Pedestrian monitoring techniques for crowd-flow prediction

Claudio Martani PhD
Research associate, Department of Architecture, Centre for Smart Infrastructure and Construction, University of Cambridge, Cambridge, UK (corresponding author: clhouse.martani@gmail.com)

Simon Stent MEng
PhD student, Department of Engineering, University of Cambridge, Cambridge, UK

Sinan Acikgoz PhD
Brunel Research Fellow, Department of Engineering, Centre for Smart Infrastructure and Construction, University of Cambridge, Cambridge, UK

Kenichi Soga PhD
Professor, Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA

Dean Bain
Survey manager, Costain, London, UK

Ying Jin PhD
Senior lecturer, Department of Engineering, Centre for Smart Infrastructure and Construction, University of Cambridge, Cambridge, UK

The high concentration and flow rate of people in train stations during rush hours can pose a prominent risk to passenger safety and comfort. In situ counting systems are a critical element for predicting pedestrian flows in real time, and their capabilities must be rigorously tested in live environments. The focus of this paper is on evaluating the reliability of two alternative counting systems, the first using an array of infrared depth sensors and the second a visible light (RGB) camera. Both proposed systems were installed at a busy walkway in London Bridge station. The data were collected over a period of 2 months, after which, portions of the data set were labelled for quantitative evaluation against ground truth. In this paper, the implementation of the two different counting technologies is described, and the accuracy and limitations of both approaches under different conditions are discussed. The results show that the developed RGB-based system performs reliably across a wide range of conditions, while the depth-based approach proves to be a useful complement in conditions without significant ambient sunlight, such as underground passageways.

1. Introduction

The high concentration and flow rate of people in stations during rush hour pose a prominent risk to passenger comfort and safety. Passenger counters and flow prediction tools have the potential to reduce these risks by anticipating upcoming overcrowded situations and allowing preventive mitigation actions. To date, their use in station design and crowd monitoring has resulted in improved passenger safety and comfort (Hashimoto et al., 1998; Li, 2000). However, changes in train platforms and schedules and related operational problems in the rail network continue to cause unexpected crowds. Therefore, there is a need for passenger flow-prediction tools that can quickly respond to changes in the dynamic station environment. Such tools can help asset managers to decide when to close the ticket barriers and guide the passengers to alternative routes.

Existing pedestrian passenger prediction tools use conservative estimates of average count statistics to simulate flows. However, these statistics are typically calculated over a short time span and are not updated. Therefore, they do not capture dynamic changes in the station, which might result in crowding (Batty and Torrens, 2001; Li and Zhu, 2016; Timmermans, 2009; Zachariadis, 2005). To capture these changes, it is necessary to utilise real-time passenger-counting tools.

One popular passenger-counting tool is the infrared beam counter (Bu et al., 2007; Stogdale et al., 2003). However, this solution is not optimal as singular-beam counters do not provide directional counts; these sensors can detect only the total number of passengers crossing a predefined line, irrespective of their direction (Amin et al., 2008). In order to count passengers and determine their direction of movement, image processing techniques that utilise video data can be used (Chow and Cho, 2002; Hu et al., 2013, 2015; Sacchi et al., 2001; Schofield et al., 1997). In this approach, the movement of passengers within the field of view of the camera is tracked by comparing the positions of identified features from sequential instances.

Many different types of cameras can be used for counting passengers. Infrared depth sensors may be used to sense pedestrians by volume recognition, independent of their visual appearance. When the sensor faces downwards, only the tops of passengers' heads are visible and related features can be identified and tracked. The ability of these sensors to count pedestrians was only recently demonstrated (Zhang et al., 2012) and tested (Del Pizzo et al., 2016; Vera et al., 2016). In contrast, pedestrian sensing by using visible light is better established. During data retrieval, image features such as histograms of oriented gradients (Dalal and Triggs, 2005) can be used to detect and track pedestrians in real time (e.g. Xu et al. (2015)). Similar to depth sensors, when these cameras are mounted downward facing, there are minor occlusions and feature identification and tracking can be achieved more reliably. However, there is limited information in the literature on the performance of these sensors under different lighting and crowding conditions that might be experienced in a station.

To predict and localise congested 'hotspots' in crowded stations in real time, a new method has been recently proposed (Martani et al.,

Smart Infrastructure and Construction
Volume 170 Issue SC2

**Pedestrian monitoring techniques for
crowd-flow prediction**
Martani, Stent, Acikgoz *et al.*

2015) (see Figure 1). This method consists of three components: (*a*) a system for real-time people counting at all origin and destination points, (*b*) a system for estimating passenger movements between origin and destination points and (*c*) a microsimulation model.

For the method to be implemented, it is essential that the components handle the difficult working conditions of a busy train station. Within this context, the primary aim of this paper is to evaluate the suitability of two counting technologies in a rail station environment (grey box in Figure 1). These two technologies utilise cheap and highly portable sensors, and the associated crowd counting techniques require minimum computational power. These sensors are (*a*) an array of infrared structured light, depth-only sensors and (*b*) a wide field of view visible light camera, and their pertinent counting processes are discussed in detail in Section 2. The test deployment was conducted in one of the busiest railway stations in the UK, London Bridge station. The test, which is reported in Section 3, spanned a 2-month period and provided a rich data set, which allowed a methodical evaluation of the accuracy of the two approaches, described in Section 4. To the best of the authors' knowledge, the present work is the first effort to test the reliability of these novel counting techniques in the working conditions of a busy train station.

2. Two approaches to real-time crowd counting

In this paper, automatic and real-time crowd counting is performed using two alternative approaches: using depth-only infrared sensors and using a visible light (RGB) camera.

2.1 Depth sensor-based counting approach

The depth sensor-based counting approach consists of two steps: firstly, human-sized volumes are extracted from an array of calibrated raw Xtion Pro depth data, and then the volumes are recognised and tracked using software provided by Counterest Ltd, which uses decision trees for recognition training using an active learning approach described by Bonnin *et al.*, (2011). The output of the counting process is the bidirectional counting data – that is, the number of passengers crossing a predefined line in both directions within a certain time (see the bar chart in Figure 2). Other types of information can also be collected as a byproduct of this, such as pedestrian orientation and speed.

