Automatic Surveillance in Transportation Hubs: No Longer Just About Catching the Bad Guy

Simon Denman[a,b], Tristan Kleinschmidt[a], David Ryan[a,b], Paul Barnes[a], Sridha Sridharan[a,b], Clinton Fookes[a,b]

[a]*Airports of the Future Project, Queensland University of Technology (QUT), GPO Box 2434, 2 George Street, Brisbane, QLD 4001*
[b]*Image and Video Laboratory, Queensland University of Technology (QUT), GPO Box 2434, 2 George Street, Brisbane, QLD 4001*

Abstract

As critical infrastructure such as transportation hubs continue to grow in complexity, greater importance is placed on monitoring these facilities to ensure their secure and efficient operation. In order to achieve these goals, technology continues to evolve in response to the needs of various infrastructure. To date, however, the focus of technology for surveillance has been primarily concerned with security, and little attention has been placed on assisting operations and monitoring performance in real-time. Consequently, solutions have emerged to provide real-time measurements of queues and crowding in spaces, but have been installed as system add-ons (rather than making better use of existing infrastructure), resulting in expensive infrastructure outlay for the owner/operator, and an overload of surveillance systems which in itself creates further complexity. Given many critical infrastructure already have camera networks installed, it is much more desirable to better utilise these networks to address operational monitoring as well as security needs.

Recently, a growing number of approaches have been proposed to monitor operational aspects such as pedestrian throughput, crowd size and dwell times. In this paper, we explore how these techniques relate to and complement the more commonly seen security analytics, and demonstrate the

Email addresses: `s.denman@qut.edu.au` (Simon Denman), `kleinschmidt__11_tf@yahoo.com.au` (Tristan Kleinschmidt), `david.ryan1@gmail.com` (David Ryan), `p.barnes@qut.edu.au` (Paul Barnes), `s.sridharan@qut.edu.au` (Sridha Sridharan), `c.fookes@qut.edu.au` (Clinton Fookes)

value that can be added by operational analytics by demonstrating their performance on airport surveillance data. We explore how multiple analytics and systems can be combined to better leverage the large amount of data that is available, and we discuss the applicability and resulting benefits of the proposed framework for the ongoing operation of airports and airport networks.

Keywords: Transportation, Security, Operational performance, Video surveillance, Operational analytics, Airport Operations

1. Introduction

The scale of challenges facing society in providing more advanced critical infrastructure, and in particular transportation hubs, is substantial. An increased importance has been placed (and will continue to be placed) on transportation hubs to accommodate increased demand as cities continue to expand, and global air travel becomes more accessible. Consequently, the complexity of these hubs is increasing; airports are a perfect example where a multitude of factors are continually in play, including technological advancements, changes in regulations, and the interaction of multiple stakeholders including (but not limited to) government agencies, airport operators, airlines, security contractors, commercial operators, and of course, the travelling public (Ashford et al., 2013). With this increase in demand and complexity, it can be argued that it is becoming more important to monitor the operational performance of the system. This is especially true for many privatised airports who often depend on non-aviation revenue sources to a greater extent than traditional aviation revenue.

Unfortunately, with millions of passengers passing through these hubs on a daily basis, the infrastructure itself become prime targets for terrorist activities (Tsai et al., 2009). Recent examples of such attacks on transportation hubs include the London train bombings in 2005, the Glasgow Airport car bombing in 2007, the Domodedovo International Airport bombing in 2011; the Peshawar airport attack in 2012l and the Jinnah International Airport attack in 2014. With this constant threat in place, it is extremely important to ensure transportation hubs are safe and secure for all involved.

Over the past decade, surveillance cameras have become commonplace in public locations, including transportation hubs (Wells et al., 2006; Welsh and Farrington, 2009; Adrem et al., 2007). This is a direct result of an increased focus on public safety and security, but can also be attributed to the reduced cost of cameras and their associated computing infrastructure which help to keep overall security costs down (Adrem et al., 2007). The increase in surveillance cameras has given rise to an increase in computer systems to manage, and in many cases to analyse the incoming camera feeds. At present, these feeds are typically monitored by staff, and detecting events as they happen is very challenging due to the sheer amount of data being presented to each operator.

