Australasian Transport Research Forum 2015 Proceedings
30 September - 2 October 2015, Sydney, Australia
Publication website: http://www.atrf.info/papers/index.aspx

Sensing and perception technology to enable real time monitoring of passenger movement behaviours through congested rail stations

Alexander Virgona, Nathan Kirchner, Alen Alempijevic

Centre for Autonomous Systems, University of Technology, Sydney, Australia

Email for correspondence: alexander.virgona@uts.edu.au

Abstract

Passenger behaviour can have a range of effects on rail operations from negative to positive. While rail service providers strive to design and operate systems in a manner that promotes positive passenger behaviour, congestion is a confounding factor, which can create responses that may undermine these efforts. The real time monitoring of passenger movement and behaviour through public transport environments including precincts, concourses, platforms and train vestibules would enable operators to more effectively manage congestion at a whole-of-station level.

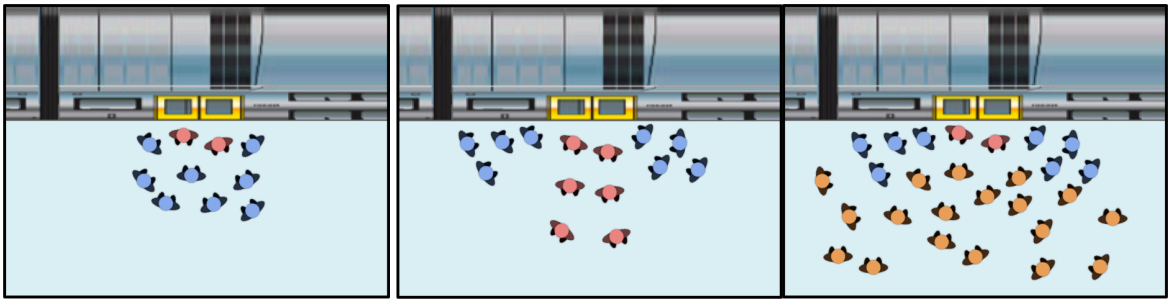
While existing crowd monitoring technologies allow operators to monitor crowd densities at critical locations and react to overcrowding incidents, they do not necessarily provide an understanding of the cause of such issues. Congestion is a complex phenomenon involving the movements of many people though a set of spaces and monitoring these spaces requires tracking large numbers of individuals. To do this, traditional surveillance technologies might be used but at the expense of introducing privacy concerns. Scalability is also a problem, as complete sensor coverage of entire rail station precinct, concourse and platform areas potentially requires a high number of sensors, increasing costs. In light of this, there is a need for sensing technology that collects data from a set of 'sparse sensors', each with a limited field of view, but which is capable of forming a network that can track the movement and behaviour of high numbers of associated individuals in a privacy sensitive manner.

This paper presents work towards the core crowd sensing and perception technology needed to enable such a capability. Building on previous research using three-dimensional (3D) depth camera data for person detection, a privacy friendly approach to tracking and recognising individuals is discussed. The use of a head-to-shoulder signature is proposed to enable association between sensors. Our efforts to improve the reliability of this measure for this task are outlined and validated using data captured at Brisbane Central rail station.

1. Introduction

Passengers are at the core of rail transport and their presence and behaviour can have a range of effects on rail operations (Wang & Legaspi 2012). The time taken for passengers to board and alight from a train at a platform can directly impact the total train dwell time, which in turn affects the reliability and frequency of services provided (Veitch et al. 2013). Undesirable passenger behaviour such as that exemplified in Figure 1(a), where passengers waiting on the platform block the doors of the train as it arrives, will slow passengers alighting from the train and increase dwell time.

Figure 1: Passenger behaviour can have a range of effects on rail operations



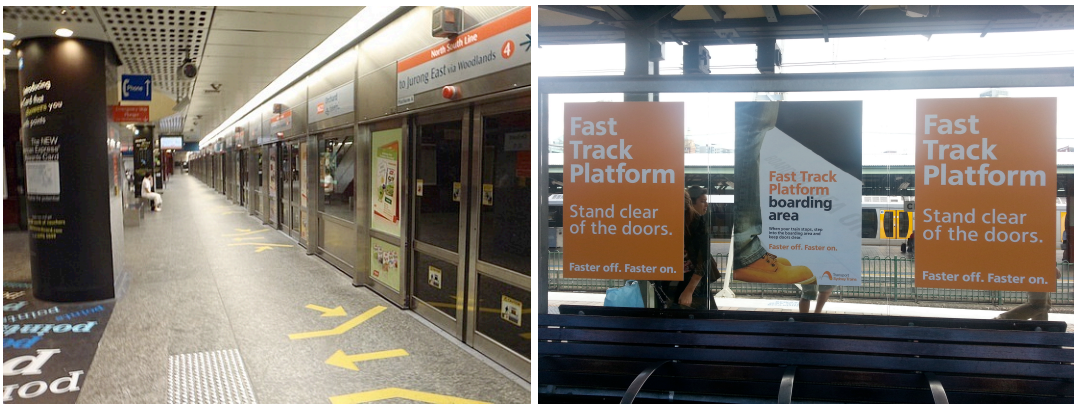
(a) Boarding passengers (blue) stand in front of doors, blocking alighting passengers (red), leading to increased dwell time.

(b) Boarding passengers (blue) stand to the sides of doors, allowing alighting passengers (red) to exit the train promptly, leading to reduced dwell time.

(c) Platform congestion (orange) undermines positive behaviour of some passengers (blue), increasing dwell time.