The Xtion Pros are low-cost active light depth sensors that are able to provide real-time range videos in a range of 0·8–3·5 m. This is possible by the projection of a known infrared dot pattern into the scene and stereo matching the pattern with a camera offset from the projector to resolve depth. This active-light depth sensing is known to be very good at detecting depth in untextured scenes (Chen *et al.*, 2016). When mounted downward facing, the VGA resolution depth images can allow easier segmentation of crowded scenes since depth discontinuities clearly mark the boundaries between pedestrians (Figure 2(a)). The technology is particularly well suited to low-lit scenes, where visible light may struggle to discriminate pedestrians. In contrast, the volume recognition ability of Xtion sensors is known to suffer in daylight conditions, as ambient sunlight interferes with the projected light pattern and can cause noisy, incomplete depth maps.
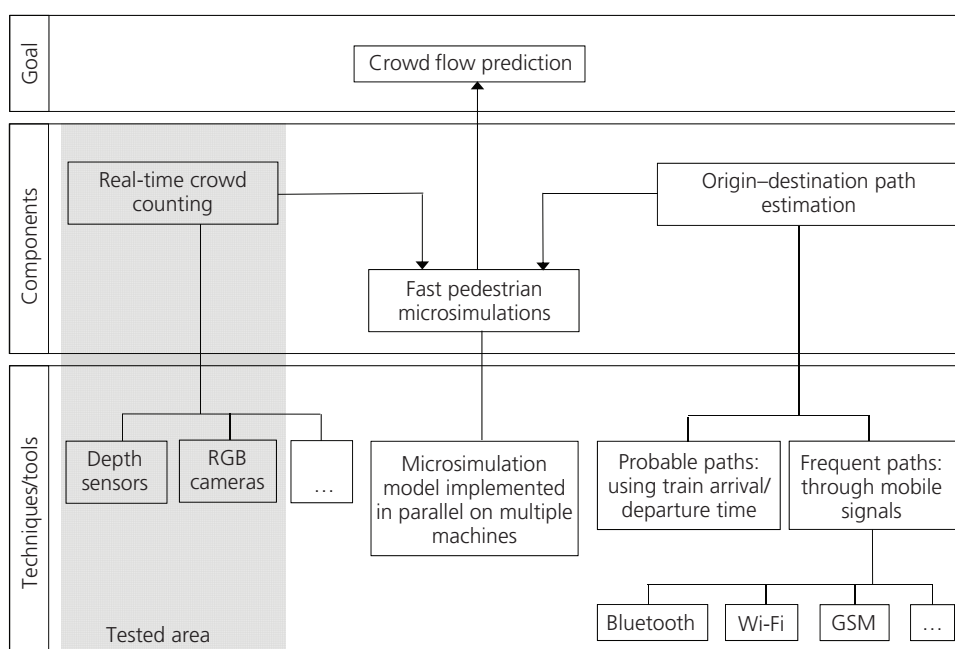

Figure 1. General concept of combined sensing and microsimulation for predicting crowd flows in stations. GSM, Global System for Mobile Communications (diagram developed by Vassilis Zacharidis as part of Martani *et al.* (2015))
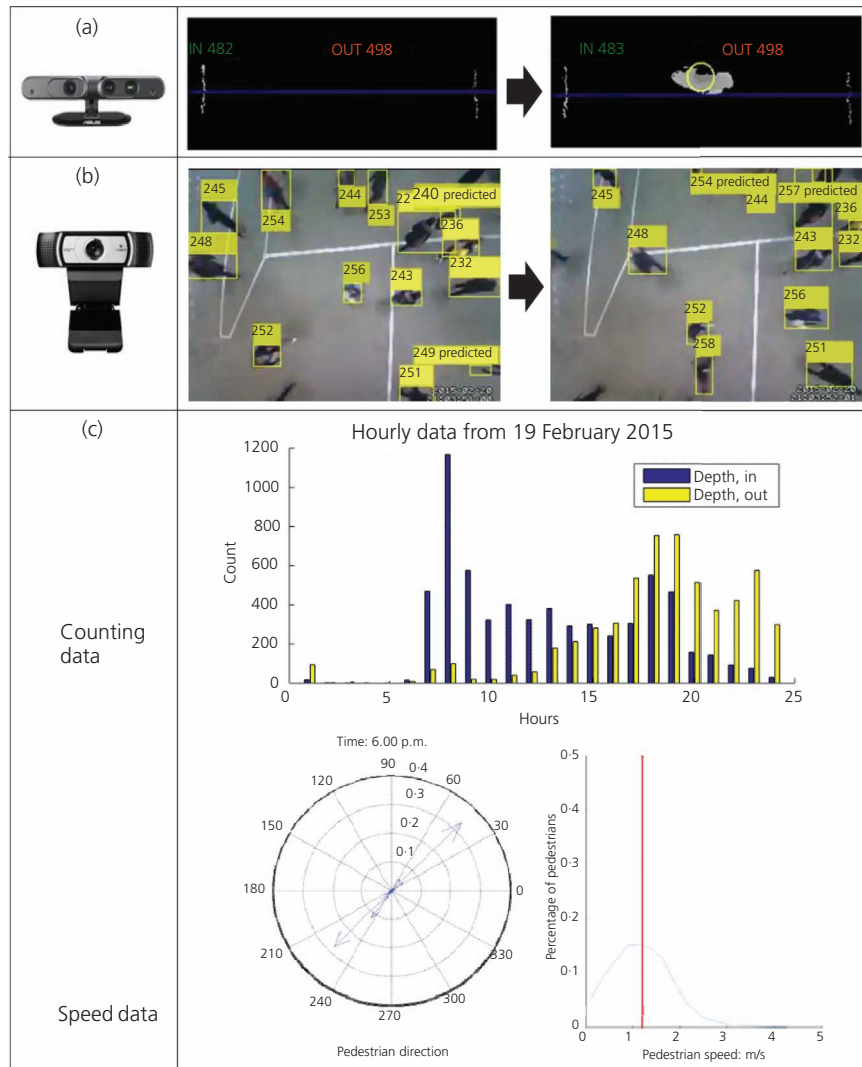
Smart Infrastructure and Construction
Volume 170 Issue SC2