To assist human operators in monitoring large CCTV networks, there has been a significant increase in computer vision research and development, to create algorithms to analyse and extract information from the CCTV feeds. These developments have tended to focus on security related tasks such as object/person tracking, perimeter surveillance, motion segmentation, abnormal event detection and recognition, and biometrics (e.g face, iris, fingerprint) for person identification (Fookes et al., 2010). Many of these algorithms have begun to be implemented within video management and analytic systems, giving commercial video analytic packages a wide range of (primarily) security capabilities.

Operational analytics however, such as crowd counting and queue monitoring, have received less attention and while commercial systems do exist to perform these tasks, they are fewer in number and often require specially placed cameras making them difficult to integrate with existing systems. This is despite the comparatively poor performance of security based systems, which are prone to missed detections and false alarms.

Furthermore, the integration of these emerging techniques into airports or other critical infrastructure has received limited attention. When discussing the implementation of surveillance techniques, existing research has focussed on how a single surveillance task may function within a piece of infrastructure in isolation (i.e. Li et al. (2014) considers person re-detection in an airport environment; while Arroyo et al. (2015) consider suspicious behaviour detection in a shopping mall. Similarly, when considering the system wide implications of video surveillance, the literature has focussed on aspects such as the data and networking requirements of such large scale systems (Ajiboye et al., 2015; Chang et al., 2012); interfaces to retrieve and display results (Ye et al., 2015); or the needs of researchers and developers to aid in the development of such techniques (Nazare et al., 2014).

Within this paper we propose an automated surveillance framework for both operational and security tasks for on-site and across-site monitoring.

Whilst ensuring that a wide range of possible surveillance technologies are included in this framework, we specifically discuss the overlap between video analytics for security and operational analytics for operational monitoring which can be exploited to make better use of CCTV networks in public spaces. We present an overview of intelligent surveillance techniques for security and operational tasks, and show that although security has long been the focus of surveillance deployments, the operational video analytics currently in development are in many ways, more appropriate for deployment.

We show, on airport surveillance data, how recent approaches can be used and combined to extract measures of operational performance such as crowd sizes, processing rates and dwell times. The performance of these approaches, as well as their strengths and weaknesses from a real-world standpoint (i.e. deployment requirements and challenges) are discussed, and we explore how these techniques can be used in tandem with other statistical modelling approaches to provide better situational awareness. To demonstrate how such a combined security and operations framework could benefit a transport hub, we develop a case study around airport security, incident response and level of service monitoring to demonstrate the potential of video analytics as a solution to both these needs.

The remainder of this paper is structured as follows: Section 2 presents an overview of intelligent surveillance and provides an outline of the current abilities of security analytics; Section 3 presents our proposed surveillance framework, and an overview of operational analytics and how they can be applied to a transport environment; Section 4 presents two case studies examining how our proposed framework could be applied to an airport environment; and Section 5 concludes the paper.

2. Automatic Surveillance: A History in Security

Surveillance systems are an essential and integral component in transportation networks, public places, and other critical infrastructure where it is necessary to monitor activities, threats, and to prevent or investigate criminal or other unwanted activity.

An increased focus on security coupled with falling costs of hardware has seen an increase in the number of CCTV management products available. Some systems (such as Iomniscient [1], BlueEye Video [2], Agility Video [3], ObjectVideo [4]) also offer video analytics: algorithms which can extract information from the incoming feeds in real-time or near real-time. The capabilities of such products vary significantly and as such, it is helpful to categorise them through a high-level classification outlined below.

- 1st Generation: Traditional analogue CCTV systems with recording facilities through tape or digital video recorders.
- 2nd Generation: Highly capable "Video Management Systems" utilising large IP networks (cameras may be digital or analogue with encoders). These systems have a suite of low-level image processing tools (such as perimeter intrusion detection, loitering, abandoned object detection, etc).
- 3rd Generation: True multi-view capable intelligent surveillance systems with robust semantic information extraction.