For this reason operators try to promote positive passenger behaviours, such as those in Figure 1(b) where passengers wait to the sides of train doors, through the use of signage (as in Figure 2), announcements and spatial design. Such treatments may be effective when the infrastructure is operating below capacity, but in peak times overcrowding undermines these efforts, as in Figure 1(c), and causes negative outcomes for passengers and rail operators alike. The negative effects of crowding are not limited to dwell time delays, with congestion in thoroughfares around the station delaying passengers from reaching their platform in time (Veitch et al. 2013), and further contributing to platform congestion while they wait for the next service.

Figure 2: Operators use platform signage to promote positive passenger behaviours.



(a) Platform markings used by Singapore Mass Rapid Transit

(b) Signage used by Sydney Trains

Passenger crowding in peak travel times clearly poses challenges for rail operators, causing numerous researchers to develop models to estimate the costs of overcrowding (Wang & Legaspi 2012, Veitch et al. 2013) and its effect on operational capacity (Gray 2013). Creative approaches to the problem such as fare differentiation to spread out peak travel times have been suggested and trialled with some success (Liu & Charles 2013), however as populations globally continue to rise such methods may not be sufficient to overcome these challenges. In order to combat overcrowding in the urban rail transport environment more information is needed about how, when and where overcrowding can occur. If operators could directly monitor the movements of people through the train station in real time this could lead to a deeper understanding of passenger behaviour and the causes of overcrowding.

The traditional approach to monitoring people in transport environments and other public spaces uses closed circuit television (CCTV) systems. Such systems typically involve a large number of cameras positioned throughout a train station, monitored from a control room by station staff such as in Figure 3. This type of system is ubiquitous in the transport industry and highly useful for security purposes, however the attention to detail required of operators and the high volume of visual information to be monitored in a large train station mean that there is a limit to the effectiveness of manual, CCTV based, crowd monitoring (Boghossian & Black 2005). This has led to the development of automated crowd monitoring systems based on computer vision, allowing real time crowd monitoring based on numerous video feeds.

Figure 3: The volume of video information captured and high level of attention required limit the efficacy of manual, closed-circuit television based crowd monitoring.

There are a number of computer vision based crowd monitoring systems currently available that provide real-time reporting of crowd densities observed by CCTV cameras. These systems allow operators to monitor the occupancy of trafficable areas throughout a train station and respond to overcrowding incidents as they occur. Whilst this technology presents practical benefits to operators, the level of information may not be sufficient to fully understand the complex causes of overcrowding. Beyond estimating gross crowd densities, a system capable of simultaneously tracking the real-time movements of many individuals throughout a crowded train station could give operators deeper insights into the specific causes of crowding issues and into passenger behaviour generally. This level of information could enable operators to find solutions to congestion issues at a whole-of-station level through changes to spatial design, and the development of responsive passenger information systems.

The task of autonomously tracking individuals has been approached by many researchers in the field of computer vision, with most methods relying on two main stages: person detection, and motion tracking. Successful tracking relies on detecting people with a great enough accuracy and frequency that their motion can be reliably predicted between observations for the purpose of associating each observation with a continuous track. There are many methods capable of detecting persons with sufficient accuracy and frequency to enable tracking within the field-of-view (FOV) of a single sensor however the problem remains that a train station is typically much larger than this.

In their recent work using 3D depth sensors for person tracking, Brscic et al. (2013) overcame the limits of their sensors' FOV by constructing a network of sensors with overlapping FOV covering a large area inside a shopping centre. By calibrating their tracking

system to account for the relative 3D positions of each sensor, they were able to successfully track individuals using multiple sensors as they move through the shopping centre. Although this example demonstrates the promise of using multiple 3D sensors for large-scale person tracking, the sheer number of sensors that would be required to achieve complete coverage of a major train station may limit the practicality of such an approach for our application. A more practical solution would be a system capable of tracking individuals across a network of sparse sensors, whose FOV do not necessarily overlap.

To achieve tracking of individuals across sparse sensors, a method is needed for associating observations of an individual made by one sensor with observations of the same individual made by another sensor. In their review of state-of-the-art person re-identification methods, Mazzon et al. (2012) break the task down into 4 main stages: multi-person detection, feature extraction, cross camera calibration and association. Assuming a system capable of person detection, the authors discuss a number of approaches to feature extraction with the most common being colour, texture and shape. The challenges typically faced by these methods are cited as "changes in pose, scale and illumination that modify the perceived appearance of a person across cameras" and the authors state that "In general, methods solely based on appearance features extracted on full body have performances close to random" (Mazzon et al. 2012). Face detection algorithms have also been applied to the problem of person tracking (Zhao et al. 2009) but are limited by the requirement for the face to be visible in every sensor FOV, an assumption which cannot reasonably be made in our context.

Besides the technical challenges of such vision based re-identification methods, this type of system is also likely to raise privacy concerns. When video surveillance is used the public are typically concerned with the protection of their personal information, and in a system which tracks the movements of individuals these concerns are likely to be amplified. It is plausible that video data captured by a vision-based person tracking system could be used to link people's monitored behaviour with their identity. If this capability is not the intended purpose of the system, it would be preferable to limit the potential for such misuse by design. In light of this, there is a need for a system capable of real-time tracking of individuals across a network of sparse sensors, in a way that respects their privacy. This paper discusses our work towards the core sensing and perception technology needed to create such a system. Building on our previous work in person tracking using 3D depth cameras, we explore the potential of our previously developed Head-to-Shoulder signature (Kirchner et al. 2012) for the task of track associations.

Following this introduction, Section 2 will give the background for this work, including a discussion of privacy in this context, and an overview of our previous work in person tracking on which we have built. Section 3 will detail our technical contributions including the proposed system framework and feature filtering method, which leverages intra-sensor person tracking to improve inter-sensor associations. Following this Section 4 describes our empirical validation of this method, based on data collected at Brisbane Central train station. Finally Section 5 will discuss our conclusions from this work and intended future work.