**Pedestrian monitoring techniques for
crowd-flow prediction**
Martani, Stent, Acikgoz *et al.*

Figure 2. (a) A depth sensor and (b) an RGB camera, both with two sequential snapshots showing a sample of the two counting processes. Shown at the bottom are sample outcomes (c) in terms of counting data, average pedestrian walking direction and speed

The system automatically performs pedestrian identification, tracking and counting. Firstly, the infrared data from the depth sensors is thresholded by height to a calibrated ground plane to identify pedestrian heads. The software then tracks the heads and evaluates if they cross a preset line. Tracking within the frame allows determination of the direction of the passenger (in against out) as well as preventing repeat counts when pedestrians cross the line several times while walking or waiting in close proximity (Bonnin *et al.*, 2011). Thus, the depth data of each frame are only temporarily stored and every time a pedestrian crosses the line, data are logged on the machine with a unique time stamp and a binary indicating direction. Sample passenger counts for passengers leaving and entering a train platform, gathered using this technique, are shown in Figure 2 (first row). This enables the depth sensor to function as a highly efficient real-time counter.

The visible light camera-based counting approach consists of six steps: image capture, background subtraction, pedestrian detection, pedestrian tracking, data association and count updating. In this section, the main features of this process are outlined. Similarly to depth sensors, the RGB video sequence allows the direction and speed of passengers to be estimated (see graphs in Figure 2(c)).

2.2.1 Image capture
Greyscale images are captured using a Widecam F100TL 1080 pixel high-definition web camera, which has a 120° horizontal field of view, and the Linux *motion* package. The image resolution and frame rate were chosen to trade off data storage/ processing costs with the difficulty of detecting individual pedestrians and performing frame-to-frame data association. It

Smart Infrastructure and Construction
Volume 170 Issue SC2

Pedestrian monitoring techniques for
crowd-flow prediction
Martani, Stent, Acikgoz *et al.*

was found empirically that a frame rate of 2 Hz and an image resolution of 320 × 240 pixels were suitable for the task.

2.2.2 Background subtraction

At each time instant, the input image is read and its background image is subtracted to yield a foreground difference mask, which highlights moving objects in the scene with a different intensity to the background. The background image is calculated using the mean image from the previous N frames, where $N = 50$ for the experiments, corresponding to 25 s. This allows both a regular update of the background condition as lighting conditions change throughout the day and the exclusion of stationary objects left in the field of view of the camera. The resulting absolute difference mask is thresholded to yield a noisy binary foreground map. The threshold is set at 0·15 in all experiments (where the image ranges from 0 to 1) by calibration on a training data set.

2.2.3 Pedestrian detection

The binary foreground map is then filtered using a sequence of open, close and fill morphological operations in order to remove salt-and-pepper noise and fill in holes. Rectangular filters measuring 3 × 3 and 5 × 5 were used for the open and close operations, with a four-connected neighbourhood to fill holes. The resulting binary map is segmented by connected components to yield a set of bounding boxes and centroids corresponding to individual shapes (pedestrians) in the map. While it is pointed out that more complex pedestrian detectors exist (e.g. Dollar *et al.* (2012), Tian *et al.* (2014)), those detectors are designed for the more challenging case of frontal pedestrian detection, where pedestrians often occlude one another or are occluded by other objects, and there is more significant visual variability in the scene. In this work, downward-facing cameras are used to remove much of this visual complexity.

2.2.4 Pedestrian tracking and data association

A Kalman filter (Kalman, 1960) was used to model each pedestrian, as described in similar approaches (Li *et al.*, 2010) and as shown in Figure 2(b). It was observed that in the case of top-down data of pedestrians moving in a walkway, the linear Kalman model was sufficient for tracking; therefore, the use of non-linear methods such as extended Kalman filters or particle filters was not investigated. When dealing with multiple pedestrians, some of whom may be missing data at particular frames and whose trajectories can occasionally collide, the problem of assigning pedestrian proposals (observations) to existing pedestrian tracks (models) must be addressed. In the presented approach, the Euclidean distances between predicted and measured pedestrian positions are taken into account, as well as the confidence of the predictions (the covariance of each Kalman model) and the cost of non-assignment, which is set experimentally to reduce the number of false-positive proposals that generate new tracks. The cost of data assignment is minimised using the Kuhn–Munkres, or Hungarian, algorithm (Munkres, 1957), which is the standard method for finding the optimal set of detection-to-track assignments in polynomial time.

Since pedestrian motion in the image is relatively small over 0·5 s intervals (at 2 Hz) and the downward-facing camera means occlusions are very rare, the process is observed to work well. Pedestrians are counted when they cross over a partition in the image and disappear from view and contain more than five assigned observations, to reduce spurious readings. The initial and final observations of those tracks are used to generate velocity statistics for each pedestrian.

3. Application

The proposed methods were applied in London Bridge station to assess the accuracy of the counting technologies discussed in Section 2, as well as their stability over time under continuous operation. The details of the application are presented in this section. The application has been implemented in two steps: sensor deployment and data gathering and validation of the automatic counting.

3.1 Sensor deployment and data gathering

Three depth sensors and one RGB camera were mounted to the steel-and-glass roof structure at the eastern end of platforms 12 and 13 of London Bridge station. The sensors were strategically placed to capture a large number of incoming and outgoing passengers. The sensed area was utilised by the vast majority of passengers proceeding from platforms 12, 13, 14 and 15 to the ticket barriers and vice versa (see Figure 3). After the ticket barriers, there is access to the station, the underground lines, the bus station and the connection to the other platforms. Depending on their direction, the passengers enter the field of view of the camera through the different sides of the image. For instance, passengers using platform 12 appear on the top left side, while passengers from platform 13 may be seen entering the frame on its top centre and top right. On the other hand, passengers going towards the platforms 13–15 typically enter the image frame around its bottom-left corner. Since the location experiences a variety of crowding and lighting conditions throughout the day, it provided an extensive data set for assessing the reliability of the technology.