We argue that most commercial solutions are still only 2nd generation systems (with a select few 2.5 generation) and are often characterised by high false-alarm rates, and limited knowledge of the environment in which they are deployed (i.e. camera calibration). A significant advancement in capabilities is still required before 3rd generation systems are reached, i.e. "cognitive" systems that can track, identify and explain what is taking place (Bellotto et al., 2009). This includes the development of: true multi-view capability (rather than single-view with simple camera network topologies); automatic camera calibration; robust tracking and recognition of people and events that are invariant to the challenging day-to-day operating conditions including illumination, pose, viewpoint; and invariance to noisy, cluttered complex environments. Recent advances in deep learning and convolutional neural networks indicate one direction that may advance these goals. Significant gains have been made in fields such as speech recognition (Deng and Yu, 2014), natural language processing Manning et al. (2014), object recognition Erhan et al. (2014) and pedestrian detection (Luo et al., 2014) by

[1] http://iomniscient.com/

[2] http://www.blueeyevideo.com/

[3] http://www.vidient.com/

[4] http://www.objectvideo.com/

leveraging very large amounts of data to automatically learn complex relationships within the data. The data requirements of deep-learning have to date meant that it's applications has been restricted to data rich domains; however it offers a promising direction to enable the development of true '3rd Generation' systems.

As products move towards becoming 3rd generation, many video analytics are becoming incorporated and/or integrated into video management systems. The majority of analytics can broadly be classified into two groups:

1. Security - such as (but not limited to) perimeter protection, suspect tracking, loitering detection, and abandoned luggage detection;
2. Operational - such as (but not limited to) crowd counting, queue length estimation, throughput analysis, service rates, utilisation rate estimation, trajectory analysis, and travel time estimation.

There is some overlap between the two categories of analytics, for example crowd counting can be used to obtain both operational measures as well as indicate when the crowd has reached a dangerous size. Despite this overlap of capabilities, surveillance systems are still seen primarily as a security tool, and as such the majority of the analytics available in commercial products are security related. It is also important to note that other technologies aside from CCTV analysis are being used for both operational and security tasks. Wireless sensors such as BlueTooth or WiFi can be used to track assets or measure performance (Shen et al., 2008; Versichele et al., 2012; Woo et al., 2011; Patil and Kokil, 2015) (either by tracking a dedicated tag or another device that incorporates the technology such as a smart phone). Other sensor based technologies such as RFID are becoming widely adopted for inventory management problems, such as luggage tracking within the transport domain (Medeiros et al., 2011; Ting et al., 2006; Zhang et al., 2008). At present, technologies such as these are typically implemented independently of any CCTV systems and are restricted to operational tasks, while CCTV is used for security.

A broad range of video analytics are available for security, however, they can be broadly classified as follows:

1. Object Tracking Derived - Analytics such as perimeter intrusion and out of bounds detection, loitering, and tail-gaiting detection;
2. Change Detection Derived - Analytics such as abandoned object detection, stopped vehicle detection, and theft detection.

Furthermore, other key surveillance tasks are attracting significant attention from the research community, including:

1. Event recognition in crowded scenes;
2. Person re-detection and a semantic search for a specified person.

While these techniques are less mature than those that underpin tracking and change detection analytics, applications of these are beginning to filter into commercial systems, such as Iomniscient's detection of slips and falls.

A brief overview of these four areas is presented in Sections 2.1, 2.2, 2.3 and 2.4.

2.1. Object Tracking

Object tracking is the task of continuously detecting an object through a video sequence, and can be broken down into two types:

1. Single object tracking;
2. Multi-object tracking.

Single object tracking is the problem of following a region of interest in video. Typically, an initial detection is provided by a user, and this region is then tracked through a video clip. Two broad approaches exist for this problem: generative trackers (Liu et al., 2010; Zhang et al., 2015b; Sevilla-Lara and Learned-Miller, 2012; Felsberg, 2013) which seek to find a region that best matches a target model; and discriminative trackers (Hare et al., 2011; Danelljan et al., 2014; Henriques et al., 2015) which treat the tracking problem as a binary classification task, and train a model to classify a region as being the target of interest.