2. Background

2.1 Privacy in Crowd Monitoring

Surveillance in its various forms is increasingly embedded in all parts of our life. Much of our daily activity both online and offline is monitored in some way, generally with practical reasons in mind such as security, marketing and operational optimisation. However as surveillance increases, so too does the risk of misuse of personal information gathered, both intentionally and unintentionally. As such, the general public tend to be sceptical when additional surveillance measures are introduced in public spaces.

The issue of unintentional video surveillance is addressed in many cases by image processing systems built into surveillance equipment, capable of removing or scrambling

areas of an image which should not be monitored. While this may be helpful in certain circumstances, in the context of video based crowd monitoring where members of the public are intentionally observed, there is a risk the information gathered has utility beyond the intended purpose of the system. Perhaps a preferable scenario would be one where the information gathered was adequate only to perform the intended task.

The recent advent of economical 3D depth cameras offers a new type of sensing technology to the problem of crowd monitoring. 3D depth cameras can be used in a similar fashion to regular CCTV cameras, in that they have a similar FOV and can be positioned throughout an environment to observe regions of interest, with frame rates of up to 30fps. Where this technology differs from regular CCTV cameras is in the type of information captured. In a digital image taken from a standard colour camera, each pixel represents the colour and intensity of light coming from that part of the cameras FOV. In a depth image, on the other hand, each pixel represents the distance from the sensor of a visible surface in that part of the cameras FOV. Given some understanding of the optics of the camera, each depth pixel can be converted to a point in 3D space and the collection of points (pointcloud) generated by each depth image can be used to interpret the 3D location of objects in the FOV. This type of information has obvious advantages when applied to the task of tracking the locations of people in the scene, but less obvious are the privacy implications of the technology.

For the sake of discussion we can consider privacy concerns around surveillance technology under two broad questions: ”what personal information is observed by the system?” and ”how can this information be used?”. Starting with the first question, we can compare a system using standard colour CCTV cameras to one using 3D depth cameras. Figure 4 presents an example of each type of image. Looking first at the colour image, the types of personal information visible are: skin colour, hair colour, face, sex and some indication of age. In comparison the depth image gives us a silhouette of each person from which we could gather information about the height and build of a person, but little else of personal significance, and with modern computer vision techniques and some knowledge of the scene, height is obtainable from the colour image also.

Figure 4: 3D depth cameras inherently capture less personal information that standard colour CCTV cameras.

(a) Image from a standard colour camera (b) Image from a 3D depth camera

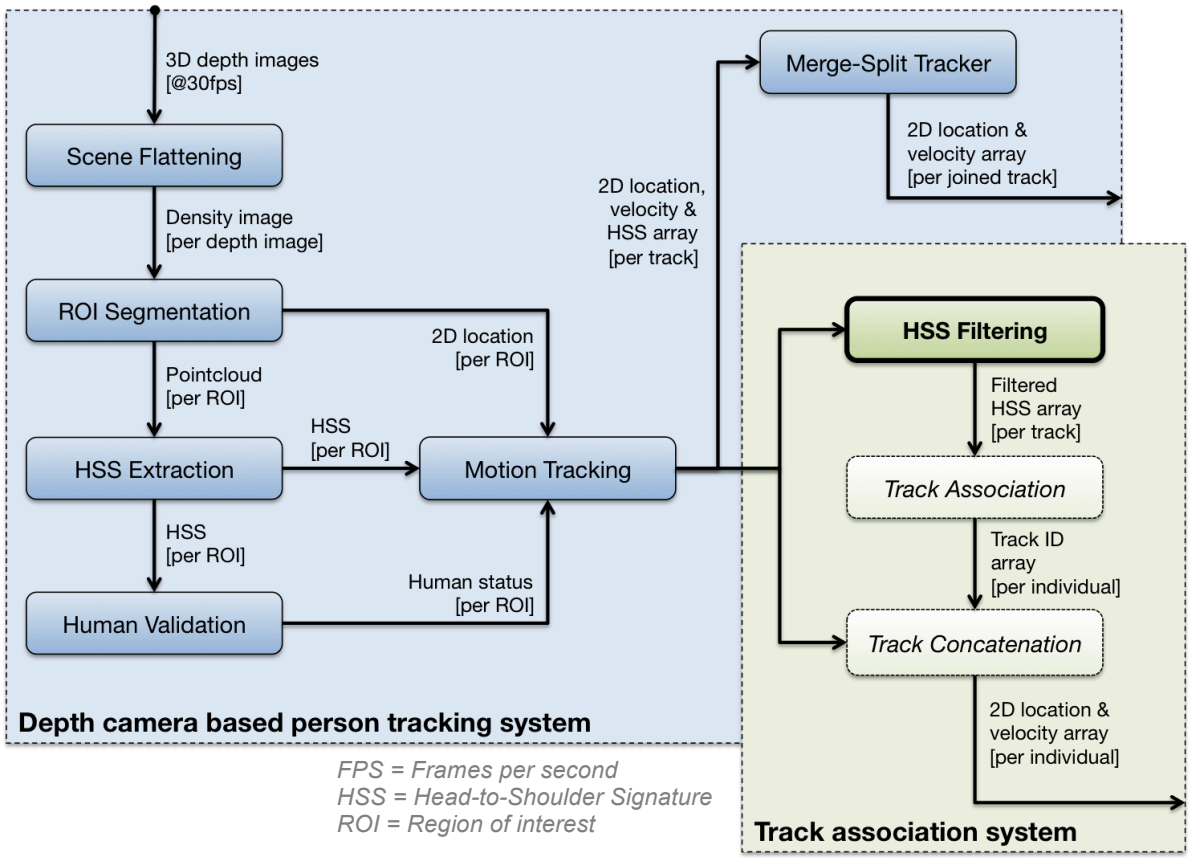
If we consider now how this information can be used, in the case of a system based on standard CCTV cameras, the movements and behaviour captured by the system would, apart from providing the desired crowd tracking information, be linked to an array of photographic images of the individual. The personal information contained in these images could be compared with a database of known persons containing similar types of information, such as that maintained by police, and combined with the time and place of observation could be used to resolve their identify. Comparing this with the 3D depth camera case, each individual’s activities would be associated with information about their height and shape, and while this type of information may also be contained in some identification databases it is