All sensors were mounted, downward-facing, 3·7 m above the region of interest. The three depth sensors were placed 1·5 m apart to capture the 8 m^2 region of interest, while the RGB camera was placed centrally to capture the same area. Note that a single RGB camera provided the same effective field of view as three depth sensors in series. This is due to the limited field of view of the depth sensor due to its stereo camera arrangement. The lightweight sensors were attached to temporary steel beams affixed to the permanent roof structure. The steel beams of the roof also supported a temporary unit to house a mini-personal computer (mini-PC) which controlled automatic data acquisition, processing and networking. The mini-PC was powered from a 110 V supply, and the sensors were powered over universal serial bus. The sensors were deployed, and data acquisition was initiated during a single night shift. This required the attendance of specialist engineers, however, no additional maintenance was required after the installation.
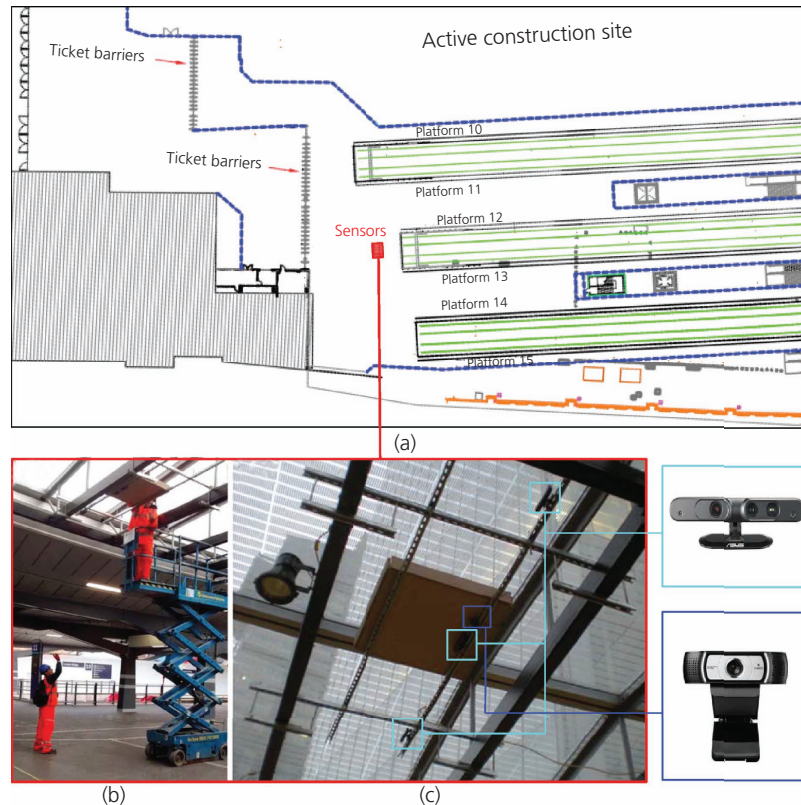
Figure 3. (a) Pedestrian counter sensor installation area in London Bridge station, (c) the utilised sensors and (b) their arrangement

The total equipment cost was less than £1000, with the PC accounting for £200, the three depth sensors £350 and the RGB camera £30. The pedestrian counting was trialled over February–March 2015. During this time, the hardware and counting algorithms operated continuously, demonstrating their robustness over time with variable temperatures. Remote access to the PC through 3G allowed periodic transfer of counting data and images for backup and off-site analysis.

3.2 Validation of the automatic counting approaches
Results from sensors are validated by comparing pedestrian count estimates against ground truth, where 'ground truth' refers to the manually counted estimates of people passing under the sensors over the time intervals of interest. The manual counting is performed using the image frames recorded from the RGB camera. This was achieved by (*a*) selecting the image sequences from the time intervals that represent the most relevant stress conditions (i.e. time windows with different exposure to ambient sunlight and crowd concentration) and (*b*) performing the manual count of people passing the counting line of the sensors by advancing one frame at the time. The whole operation was conducted by one person in approximately 6 h by using a script written for this specific purpose, while some counts were repeated several times for validation.

Results from both types of sensors were compared with the manual count results for 14 time intervals over 7 d with varying ambient sunlight and crowd conditions (see Figure 4). For each day, two 10 min intervals were chosen. During these intervals, passengers may be travelling in either direction between the station entrance and platforms 12, 13, 14 and 15. The first slot is in the morning from 9 to 10 a.m., when ambient sunlight is strong and arriving passenger numbers are typically high. In some of these slots, direct sunlight may also be observed. The next slot was chosen in the afternoon rush hour from 6 to 7 p.m., where the ambient and direct sunlight influence on the scene is negligible. However, during this time, the station is well lit by artificial lighting and is used by a considerable number of passengers who travel to and from London Bridge station. Results of this validation exercise are reported in the next section.

4. Results and discussion
This section compares and contrasts the passenger counting data from the two approaches described in Section 2. Both techniques have demonstrated good accuracy in counting individual pedestrians with different physical characteristics. In particular, the accuracy of the RGB camera-based system can be visually demonstrated with an output tracking video, which is attached as Supplementary Material (see also Stent (2015a)). In the video, the counting line is shown as a thick horizontal line, the marker × represents identified passengers and the green lines show their direction. The passenger direction line becomes red as passengers pass through the crossing line to indicate that they are exiting the platform. Correspondingly, the line becomes
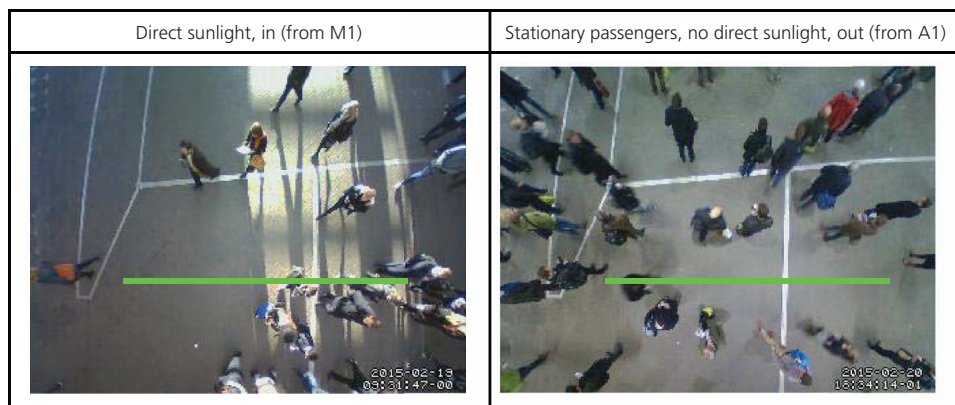
Figure 4. Snapshot samples from the RGB camera showing various capturing conditions. The horizontal line shows the extent of the crossing line: when pedestrians cross over this line, both systems considered them counted. The codes M1 and A1 refer to 2 of the 14 time intervals analysed in Table 1