Generative approaches seek to build a model to represent a target's appearance. Multi-channel representations such as a distribution field (Sevilla-Lara and Learned-Miller, 2012), or channel representation (Felsberg, 2013) have proven successful, while sparse representation has also been very effective. Sparse representation allows the target's appearance to be modelled as a linear combination of templates that can be updated on-line to allow for appearance changes (Liu et al., 2010). A recent approach by Zhang et al. (2015b) sought to extend this further to better incorporate the underlying target structure by learning a joint dictionary to represent both the overall and part-wise appearance of the target object, allowing the overall appearance and structure to be incorporated, improving performance.

Discriminative approaches (Hare et al., 2011; Danelljan et al., 2014; Henriques et al., 2015) have typically used intensity based features due to the demands of training and updating a classifier each frame. Although recent research by Danelljan et al. (2014) demonstrated how low dimensional colour features could be incorporated to improve performance; while Henriques et al. (2015) proposed a highly efficient correlation filter that leveraged the circulant nature of the data to greatly improved computational efficiency and storage, and thus enabled the use of multiple channels to further improve performance.

Multi-object tracking is typically a two stage problem, where detection of objects is followed by a matching and updating step. Detection may be performed either through the analysis of motion features (Huang and Barth, 2010; Ottlik and Nagel, 2008; Zhao and Nevatia, 2004; Lu and Tan, 2001; Haritaoglu et al., 2000), or through the detection of a previously learned model (**?**Tamersoy and Aggarwal, 2009; Yang et al., 2005; Okuma et al., 2004). The available methods to match and update the tracked objects are many and varied, and depend on the features being extracted from the objects. Various geometric features (i.e. object position, size), colour and appearance features (i.e. histogram), edge- or graph- based features can be used, either in isolation or combination (**?**Denman et al., 2006a; Haritaoglu et al., 2000).

Technologies such as the detection of people inside 'out-of-bounds' areas, and loitering detection can all be achieved through object tracking, by configuring simple rules to trigger alerts within the object tracking system. For instance:

- A given area may be marked as 'out-of-bounds', if a person enters this region they are an intruder.
- If a person is observed and tracked for a long period of time within a single area, they are loitering.

Within the research space, there has been much work focused on object tracking. This research has covered person tracking (**?**Zhao and Nevatia, 2004; Lu and Tan, 2001; Haritaoglu et al., 2000), vehicle tracking (Huang and Barth, 2010; Brulin et al., 2010; Tamersoy and Aggarwal, 2009; Ottlik and Nagel, 2008; Denman et al., 2006b), tracking multiple types of objects (Denman et al., 2006a), handling of groups of objects (Bazzani et al., 2015; Kooij et al., 2015; Denman et al., 2010; Galoogahi, 2010; Haritaoglu et al., 1999), tracking people in a sports environment (Vermaak et al., 2003; Okuma et al., 2004), and recently tracking people in crowds (Tang et al., 2015; Ben Shitrit et al., 2014; Eshel and Moses, 2008; Pirsiavash et al., 2011; Berclaz et al., 2011).

The primary limitation of tracking systems is their ability to handle large (or even moderately sized) crowds effectively, with occlusions and people crossing paths causing frequent errors. The approach of **?**, which combined the detector confidence from multiple object detection routines with object specific classifiers in a particle filter framework, showed progress in addressing these issues, demonstrating improved performance over a variety of other techniques (Huang et al., 2008; Leibe et al., 2008; Okuma et al., 2004), across a suite of databases that include small crowds and frequent occlusions. However, **?** still has a false negative rate (i.e. missed detections and tracks) of 15%-30% for most databases, and ID switches are still common in several data sets which would likely lead to frequent errors if used in a live analytics system. Explicitly modelling groups has been pursued to obtain further improvement, with Bazzani et al. (2015) proposing the the joint modelling of individuals groups jointly using a decentralised particle filter. This allows the groups and individuals to be tracked within the same framework, with information shared between the two processes. Kooij et al. (2015) proposed the use of Latent Dirichlet Allocation (LDA) to resolve unreliable object detections into the true individual targets that generated them. By effectively treating the targets in the scene as the 'topics' within the LDA model, the targets present in each frame can be detected, and the LDA model can be back-projected into the source frame to segment the individual targets.