unlikely to be sufficient to provide a definitive identification. From this it seems that if the information available in the depth image is sufficient to perform the intended crowd monitoring task then there are inherent privacy benefits in using this technology over standard cameras.

2.2 Using Three-Dimensional Depth Cameras for Person Tracking

In order to provide context for later sections of the paper, this section outlines our previous work using 3D depth cameras to detect, track and count people within the FOV of a single sensor. The foundations of this work are in the field of human-robot-interaction (HRI) where a system was developed for detecting people from 3D depth data for use on a mobile robot in a domestic environment (Hordern & Kirchner 2010). The use of a 3D sensor in this work was motivated by the robustness to lighting variation achieved by the self-illuminating design of the sensor as well as the high utility of 3D data for extracting the location of the people detected. These benefits combined with success of the approach in robustly detecting people from a single point-of-view lead to additional research into extending the work for recognising individuals from 3D data (Kirchner et al. 2012), which will be explained in more detail in Section 2.3. Building on these foundations in the field of HRI, more recent work saw the development of a system for detecting, tracking and counting people in a public train station (Kirchner et al. 2014). Figure 5 shows the basic structure of the system and type of information passed between each of the functional components.

Figure 5: A system for real time person detection and tracking with proposed developments for tracking across sparse sensors



The *Scene Flattening* block converts the received depth images into 3D pointclouds and localises itself relative to the ground by locating the ground plane. After this localisation step, subsequent pointclouds are reoriented using the obtained ground-to-sensor transformation and flattened into a density image using the bivariate histogram technique presented by Hordern & Kirchner (2010). The intensity of each pixel in the density image represents the concentration of 3D points at a particular horizontal location in the sensors FOV. The *ROI*

Segmentation block detects regions of interest (ROI) potentially representing people, by performing blob detection on the density image, leveraging the assumption that people will present in the 3D data as vertical surfaces. As well as people this technique detects other vertical surfaces such as walls, poles and furniture. To deal with this, the overall dimensions of each ROI are checked and those considered too large or too small to represent people are eliminated, removing most but not all of the false positives generated in blob detection. The pointcloud segments contained by each of the remaining ROI are passed to the *HSS Extraction* block which constructs a descriptive feature vector called the Head-to-Shoulder Signature (HSS), using the method presented by Kirchner et al. (2012). These HSS are checked against a collection of known human HSS in the *Human Validation* block to determine if the ROI is human or not. The 2D location of each ROI along with its human status are passed to the *Motion Tracking* block which uses a particle filter, as described by Alempijevic et al. (2013) to track the detected people. Tracks are validated as belonging to a person once at least one of the associated observations is positively confirmed by the *Human Validation* block.

The high density of people in the public transport environment introduces the problem of merged detections due to the close proximity of people to one another. The motion model used by the *Motion Tracking* block provides some robustness to this issue however the problem remains that in a densely populated environment people walking near one another may be repeatedly detected as a single entity causing erroneous tracking results. The *Merge-Split Tracker* treats this by recognising merge and split events. When a track terminates near (within 0.8m of) another track the two tracks are considered merged and last observed HSS for each one are stored and associated with the continuing track. When a new track appears near a previously merged track the HSS for each track is obtained and compared to the previously stored HSS to resolve the identity of each post-merge track against the possible pre-merge tracks allowing the tracks to be repaired.

2.3 The Head-to-Shoulder Signature

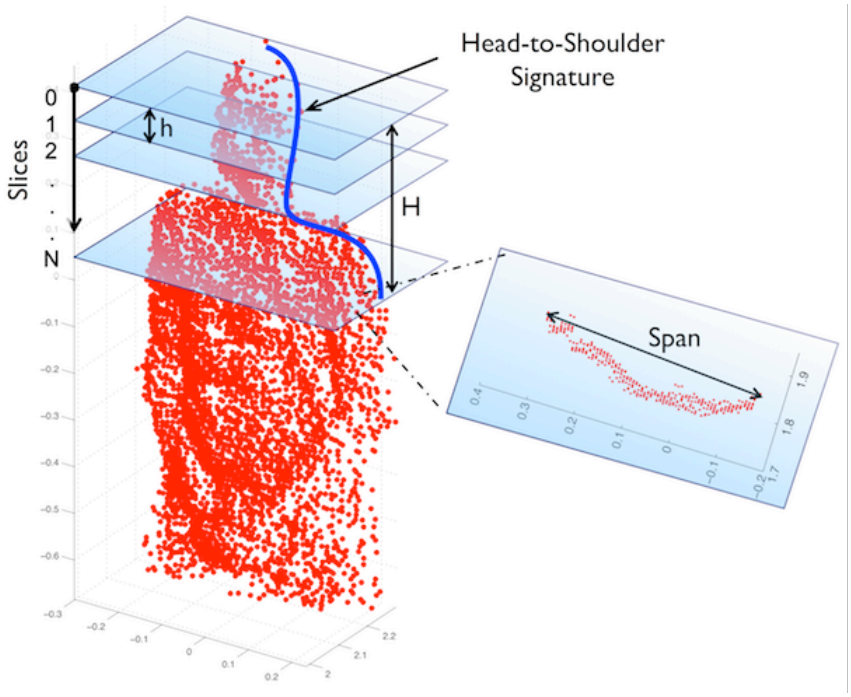
The Head-to-Shoulder Signature (HSS) is a descriptive feature vector that can be extracted from a 3D pointcloud of a person. The HSS was presented by Kirchner et al. (2012) as a method for encapsulating the 3D shape of a persons and head and shoulders for the task of person recognition in human-robot-interaction (HRI). Although less accurate for person recognition when compared with other biometric identification methods such as fingerprinting or face matching, the HSS was designed to be captured by a mobile robot from a single point of view without requiring the robot to disrupt the person, an important consideration in HRI. The method was shown to perform significantly better than random when differentiating between 25 individuals, achieving a mean classification accuracy of 78.15% on stationary participants and 52.11% while moving.