blue when they are entering. At 0:01, a group of three people walking closely pass under the counting line. At 00:10 in the video, a passenger with a pushchair passes under the camera. At 00:39, a child and an adult walk underneath. At 00:45, a passenger with a rolling suitcase exits the platform, but does not pass under the counting line. In most cases, passengers walking with different speeds and paths were counted successfully. Rarely, passengers can be in the camera field of view for only a handful of pictures, due to either their high speed or a particular path. In these cases, passengers could not be identified quickly enough and this resulted in false negatives. The errors introduced by these occurrences were not significant and can be remedied with a higher frame rate and an RGB camera with a wider field of view. False positives caused by rolling suitcases, pushchairs or bicycles were also observed. However, these incidents were relatively rare. In the case of the depth sensing approach, this was because such objects are close to the ground and below the threshold to be considered a pedestrian. In the case of the RGB-based method, this was because such objects typically move very closely and in tandem with a human and hence tend to be counted along with their owners as a single pedestrian. However, the registration of close objects as a single feature caused the RGB-based method to undercount people who are walking together, such as arm-in-arm couples. The influence of such errors on counting accuracy will be discussed later.

In Figure 5(b), passenger counts over a two week period are reported for both sensors. A clear cyclical pattern over the course of each week is evident, where a drop in passenger counts characterises weekends. Averaging these charts over the course of a day (Figure 5(a)), it can be seen that the rise in pedestrian numbers is largely due to pedestrian influx during morning hours, whereas afternoon hours indicate fewer people arriving and more people leaving the station at this time.

Figure 5 also demonstrates significant differences between the counts registered by the sensors. A clear pattern emerges from the data – that is, the counting difference between the two sensors is most significant during daylight hours, particularly during the morning, when direct sunlight effects may be prominent. In addition, notable discrepancies are also observed for the afternoon hours, where passenger crowding is significant. Table 1 explores this issue further by examining counting data from these two periods. In particular, valuable information is obtained from comparisons with manual counts, which provide the 'ground truth'. The counting data from 14 videos, each of 10 min duration, which were used to evaluate counting accuracy are reported in Table 1, where M1–M7 stand for the seven morning intervals and A1–A7 stand for the seven afternoon intervals.

Results of the comparison in Table 1 demonstrate that the RGB-based system performs better than the depth-based system on average. This can be quantified by taking the ratio of the estimated and observed passenger counts and expressing it as a percentage for each investigated case (Table 1). The absolute percentage error across all investigated cases may then be summed and averaged to yield a measure of the system's counting accuracy. When using this approach, average errors of 64·9 and 10·0% are obtained for the depth-based system, for morning and afternoon hours. The corresponding errors for the RGB-based system are 21·6 and 9·0%. These values demonstrate that the accuracy of the depth-based system suffers strongly with increasing levels of ambient sunlight. However, the system demonstrates accuracy similar to that of the RGB-based approach in the case of artificial and uniformly lit scenes in the afternoon hours.

It is informative to identify the reasons for counting errors observed in both approaches in further detail. Table 1 indicates
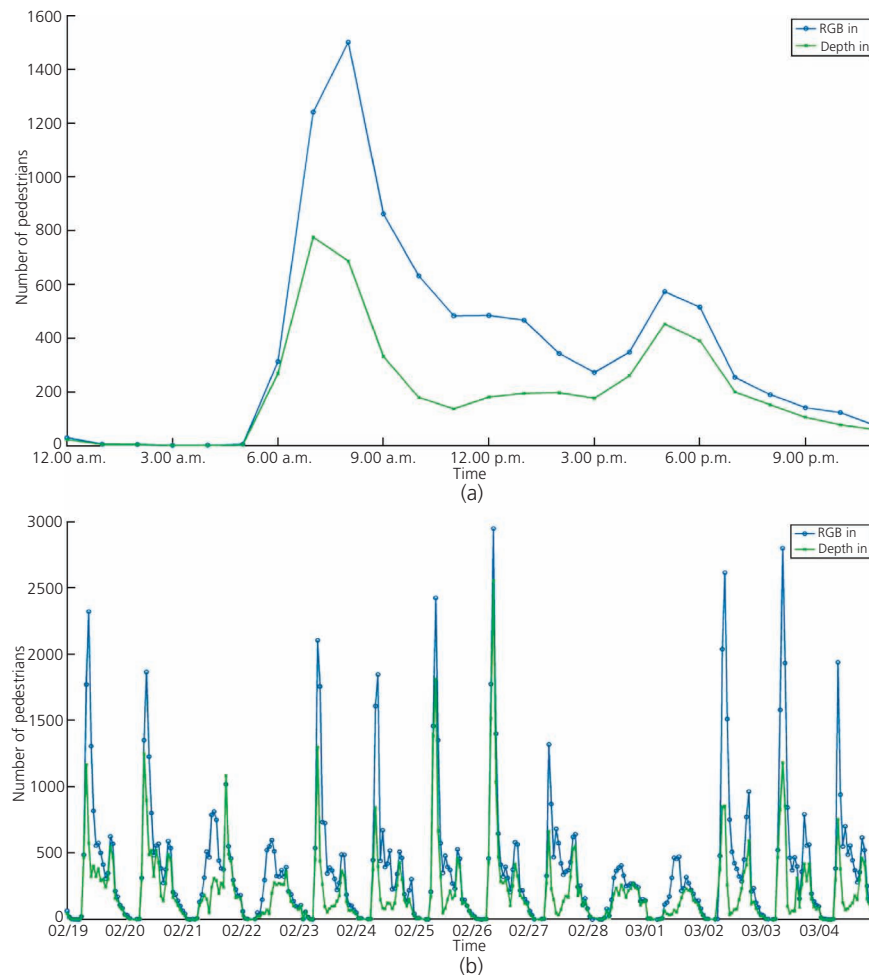
22

Figure 5. (a) The average hourly depth-based system and RGB-based system count for a day and (b) the daily total count over the 2-week monitoring period. Both graphs refer to a one-direction count only (the flow in, i.e. from the gates to the platform)

that direct sunlight influence is evident for the data sets M1, M2, M4 and M5. In particular, case M1 features the brightest scene (see Figure 4), where half of the image is flooded by direct sunlight. Therefore, it is not surprising that the lowest counting accuracy (7·7%) for the depth-based system is obtained in this sequence. In contrast, cases M3, M6 and M7 represent data taken from cloudy days. The error with depth sensors for these cases are 55·4, 33·7 and 22·4%, which represent a significant improvement over the cases with direct sunlight. However, it is noteworthy that the counting accuracy for depth sensors in the afternoon hours is consistently superior to the counting accuracy for cases during morning hours without sunlight. During the afternoon, the effects of ambient sunlight, which interferes with the infrared speckle pattern, disappear, and the count estimates with depth sensors become more reliable.