An alternate approach is to treat the tracking problem as an optimisation task, where the goal is find the optimum configuration of trajectories across a video sequence for a given set of detections (Pirsiavash et al., 2011; Berclaz et al., 2011; Ben Shitrit et al., 2014; Tang et al., 2015). Such approaches offer increased robustness to detection errors, however they are non-causal in nature, making such techniques difficult to apply in a live environment. These approaches typically first form tracklets (i.e. tracks of a few frames length) based on spatial constraints, and then cluster these over time (Pirsiavash et al., 2011; Berclaz et al., 2011; Ben Shitrit et al., 2014), however this can lead to duplicate tracklets when multiple detections occur for a single person. Tang et al. (2015) overcame this and achieved improved performance by jointly clustering all observations in time and space.

Extensive research has also focused on issues relating to multi-camera tracking, and in particular being able to accurately re-detect people across multiple disjoint camera views. Various techniques based around colour (Bazzani et al., 2013) and texture (Bak et al., 2010) features have been proposed to re-detect people in different camera views. While these techniques show promise, performance is still limited with Bazzani et al. (2013) achieving Rank-1 and Rank-10 matching accuracies (for an identification task) of 20% and 50% for the Viper data set (Gray et al., 2007) (although, synthetic recognition rates of approximately 88% and 75% are achieved for 5 and 10 subjects respectively); while Bak et al. (2010) reported Rank-1 and Rank-10 accuracies of 41% and 80%, although it should be noted that this is for a much smaller database than Viper (44 subjects verses 632).

One major problem in person re-identification is that of pose changes, which cause a person to look different from different angles. Bak et al. (2015) proposed learning a metric pool, where each metric is designed to best match a pair of poses, as a method to improve performance in the presence of pose variation. By first estimating the pose of the image pairs to compare, improved performance could be achieved. Richer, more diverse feature sets offer another avenue to improve performance and Bedagkar-Gala and Shah (2014) explored how gait could be combined with appearance for recognition in surveillance imagery. The nature of gait, in that it can be obtained at a distance without cooperation from the subject, makes it appealing for surveillance. A sparse representation based method was proposed, which allowed gait to be used without incorporating view-angle estimation, and also allowed for missing data (in the event that gait could not be captured). It was shown that despite the difficulties in extracting gait information, gait could be used to improve re-identification performance.

A second major challenge in person re-identification is the open world nature of the problem. The majority of existing work assumes a closed-world scenario, i.e. every subject in one camera has a matching image in a second. This is a highly unrealistic view, and the recent work of both Cancela et al. (2014) and Kenk et al. (2015) has sought to address this and re-formulate person re-identification as an open world problem. Cancela et al. (2014) proposed a conditional random field inference based framework to deal with the open-world nature of the problem, by learning possible transitions between cameras to improve matching across complex networks. Kenk et al. (2015) proposed an on-line distributed system, that incorporated novelty detection and a forgetting mechanism to add new potential identities and remove those that are no longer required.

2.2. Change Detection

Change detection is the process of identifying medium to long term changes in the scene. This process is different from motion segmentation in that it is not concerned with moving objects; instead changes to the underlying background are of interest, such as an object missing (i.e. theft), or a new item being added (i.e. abandoned luggage).

The vast majority of change detection algorithms have been developed and demonstrated as abandoned object detection systems. Early systems used long term change detection (Sacchi and Regazzoni, 2000; Stringa and Regazzoni, 2000), double background subtraction (Herrero et al., 2003; Singh et al., 2009), or multiple layers of motion segmentation (Denman et al., 2007) to cope with occlusions; with recent methods adding filtering stages to reduce false alarms (Wen et al., 2009; López-Méndez et al., 2014; Nam, 2015). Algorithms designed for multi-camera environments have also been proposed (Auvinet et al., 2006; Krahnstoever et al., 2006), although these use a similar methodology to typical single camera approaches in that they simply use motion segmentation information to locate regions of interest and the data is aggregated across the multiple cameras (either prior to detection in the case of Auvinet et al. (2006), or after detection and tracking as in Krahnstoever et al. (2006)).