The HSS is constructed from a segment of a pointcloud containing a person as illustrated in Figure 6. The top 40cm of the pointcloud is divided into 20 horizontal slices each 2cm thick. For each slice the maximum horizontal distance between any two points in the slice is found, referred to as the span. In this way some robustness to viewing angle is provided, as the span can be in any direction, not only orthogonal to the observation angle. The 20 obtained spans are stored as an ordered vector forming the HSS.

Many of the characteristics that make the HSS suitable to the home environment transfer well to the transport environment making it a suitable candidate for track association in this context. The capability of the HSS to accommodate a variety of viewing angles, including viewing from behind, and the robustness to lighting variations provided by the self-illuminating function of the depth sensors used, make the method preferable to methods mentioned above which typically require more controlled input data. Additionally the use of 3D sensing in this context carries with it privacy benefits as discussed in section 2.1.

In our person tracking work (Kirchner et al. 2014) discussed in Section 2.2 the HSS is used to resolve the identity of tracks after a consecutive merge and split event. Whist this approach is quite effective in discriminating between two possible individuals, in the case of inter-sensor track association in a crowded train station it is likely there will be many more than two possible outcomes. As the number of potential association outcomes increases so too will the chance of miss-association. This increased complexity in the association problem places greater demand on the discriminative power of the HSS obtained for each track and therefore requires a method more robust to potential erroneous HSS measurements. Fortunately the motion tracking stage of the system presents an opportunity to improve this robustness.

Figure 6: The Head-to-Shoulder signature is constructed from a 3D pointcloud of person, summarising their shape via a series of horizontal slices



3. Leveraging motion tracking for Head-to-Shoulder Signature based track association

The HSS is designed to be robust to a range of viewing conditions however there are known failure modes and additional factors present in a crowded train station which may undermine the consistency of the HSS. This section explores the possibility of leveraging intra-sensor motion tracking to improve the robustness of HSS collected for the task of inter-sensor track associations. Specifically by exploiting assumptions about the relationship between velocity and body pose we propose to filter the HSS observed for each track to provide a consistent representation of the observed individual. The remainder of this section is divided into two parts: Section 3.1 proposes a framework for privacy sensitive, sparse 3D-sensor based, real-time person tracking which builds on our previous work and provides the context to present our method for HSS filtering; Section 3.2 discusses our method for HSS filtering which leverages motion tracking to improve HSS based track association.

3.1 A system for tracking people across sparse sensors

Our proposed system for developing a privacy friendly, sparse 3D-sensor based, multi-person tracking system is built upon our previous work into multi-person tracking and is comprised of some previously developed components, as well as some newly proposed components. Figure 5 enumerates the components and describes the information passed

between them. The *Scene Flattening*, *ROI Segmentation*, *HSS Extraction*, *Human Validation* and *Motion Tracking* blocks work as described in Section 2 of this paper. Together they provide an output of continuous tracks of location, velocity and HSS observations for each person that moves through the FOV of the 3D sensor.

The three newly proposed blocks are intended to concatenate the partial tracks output by the Motion Tracking block into complete trajectories of individuals through an entire train station. The role of the *HSS Filtering stage* is to process the full set of HSS to produce a representation that is descriptive of the observed individual and robust to errors in the individual HSS measured. Our method for HSS filtering is discussed in detail in Section 3.2. The Track Association block has the task of associating each track with an individual based on a combination of the filtered HSS measurements and available spatio-temporal information belonging to each track. This is a complex task and outside the scope of work for this paper. Finally the Track Concatenation block combines the tracks associated with each individual to form complete trajectories of their motion throughout the sparse sensor network. This final output can be stored and interrogated further to provide operators with a detailed understanding of passenger behaviour.

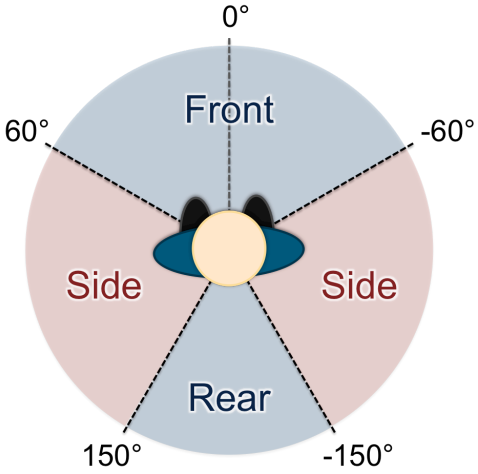
3.2 Filtering Head-to-Shoulder signatures for better track association

The object of HSS filtering in the context of the system described above is to produce a representation for each track that is descriptive of the observed individual and robust to errors in the HSS measurements such that it can be used by the Track Association block. Assuming the presence of some error in the HSS observed, collecting all the HSS together per track would allow for the extraction of robust statistics for each track, such as a median HSS to obtain a more consistent measurement. Better yet, statistical methods could be applied to compare entire distributions of HSS to one another and model the boundaries that separate one person's HSS from another's. This type of task is the role of machine learning classifiers and there exist several well developed methods, such as the Support Vector Machine used by Kirchner et al. (2012), which can be trained on labelled input data and used to classify subsequent inputs as belonging to one of the trained classes. The role of the *HSS Filtering* block then becomes to select the set of HSS that will best represent the underlying individual, to use as input data to a classifier.