The presence of ambient sunlight is handled better by the RGB camera, where the average error is 21·6% in the morning.

However, the presence of direct sunlight also affects the performance of the RGB-based approach. In particular, the quickly changing lighting conditions and shadows result in the misidentification of pixels representing passengers, which results in erroneous tracking in the following frames. In these cases, most of the passengers travelling from platforms 14 and 15 on the right (sunny) side of the image (see Figure 4) could not be tracked reliably. This explains the undercounting of the RGB based method for data sets M1 and M2. The counting accuracy is better for sequences M4 and M5, where direct sunlight effects are also observed. However, in these cases, passengers were travelling less on the sunny side and tracking performance improves. The aforementioned factors, such as false positives due to objects such as rolling suitcases and false negatives due to groups of people travelling together, also affect counting accuracy. These explain the undercounting and overcounting of the RGB-based method for the data sets M3, M6 and M7. Here, counting errors up to 15% are observed due to these factors.

Smart Infrastructure and Construction
Volume 170 Issue SC2

**Pedestrian monitoring techniques for
crowd-flow prediction**
Martani, Stent, Acikgoz *et al.*

Table 1. The two counting systems compared against manual counting over various 10 min intervals to examine their absolute and relative accuracy under different conditions

Data set name	Scene conditions			Counts			Error against the ground truth	
	Dominant direction and entry location	Direct sunlight	Stationary passengers	Ground truth	Depth	RGB	Depth: %	RGB: %
M1	Out, right	Yes	No	635	49	382	92·3	39·8
M2	Out, centre-right	Yes	No	241	51	75	78·8	68·9
M3	Out, centre-right	No	No	74	33	78	55·4	−5·4
M4	Out, centre-right	Yes	No	542	75	465	86·2	14·2
M5	Out, centre-right	Yes	No	394	56	406	85·8	−3·1
M6	Out, right	No	No	392	260	375	33·7	4·3
M7	Out, centre-right	No	No	732	568	617	22·4	15·7
Morning average							64·9	21·6
A1	In, centre-left	No	Yes	313	281	347	10·2	−10·9
A2	In, centre-left	No	No	94	68	101	27·7	−7·5
A3	Out, centre-right	No	No	334	315	276	5·7	17·4
A4	In, centre-left	No	Yes	365	429	365	−17·5	0·0
A5	In, centre-left	No	Yes	332	304	302	8·4	9·0
A6	In, centre-left	No	Yes	294	293	318	0·3	−8·2
A7	In, centre-left	No	Yes	459	459	413	0·0	10·0
Afternoon average							10·0	9·0

The last two columns report the system counts as a percentage of ground truth counts: positive values indicate a net undercounting (i.e. system false negatives), while negative values indicate net overcounting (i.e. system false positives)

A similar investigation can also be made for the afternoon data sets A1–A7. Here, the average accuracy of the two systems are similar and above 90%. From these results, it appears that crowding does not significantly affect the counting performance. However, counting errors up to 30% are observed for the depth-based system, whereas counting errors up to 20% are observed for the RGB-based system. For the latter, these errors are higher than the 15% error previously identified for the cloudy morning hours. These slightly increased errors are attributed to the influence of stationary passengers, who wait for their train in the area leading to the platforms (see Figure 4). These passengers often cross back and forth across the crossing line over the course of several minutes while waiting, leading to overestimation of the count. In other cases, when passengers walk in close vicinity to the stationary passengers, passengers' paths cross. This can be illustrated with a count video which is included as Supplementary Material (see also Stent (2015b)). The video demonstrates an application of the RGB-based counting method for data set A1. The meaning of markers and lines superposed on the video is explained in the first paragraph of this section. In the video, note the stationary passengers who wait for their train near the counting line. They cannot be reliably tracked and are counted several times. In crowded scenes, their presence may affect the tracking of the path of other passengers. However, the data in Table 1 demonstrate that the influence of stationary passengers does not cause a systematic increase or decrease in the count estimations, and a reasonable accuracy is achieved despite errors.

In order to generalise the individual comparisons between automatic pedestrian counts, in Figure 6 the depth-based system counts are compared with RGB-based system counts within a

20 min period during peak (Figure 6(a)) and off-peak hours (Figure 6(b)). The peak hours are specified as 7–10 a.m. and 4–7 p.m. The results show that, in comparison with RGB, the depth-based system undercounts pedestrians by about 15% in peak and off-peak hours in the afternoons. It is noteworthy that the undercounting trend is similar for off-peak and peak conditions. Therefore, it can be argued that crowding does not affect count accuracy systematically. In addition, Figure 6 reiterates how increased ambient light renders the depth-based system counts unreliable in the peak and off-peak morning hours. While the trend lines indicate a ratio of 1·5–2·0 between RGB-based and depth-based counts, the significant variability of results should be noted. This significant variability is likely due to differing light conditions in the morning hours. The worst comparison was noted for a case where direct sun exposure resulted in RGB/depth counting ratios of approximately 25, while ratios close to 1 were achieved on cloudy days.