Typically these systems are able to achieve high detection rates (100% for Krahnstoever et al. (2006), 85% for Auvinet et al. (2006) and Singh et al. (2009)), although evaluations are limited and restricted to small data sets such as PETS 2006 Thirde et al. (2006), which only contains seven examples. It should be noted that a common weakness of all these approaches is that they are prone to false alarms in difficult conditions, such as when large numbers of people are standing still, or the abandoned luggage is frequently occluded.

To overcome the problems of background modelling based methods, a number of approaches to filter candidates have been proposed. Wen et al. (2009) refined a set of coarse candidates through classification with a generative model incorporating colour, shape, edge and texture features. This approach is shown to be able to detect over 85% of abandoned objects with no false alarms on a larger database of 29 examples. López-Méndez et al. (2014) proposed the use of prior information (geometric information that describes the ground plane locations, and detector output that provides the likely locations of people) to help reduce the number of false alarms generated by a multi-layer background subtraction method. The approach was shown to be able to detect all abandoned objects with no false alarms on the PETS2006 database. Nam (2015) used the relationship in time and space between moving and stationary objects to filter candidate regions, achieving high performance on a number of databases. While approaches such as these reduce the susceptibility to false alarms, the reliance on background modelling means that missed detections in crowded situations are still a significant problem.

2.3. Detecting Events in Crowds

As outlined in Section 2.1, tracking people in crowded scenes is challenging and often unreliable. As such, the monitoring of a crowded scene is often better served by using event detection techniques, that extract features from the scene (such as optical flow (Wang et al., 2009a) or optical flow derived features (Ryan et al., 2011a; Nallaivarothayan et al., 2014), particle trajectories (Xu et al., 2012a), or dynamic textures (Mahadevan et al., 2010)), and use these to model the interactions of individuals and the activities present in the scene.

While simple events such as a mass evacuation or panic (such as depicted in the UMN data set [5]) can be detected with high reliability (Xu et al. (2012b) and Mehran et al. (2009) achieved AUCs of 0.97 and 0.96 respectively), more subtle events such as a cyclist or skateboarder moving through a crowd of pedestrians as depicted in the UCSD data set (Mahadevan et al., 2010) are more difficult to detect. State-of-the-art approaches such as that of Nallaivarothayan et al. (2014) have reported for detecting these more subtle events (in terms of the equal error rate) of 14.9% and 4.89% for data sets 1 and 2 respectively, while Roshtkhari and Levine (2013) achieved an equal error rate of 15% for data set 1.

For real world events typical of a transport hub, where crowd levels are high and individuals are frequently occluded, the problem becomes even more challenging. The TrecVid Surveillance Event Detection (SED) evaluations [6] seek to evaluate the performance of event detection algorithms in a crowded airport environment, and aims to detect events such as a person running,

[5] This database can be found at http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi

[6] see http://www-nlpir.nist.gov/projects/trecvid/ for details on the TrecVid evaluations

people embracing, and a person using a cell phone in a crowded airport terminal. In such environments, performance is typically very poor. For the 2014 SED evaluation[7], the performance for majority of events was limited to a miss rate of approximately 80% at a cost of 10 false alarms an hour, although one system was able to detect almost 50% of the 'person running' events at a cost of approximately 5 false alarms per hour. Of note is that for all the events in the TrecVid SED evaluation, no system is able to detect more than 50% of the instances of any event, regardless of the false alarm rate, highlighting the difficulty in detecting such events in surveillance video.

One of the major obstacles for event detection research has been the data annotation requirements. Unsupervised approaches are popular for anomaly detection, but assume that the training data only contains normal events (and contains examples of *all* normal events); while supervised approaches need a large number of carefully annotated examples, localising the event both in time and space, for the target events. Recently, weakly supervised learning (Hospedales et al., 2011) has been proposed as a means to reduce the burden of data collection, by only requiring coarse annotation (typically approximate temporal segmentation) of the events. Weakly supervised learning allows a model for the target event to be learnt using data that contains the target event as well as a number of background events, simplifying the data collection process. To further reduce data needs, the recent approach of Xu et al. (2015) is designed to require only a small number of examples of the target event relative to the number of background examples.