Potential sources of error in the HSS measurement include: image coordinate quantisation, depth resolution quantisation, partial occlusions at the edge of the sensors field of view, partial occlusions due to other people in the environment, and observation angle. Partial occlusions causing significant deformation of the measured HSS tend to be omitted by the Human Validation stage of the system. Quantisation issues are worst at long range and treated in part by only detecting people within 8 meters of the depth camera. While lesser occlusion and quantisation errors will still contribute to measurement error the largest remaining source of error is observation angle.

The HSS is designed to accommodate a range of observation angles however the method is limited by the 3D data input to the HSS extraction algorithm, meaning that if the full breadth of the shoulders is not observed by the sensor the HSS will reflect this. The ability of the HSS to accommodate different observation angles is evaluated by Kirchner et al. (2012) based on data of one of the participants of the study, and it is noted, that classification using the HSS was successful for a range of observation angles from -60° to +60° and from -140° to 160°, where 0° is a frontal view. It is expected therefore that for angles outside these ranges there is likely to be some deviation from the expected HSS measurements causing the classification to fail. Based on these results, and assuming the general case to be symmetrical we divide the full range of possible observation angles into three groups as show in Figure 7. Front refers to the angles -60° to +60° via 0°, rear refers to -150° to 150° via 180° and side refers to the remaining ranges.

Figure 7: Observation angles are divided into three cases: front, rear and side.



In order to omit potentially erroneous measurements based on observation angle, the pose of the tracked individuals must be determined. A method for shoulder pose estimation using principle component analysis, similar to that used by Brscic et al. (2013), was implemented and proved successful for a range of angles. Unfortunately the typical failure cases of this method cause it to measure extreme portrait observations (near 90° or -90°) as close to 0° rendering it unsuitable for classifying observations as either front or side views. In light of this a method was needed which would not be susceptible to errors related to observation angle.

Based on the assumption that people walking will typically align their head and shoulders with their direction of travel, the velocity of the track provided by the motion tracker was chosen as an estimate of body pose. For each HSS observation the direction of the track velocity at that point in time was assumed to be the facing direction of the person. While this assumption proved fairly reliable at a normal walking pace, for the case of a person standing still slight variations in the measured location of a person would cause velocity measurements with very small magnitude and random direction. To prevent these measurements from corrupting our experimental results a threshold of 0.5 metres per second was set and pose estimations derived from track velocities under this threshold were disregarded. For the remaining observations with velocities above 0.5m/s the observation angle was determined based on direction of the track velocity and location relative to the sensor of the person at the time of the observation. For each track, observations were divided into those observed from the front, rear and side angles.

Due to the velocity threshold used in determining observation angle there may be many HSS collected while a person is stationary which are perfectly valid measurements and yet ignored due to the limitations of the pose angle estimation technique. Additionally there may be samples selected as part of the front set which still contain significant errors due to failures in the pose angle estimation technique. For this reason the observations taken from the front for each track are used as the basis to select a subset of the total HSS collection using Mahalanobis distance. The Mahalanobis distance is a measure of the distance between a point and distribution used in multivariate statistics to test if a particular point belongs to a distribution. In this case we calculate the distance between each HSS observation in the total set and the distribution of observations observed from the front. Based on the calculated distances the inner 20% of the total collection of HSS are selected to represent the track for the purpose of Track Association. Section 4 discusses the results of this selection process on real world 3D data.

4. Experimental Results

4.1 Data collection and processing

The HSS Filtering method described above was evaluated on data captured at Brisbane's Central train station using three purpose built sensor hardware platforms (Kirchner et al. 2014), comprised of a 3D Depth camera, compact fanless PC (FitPC3), hard drive and battery system, pictured in Figure 8(a). Over the course of three days, 3D depth images were recorded at 30 frames per second in areas of passenger flow through the train station. For the sake of evaluating the HSS Filtering method discussed above, a 29 minute recording of a high passenger traffic area on one of the stations train platforms was selected for the large numbers of moving people. Figure 8 shows a depth image from the data set.

Figure 8: The sensor hardware platform mounted in Brisbane Central train station collecting 3D depth images of rail passengers

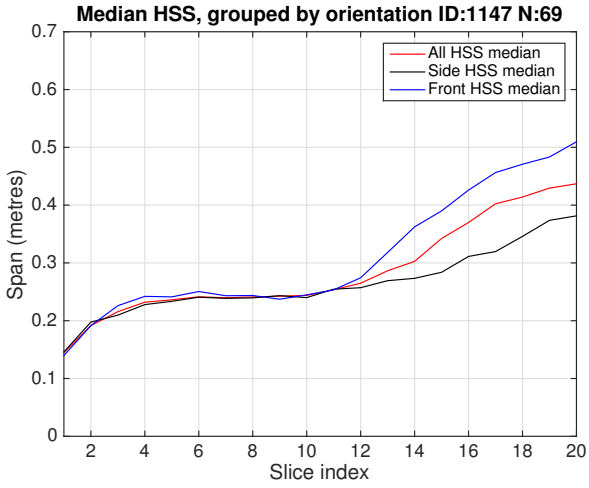
(a) The sensor hardware platform in situ (b) A depth image from the evaluation data set

The depth images from this dataset were processed by our person detection and tracking system, implemented in C++, using the Robot Operating System (ROS). ROS is an open source software framework for robotics development which provides key services such as message passing and drivers for many common robotics components and sensors. As described in Section 2.2 the output of the person tracking system is a collection of tracks containing observations of HSS, location (x,y) and velocities (x,y). This dataset was logged using ROS and imported into MATLAB in order to process and visualise the results of the HSS Filtering method.