5. Conclusions

This paper has evaluated two efficient and inexpensive counting techniques and investigated their possible usage for real-time pedestrian simulations on crowded stations. The first technique utilises downward-facing infrared depth sensor hardware and a counting system based on volumetric recognition and feature-tracking abilities. The second technique utilises a similarly mounted wide field of view RGB camera with a counting system based on change detection, feature extraction and Kalman filtering. A suitable hardware and software system was configured, and the sensing system was deployed in the London Bridge station, which allowed an effective demonstration of the suitability of these techniques for crowded transportation hubs.
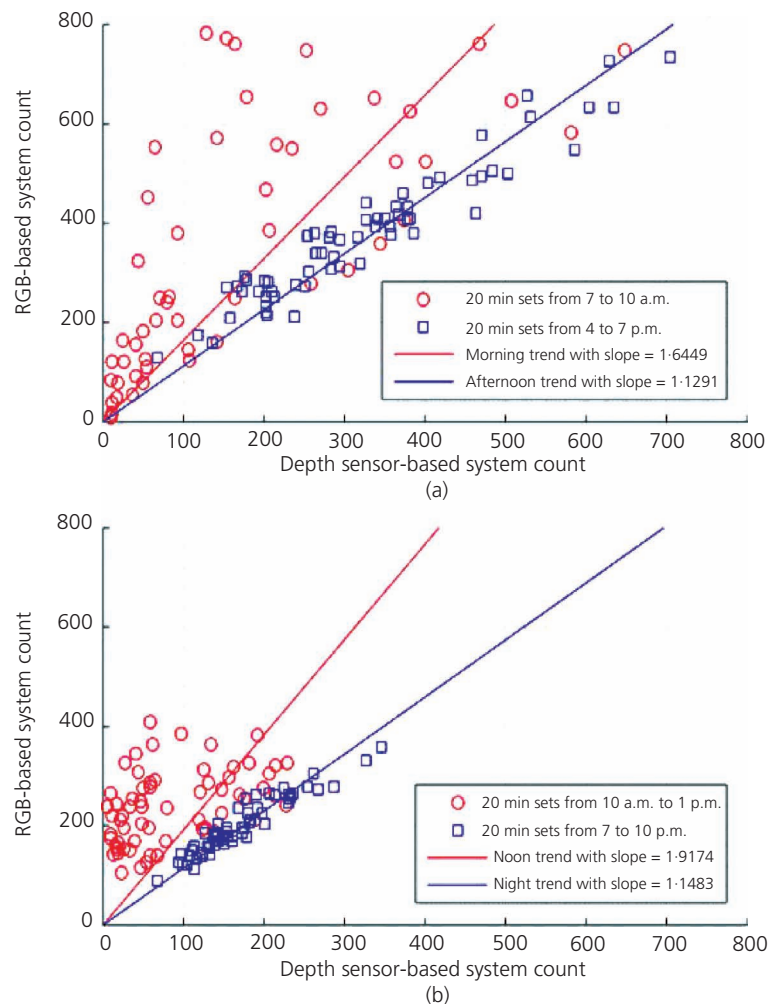
Smart Infrastructure and Construction
Volume 170 Issue SC2

**Pedestrian monitoring techniques for
crowd-flow prediction**
Martani, Stent, Acikgoz *et al.*

Figure 6. Comparison of the two counting approaches evaluated under (a) peak and (b) off-peak conditions

The counting algorithm and the hardware were stable and operated continuously for the duration of 2 months.

The focus of the paper is on understanding the sources of counting errors and evaluating the accuracy of the proposed real-time counting approaches by comparing their count estimations with manual counts. As expected, the use of depth sensors to investigate scenes affected by ambient sunlight during morning hours resulted in poor performance, with counting errors up to 90%. Comparisons to manual counts demonstrate this clearly. However, the performance of depth sensors improved drastically when counting passengers in the afternoon hours, when ambient light and direct sunlight effects were negligible, achieving typical counting errors smaller than 10%. Similarly, RGB cameras achieved errors less than 10% during afternoon hours. Crowding conditions did not affect the counting performance adversely, although it was found that improvements to the RGB-based system were required to improve the tracking for stationary passengers and direct-sunlight cases.

On the basis of these results, the following were concluded.

- Both counting technologies can provide reliable and accurate input for the simulations in the absence of strong ambient light, under any crowd condition.
- Under conditions of strong ambient light, the RGB-based system provides useful input for the simulations independently by the crowd, even though potential errors of *c.* 20% should be taken into account. The use of a depth–sensor-based system is not recommended in these situations.
- The two techniques complement each other. To investigate pedestrian flows in environments experiencing different ambient lights, the RGB sensors can be used. Alternatively, the depth sensors could provide useful counting information for dimly lit environments, where the RGB cameras may struggle to differentiate pedestrians.

Future work will explore the complementary nature of these counting technologies in further detail. In particular, the two systems

can be used together to improve count accuracy, by detecting false positives and negatives through the fusion of sensor information in a Kalman filter framework. For example, to improve RGB-based system individual pedestrian detection, depth information concerning the distance of the identified object from the ground or the distance of the object to its close neighbours may be investigated. This can help differentiate individuals among groups of people moving together and prevent false negative counts due to objects such as rolling suitcases. Moreover, the implementation of the current techniques with pattern recognition and machine learning approaches should also be explored to improve the stability of the counting under direct sunlight.

Further work is under way to expand this experiment by deploying counting technologies in critical points along the passenger routes in the station. Different counting technologies can be used to their strengths: for example, by using depth sensors in dimly lit areas without sunlight (i.e. on underground tunnels) and RGB cameras in locations subject to direct sunlight (i.e. on train platforms). Such a deployment would allow performing crowd simulations using real-time data and to autocheck their results each time by comparing the predicted outcome with the real count at destinations and refining the model calibration for subsequent simulations.

Acknowledgements

REFERENCES

Amin IJ, Taylor AJ, Junejo F, Al-Habaibeh A and Parkin RM (2008) Automated people-counting by using low-resolution infrared and visual cameras. *Measurement* **41(6)**: 589–599, http://dx.doi.org/10.1016/j.measurement.2007.02.010.

Batty M and Torrens PM (2001) Modeling complexity: the limits to prediction. *Cybergeo: European Journal of Geography*, http://dx.doi.org/10.4000/cybergeo.1035.

Bonnin A, Borràs R and Vitrià J (2011) A cluster-based strategy for active learning of RGB-D object detectors. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. Institute of Electrical and Electronics Engineers, Piscataway, NJ, USA, pp. 1215–1220, http://dx.doi.org/10.1109/ICCVW.2011.6130389.