2.4. Re-Detecting and Searching for People

A major security challenge is that of locating a person in an environment from a simple, semantic, description (i.e. 1.7m tall, blue shirt and grey trousers). To achieve this, a mapping must exist between the semantic traits and the real-world appearance of those features (i.e. we need to learn what a blue shirt looks like, and be able to detect it). It is only recently that researchers (such as Park et al. (2006); Vaquero et al. (2009); D'Angelo and Dugelay (2010); Satta et al. (2012); Denman et al. (2012)) have attempted to address this problem. However, while promising results have been obtained, techniques that are appropriate for real-world deployment are a long way off.

[7] see http://www-nlpir.nist.gov/projects/tv2014/active/tv14.workshop.notebook/tv14.sed.results/ for evaluation results

Many techniques (Park et al., 2006; Vaquero et al., 2009; Satta et al., 2012) have relied on the subject already being detected and segmented which is itself a challenging problem, especially in crowded scenes. Satta et al. (2012) proposed learning the appearance of various features (i.e. a red shirt, short trousers or a skirt), and demonstrated the recognition performance for these features using the Viper database (Gray et al., 2007). The break even point on precision-recall plots (the point where precision equals recall) is used to evaluate the performance of the proposed approach, with break-even point values of between 0.916 and 0.433 being achieved (note that for the break-even point, a score of 1 indicates perfect performance, while 0 indicates complete failure). This work was extended in Satta et al. (2014) to allow for people with specific attributes to the be located, although the requirement that the subject first be detected and localised remains.

Less constrained techniques have been proposed by D'Angelo and Dugelay (2010); Denman et al. (2012, 2015b); Halstead et al. (2014), that are aimed at allowing such a search to be performed in crowded scenes where person detection is not possible. D'Angelo and Dugelay (2010) sought to locate potential hooligans in a football stadium, and used colour descriptors and a patch-based approach to locate rival supporters congregating together. Denman et al. (2012) and Halstead et al. (2014) proposed approaches where an avatar based on size and colour features is used to locate a person in an image sequence without requiring people to be first detected. While this approach showed promise in that it could function in the presence of partial occlusions, it is troubled by ambiguities in colours and errors in segmentation. An alternate approach that built upon an efficient distribution field based single object tracking approach (Felsberg, 2013) was proposed by Denman et al. (2015b), who constructed a channel representation (CR) to describe the target subject. The CR allowed the distribution of colours, as well as (to a limited extent) the uncertainty in both the colours and their distribution, to be incorporated. This approach was shown to outperform that of Halstead et al. (2014), achieving good localisation of 74% of target subjects, compared to only 37% for Halstead et al. (2014).

3. The Emerging Role of Automatic Surveillance in Operations Management

A major limitation of existing surveillance systems is their focus on security at the expense of critical business performance data and operational information that can be extracted from the same system. Often, cameras placed to look for security threats in public areas (such as a fight or disturbance) are also able to observe an operational processes such as a queue, doorway or meeting place. These cameras can also be used to extract additional information of interest, such as the number of people in a space or queue, or the rate of people moving through a doorway. While this information may not provide any direct security information, it can be vital in determining how well the overall system is functioning, and assist in redeploying staff to deal with crowding, load balancing (Adrem et al., 2007), and forward planning.

Although some technologies do exist for monitoring queues and passenger movements, they rely on additional hardware and many require the passenger to possess a device which can be detected and tracked (such as a BlueTooth or WiFi enabled mobile phone). Furthermore, while it may be possible to capture this information from other sources, integrating this data with other security information - whether it is obtained by automatic or manual monitoring of CCTV feeds, from access control or from human observers - is far from trivial.

Emerging technologies, such as crowd counting from CCTV footage (Ryan et al., 2011b, 2015), are enabling surveillance systems to play a greater role in operations management. At present, existing commercial systems require specially configured cameras (i.e. placed directly overhead for a top-down view), however, current research is seeking to address this limitation and develop turn-key solutions. In addition, operational measures can also complement and enhance security. Techniques such as crowd counting, while offering a powerful tool for managing operations, also allows the detection of unusually large gatherings, and unusual crowd patterns which may indicate a security threat.