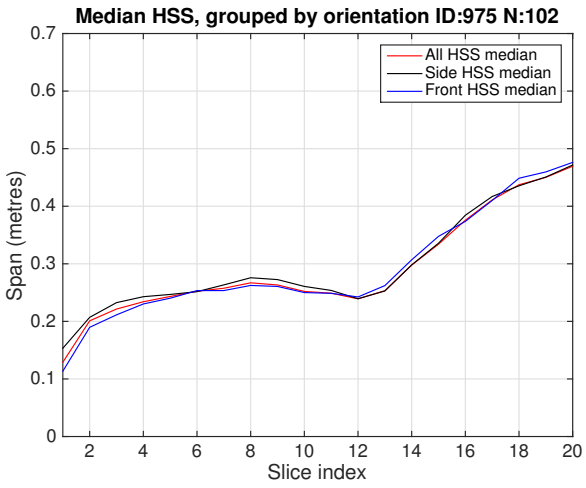
4.2 Head-to-Shoulder Signature Filtering Results

The dataset used to evaluate the HSS filtering method contained 1130 tracks with numbers of observations per track ranging from 1 to 11331 and a variety of walking paths through the sensor FOV. In order to evaluate the HSS Filtering method presented, a subset of tracks was selected for detailed analysis in which at least 30 frames (approximately 1 second worth) were captured in both front and side views. The effect of grouping HSS measurements for each of these tracks by observation angle was examined by plotting the median HSS for each group. In this way the characteristic differences between groups of HSS could be clearly visualised. The median HSS is a vector constructed from the median value of each slice index across the set of HSS. The median is used here rather than mean for its robustness to outliers. Figure 9 shows the result of this process on two tracks that are indicative of the range of results encountered.

Figure 9: Median HSS plots for two tracks, shown before grouping HSS by observation angle (red) and after grouping by observation angle into front (blue) and side (black).

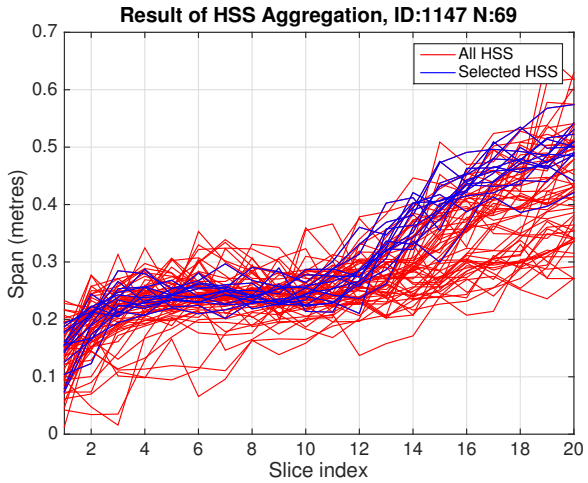


(a) In most cases, such as track 1147, front (blue) and side (black) groupings showed a clear difference in the shoulder region.

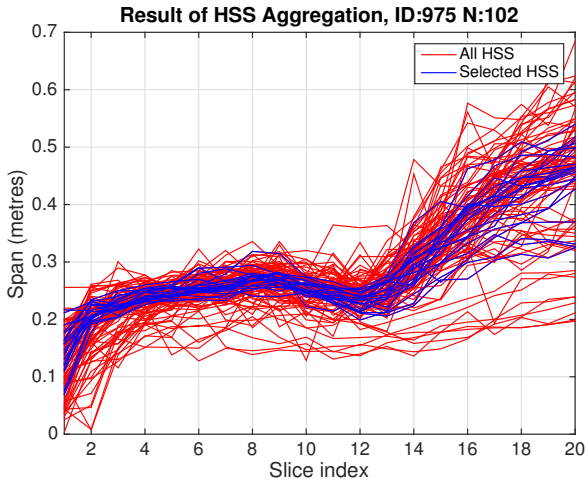


(b) In some tracks such as track 975 the difference in shoulder breadth was not visible, possibly due to less extreme side (~90°) observations.

Figure 10: HSS selected based on similarity to frontal observations (blue) plotted against all HSS observations (red) for two tracks.



(a) For most tracks, such as track 1147, selected HSS (blue) contain less variation than the complete set (red) and represent a greater shoulder breadth.



(b) In cases such as track 975 where front and side HSS were not well differentiated the selected HSS (blue) tend towards the dominant shoulder shape, omitting many clearly erroneous observations.

The plots resemble the shape of the head and shoulders of a person lying on their side, where slice index 1 corresponds to the top of a person's head and slice 20 is a point 40cm below that. In most cases the median of HSS measured from the front showed a visible deviation from those measured from the side, particularly in the shoulder region (slices 14 to 20). In most cases, such as in track 1147 in Figure 9(a), the front HSS median (blue) shows a broader shoulder region than the side HSS median (black), with the HSS median for all measurements (red) predictably showing somewhere in-between the two. The significance of this is that a track association method not taking into account shoulder orientation would tend to underestimate the shoulder breadth of individuals and misrepresent their true shape, likely leading to poorer association results. This effect is most likely explained by self-occlusion in extreme side observations (90°), where the view of the far shoulder is hidden behind the head and neck of the person. This trend is more pronounced in some samples than others

with cases such as track 975 in Figure 9(b) showing no significant difference between the two. This may simply be due to the side observations for this track being less extreme but could also be related to individual characteristics such as head size and hairstyle.