Bu F, Greene-Roesel R, Diogenes MC and Ragland DR (2007) *Estimating Pedestrian Accident Exposure: Automated Pedestrian Counting Devices Report*. Safe Transportation Research and Education Center, Berkeley, CA, USA.

Chen X, Henrickson K and Wang Y (2016) Kinect-based pedestrian detection for crowded scenes. *Computer-aided Civil and Infrastructure Engineering* **31(3)**: 229–240, http://dx.doi.org/10.1111/mice.12163.

Chow TWS and Cho SY (2002) Industrial neural vision system for underground railway station platform surveillance. *Advanced Engineering Informatics* **16(1)**: 73–83, http://dx.doi.org/10.1016/S1474-0346(01)00002-7.

Dalal N and Triggs B (2005) Histograms of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA*, pp. 886–893.

Del Pizzo L, Foggia P, Greco A, Percannella G and Vento M (2016) Counting people by RGB or depth overhead cameras. *Pattern Recognition Letters* **81**: 41–50, http://dx.doi.org/10.1016/j.patrec.2016.05.033.

Dollar P, Wojek C, Schiele B and Perona P (2012) Pedestrian detection: an evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34(4)**: 743–761, http://dx.doi.org/10.1109/TPAMI.2011.155.

Hashimoto K, Kawaguchi C, Matsueda S, Morinaka K and Yoshiike N (1998) People-counting system using multisensing application. *Sensors and Actuators A: Physical* **66(1)**: 50–55, https://doi.org/10.1016/S0924-4247(97)01715-9.

Hu X, Zheng H, Wang W and Li X (2013) A novel approach for crowd video monitoring of subway platforms. *Optik – International Journal for Light and Electron Optics* **124(22)**: 5301–5306, http://dx.doi.org/10.1016/j.ijleo.2013.03.057.

Hu X, Zheng H, Chen Y and Chen L (2015) Dense crowd counting based on perspective weight model using a fisheye camera. *Optik – International Journal for Light and Electron Optics* **126(1)**: 123–130, http://dx.doi.org/10.1016/j.ijleo.2014.08.132.

Kalman RE (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* **82(1)**: 35–45, http://dx.doi.org/10.1115/1.3662552.

Li JP (2000) Train station passenger flow study. In *2000 Winter Simulation Conference Proceedings*. Institute of Electrical and Electronics Engineers, Piscataway, NJ, USA, vol. 2, pp. 1173–1176.

Li W and Zhu W (2016) A dynamic simulation model of passenger flow distribution on schedule-based rail transit networks with train delays. *Journal of Traffic and Transportation Engineering (English Edition)* **3(4)**: 364–373, https://doi.org/10.1016/j.jtte.2015.09.009.

Li X, Wang K, Wang W and Li Y (2010) A Multiple object tracking method using Kalman filter. In *The 2010 IEEE International Conference on Information and Automation*. Institute of Electrical and Electronics Engineers, Piscataway, NJ, USA, pp. 1862–1866.

Martani C, Stent S, Acikgoz S *et al.* (2015) Real time station monitoring and modelling for peak flow management. Results of tests conducted on technology deployment at London Bridge Station and plans for future actions. *AUM 2015*, Cambridge, UK.

Munkres J (1957) Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics* **5(1)**: 32–38, http://dx.doi.org/10.1137/0105003.

Sacchi C, Gera G, Marcenaro L and Regazzoni CS (2001) Advanced image-processing tools for counting people in tourist site-monitoring applications. *Signal Processing* **81(5)**: 1017–1040, https://doi.org/10.1016/S0165-1684(00)00280-2.

Schofield AJ, Stonham TJ and Mehta PA (1997) Automated people counting to aid lift control. *Automation in Construction* **6(5)**: 437–445.

Stogdale N, Hollock S, Johnson N and Sumpter N (2003) Array-based infra-red detection: an enabling technology for people counting, sensing, tracking, and intelligent detection. In *SPIE Proceedings* (Carapezza EM (ed.)). SPIE, Orlando, FL, USA, vol. 5071, pp. 465–475.

Stent S (2015a) https://www.youtube.com/watch?v=ME0FAqR4Z6Y (accessed 05/05/2017).

Stent S (2015b) https://www.youtube.com/watch?v=iPHZYCvvfgY (accessed 05/05/2017).

Tian Y, Luo P, Wang X and Tang X (2014) *Pedestrian Detection Aided by Deep Learning Semantic Tasks*. Cornell University, Ithaca, NY, USA. See http://arxiv.org/abs/1412.0069 (accessed 05/05/2014).

Timmermans HJP (2009) *Pedestrian Behavior: Models, Data Collection and Applications: Models, Data Collection and Applications*. Emerald, Bingley, UK.

Vera P, Monjaraz S and Salas J (2016) Counting pedestrians with a zenithal arrangement of depth cameras. *Machine Vision and Applications* **27(2)**: 303–315, http://dx.doi.org/10.1007/s00138-015-0739-1.

Xu Y, Cai-nian L, Xiao-liang X, Mei J and Jian-guo Z (2015) A two-stage hog feature extraction processor embedded with SVM for pedestrian detection. In *2015 IEEE International Conference on Image Processing (ICIP)*. Institute of Electrical and Electronics Engineers, Piscataway, NJ, USA, pp. 3452–3455.

Zachariadis V (2005) An agent-based approach to the simulation of pedestrian movement and factors that control it. *Computers in Urban Planning and Urban Management (CUPUM), London, UK*.

Zhang X, Yan J, Feng S *et al.* (2012) Water filling: unsupervised people counting via vertical kinect sensor. In *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance)*. Institute of Electrical and Electronics Engineers, Piscataway, NJ, USA, pp. 215–220.

How can you contribute?

To discuss this paper, please email up to 500 words to the editor at journals@ice.org.uk. Your contribution will be forwarded to the author(s) for a reply and, if considered appropriate by the editorial board, it will be published as discussion in a future issue of the journal.

Proceedings journals rely entirely on contributions from the civil engineering profession (and allied disciplines). Information about how to email your paper online is available at www.icevirtuallibrary.com/page/authors, where you will also find detailed author guidelines.