below, and confirms that the tracks of individual one, Tracks 1, 2 and 3, match each other. The tracks are considered to match if the difference between the means of the two tracks is smaller than the lesser of the two standard deviations. The tracks corresponding to individual two, Track 4, and individual three, Track 5, are clearly different to the tracks of individual one.

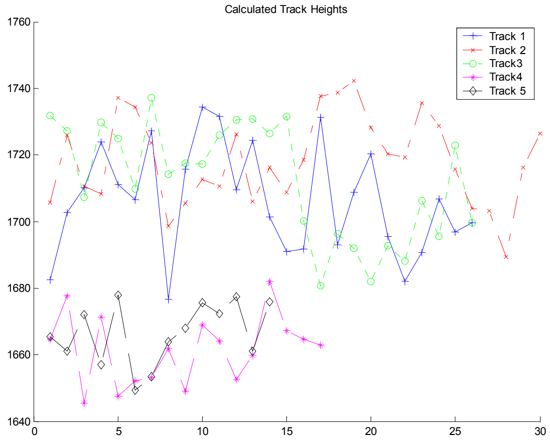

Figure 5: Estimated heights of an individual moving past the same camera twice.

Track	Average Height (cm)	Standard Deviation	Matching Tracks
Track 1	170.64	1.63	2, 3
Track 2	171.85	1.34	1, 3
Track 3	171.23	1.73	1, 2
Track 4	166.15	1.04	5
Track 5	166.64	0.92	5

Table 1: Statistical analysis of track heights

Height estimation is also a very fast session-based biometric to calculate. The camera calibration process can be time consuming to complete, but it is only required once when a camera is to be added into the building surveillance system. The time consuming components of the algorithm are actually extracting an individual from the background of the image, determining the head point and its vertical projection down the -Z direction to the ground plane. Converting these image points into 3D coordinates then requires solving the inverse of (4), and evaluating (5) for the data from each frame. Once the heights for the sequence have been determined, then their average and standard deviation can be quickly computed to compare with previous track information. This makes the height estimation fast to calculate once the individual has been extracted from the sequence. The extraction of the individual from the rest of the image can be provided by a background subtraction method, and is also required for all other biometrics [14]. Therefore this step can be considered as part of the overhead for the whole session-based biometric process rather than a component of the height estimation.

4 Conclusions

This paper proposed session-based biometrics as a means to match an individual's movements through a surveillance system. We proposed that using a session-based framework will provide two major benefits. Firstly it will provide a practical limitation on the number of potential matches of an individual within the system by only matching against those still within the system. Secondly it will allow the use of non-conventional biometrics for matching individuals as they are only required to remain stable for the surveillance session.

Height estimation has been shown here as an example of a session-based biometric. The results show that an individual's height can be matched both within, and between disjoint cameras in the surveillance system. The results also show that discrimination between individuals of different heights is also possible. Height estimation can be considered a useful lightweight biometric because it can provide a reasonable level of discrimination between individuals very quickly. Thus it can be combined with other session-based biometrics, including clothing appearance or gait, in order to uniquely identify an individual.

Acknowledgements

Special thanks to our colleague Eric Cheng for helping with the image acquisition by walking through the two cameras representing the building surveillance system.

5 References

[1] W. Hu, T. Tan, L. Wang and S. Maybank, 'A survey on visual surveillance of object motion and behaviors', *IEEE Transactions on Systems, Man and Cybernetics*, vol. 34, pp 334-352, 2004.

[2] M.-H. Yang, D. J. Kriegman and N. Ahuja, 'Detecting faces in images: a survey', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp 34-58, 2002.

[3] S. Pankanti, S. Prabhakar and A. K. Jain, 'On the individuality of fingerprints', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp 1010-1025, 2002.

[4] X. Yuan and P. Shi, 'Iris Feature Extraction Using 2D Phase Congruency', *International Conference on Information Technology and Applications*, vol. 2, pp 437-441, 2005.

[5] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother and K. W. Bowyer, 'The humanID gait challenge problem: data sets, performance, and analysis', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp 162-177, 2005.

[6] L. Lee and W. E. L. Grimson, 'Gait analysis for recognition and classification', *International Conference on Automatic Face and Gesture Recognition*, vol. pp 148-155, 2002.

[7] R. D. Green and L. Guan, 'Quantifying and recognizing human movement patterns from monocular video Images-part I: a new framework for modeling human motion', *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, pp 179-190, 2004.

[8] S. Khan and M. Shah, 'Consistent labeling of tracked objects in multiple cameras with overlapping fields of view, ' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp 1355-1360, 2003.

[9] O. Javed, K. Shafique and M. Shah, 'Appearance Modeling for Tracking in Multiple Non-Overlapping Cameras', *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp 26-33, 2005.

[10] M. Piccardi and E. D. Cheng, 'Track Matching Over Disjoint Camera Views based on an Incremental Major Color Spectrum Histogram', *IEEE Conference on Advanced Video and Signal Based Surveillance*, 2005.

[11] A. Criminisi, I. Reid and A. Zisserman, 'Single View Metrology', *International Journal on Computer Vision*, vol. 40, pp 123-148, 2000.

[12] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000

[13] R. Tsai, 'A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses', *IEEE Journal of Robotics and Automation*, vol. 3, pp 323-344, 1987.

[14] M. Piccardi, 'Background subtraction techniques: a review", *IEEE SMC 2004 International Conference on Systems, Man and Cybernetics*, 2004.