As described in Section 3 the front HSS are used as the basis to select a subset of the original HSS that will be used to represent the track. Figure 10 shows the result of this selection process on the same two tracks shown in Figure 9. The most noticeable difference between the unfiltered set (red) and those selected based on the shape of the frontal observation set (blue), is the reduced variation within the set. This is expected, as the HSS selection process based on the Mahalonobis distance, is a form of outlier removal. Importantly however, the selected set does not appear in the centre of the distribution of all HSS but rather favours the HSS shape determined by the frontal HSS set. By not only reducing the variance of the HSS used but also basing this selection on the most consistent HSS observations, those captured from the front, it is expected that the most erroneous HSS will be omitted and the performance of subsequent HSS based track matching will be improved. This effect is clearly visible in Figure 10(a) where a large number of HSS measurements clearly underestimating the shoulder breadth have been omitted. Even in the case of track 975 in Figure 10(b) where the difference in shape between side and front observations was not clearly pronounced the majority of selected HSS (blue) seem to follow the dominant wider shoulder shape.

5. Conclusions and future work

The development of the privacy sensitive, multi person tracking system described in this paper is expected to provide rail operators with a valuable tool for understanding passenger movements. Data gained from such a system could be used in a variety of ways to improve service outcomes including: informing spatial design choices, providing effective real time monitoring solutions capable of early detection of congestion issues, and enabling responsive passenger information systems capable of automated crowd management.

This paper has discussed our efforts towards developing the sensing and perception technology to enable privacy sensitive, multi-person tracking across a network of sparse sensors. Specifically it has proposed the use of the HSS for associating observations of individuals across a sparse sensor network to reconstruct their complete movement trajectories. A method for improving the efficacy of the HSS for this task has been explored which leverages data association provided by intra-sensor motion tracking to choose a reliable subset of the available HSS to be used for track association. Based on prior investigations into the effect of observation angle on the HSS, track velocities are used to estimate body pose, and hence observation angle, and choose HSS based on their similarity to those measured from known reliable angles of observation.

The proposed HSS selection method is tested on data collected at Brisbane Central train station and the results are discussed. A clear difference is noted in the HSS collected from front and side observation angle respectively highlighting the value of segmenting HSS based on observation angle. The HSS sets selected for each track give a tighter representation than the original set, centred around the reliable frontal HSS set. This improved representation of each track is expected to allow more reliable track association provided that at least some frontal HSS observations can be made for each track.

Future work will investigate the possibility of developing multiple representations per individual based on different viewing angles to account for cases where frontal images are not observed at all. Beyond this the non-trivial task of track association will be approached using a combination of HSS based classification and spatio-temporal information. The use of spatio-temporal information for inferring a person's future behaviour is also currently being explored by others within our research group and is expected to provide valuable inputs to this work.

References

Alempijevic, A., Fitch, R. & Kirchner, N. (2013), 'Bootstrapping navigation and path planning using human positional traces', *Proceedings - IEEE International Conference on Robotics and Automation* pp. 1242–1247.

Boghossian, B. & Black, J. (2005), 'The Challenges of Robust 24/7 Video Surveillance Systems', *IEE International Symposium on Imaging for Crime Detection and Prevention* (1), 33– 38.

Brscic, D., Kanda, T., Ikeda, T. & Miyashita, T. (2013), 'Person Tracking in Large Public Spaces Using 3-D Range Sensors', *IEEE Transactions on Human-Machine Systems* **43**(6), 522– 534.

Gray, J. (2013), Rail simulation and the analysis of capacity metrics, *in* 'Australasian Transport Research Forum Proceedings', number October, pp. 1–15.

Hordern, D. & Kirchner, N. (2010), 'Robust and Efficient People Detection with 3-D Range Data using Shape Matching', *Proc. of the 2010 Aust. Conf. on Robotics and Automation* pp. 1–8.

Kirchner, N., Alempijevic, A. & Virgona, A. (2012), Head-to-shoulder signature for person recognition, *in* 'Proceedings - IEEE International Conference on Robotics and Automation', pp. 1226–1231.

Kirchner, N., Alempijevic, A., Virgona, A., Dai, X., Pl, P. G. & Venkat, R. K. (2014), A robust people detection , tracking , and counting system, *in* 'Australasian Conference on Robotics and Automation'.

Liu, Y. & Charles, P. (2013), Spreading peak demand for urban rail transit through differential fare policy : A review of empirical evidence 2 . Transit Peak Spreading and Fare Differentiation, in 'Australasian Transport Research Forum Proceedings', Vol. 2007, pp. 1–35.

Mazzon, R., Tahir, S. F. & Cavallaro, A. (2012), 'Person re-identification in crowd', *Pattern Recognition Letters* **33**(14), 1828–1837.

Veitch, T., Partridge, J. & Walker, L. (2013), Estimating the Costs of Over-crowding on Melbourne s Rail System, *in* 'Australasian Transport Research Forum 2013 Proceedings', number October, pp. 1–14.

Wang, B. & Legaspi, J. (2012), Developing a train crowding economic costing model and estimating passenger crowding cost of Sydney CityRail network, *in* 'Australasian Transport Research Forum Proceedings', number September, pp. 1–15.

Zhao, X. Z. X., Delleandrea, E. & Chen, L. C. L. (2009), 'A People Counting System Based on Face Detection and Tracking in a Video', 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance pp. 67–72.