The complementary interests of security and operations gives rise to the surveillance framework presented in Figure 1. The main components of this framework are outlined in the following Sections:

1. Data Capture Layer (Section 3.1),
2. Intelligence Layer (Section 3.2),
3. Aggregation Layer (Section 3.3),
4. Reporting and Monitoring (Section 3.4).

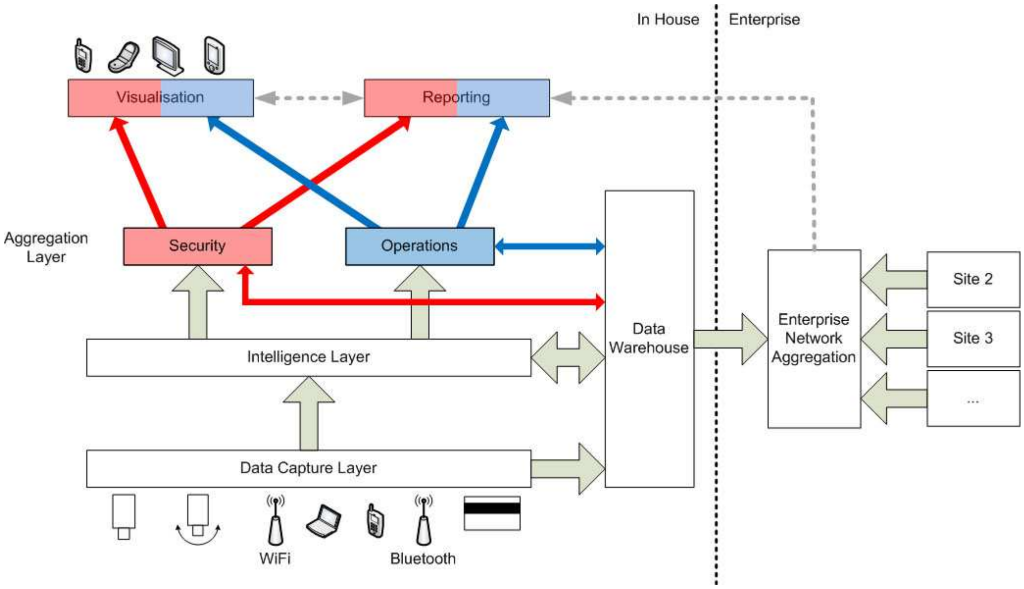


Figure 1: Diagram of the Proposed Surveillance Framework - Captured data is collected and stored, and also analysed by the intelligence layer. These results can be combined to produce both security and operational information, across both an individual site and a complete enterprise.

3.1. Data Capture Layer

Within a modern transportation environment there are many potential sources of data, from building access control systems to CCTV networks to mobile devices carried by staff and customers alike. Ideally, all these sources of data should be captured by a common framework, allowing both security and operational analytics algorithms to share the multiple data sources to improve results. Data sources that could be captured include: video feeds, audio feeds, satellite positioning systems (GPS) (Bandini et al., 2007), RFID inventory and asset management systems, WiFi and Bluetooth device ID's and locations, wireless cellular access device ID's and locations, building access control, and fire/heat alarm systems. Of these data sources, we consider the video feeds to be the richest source of data for the capture of detailed operational measures of interest.

There exists a large body of research concerned with the development of video analytics to extract information from video data, and there are several commercial systems that incorporate video analytics. We consider the other sources of data as complementary, and useful as a means to trigger an action within the video analytics (i.e. a door alarm being activated may result in a video analytic being used to locate and track the person who set the alarm off), or to provide support to the video analytics (i.e. a person may be tracked by both the video analytics, and other signals such as WiFi or BlueTooth through their mobile phone).

3.2. Intelligence Layer

Within current commercial systems, many (if not all) of the data sources outlined in Section 3.1 can be captured and stored, albeit not necessarily within the one system. However, at present there is very little automated processing of the data streams, aside from some video analytics typically focused on security (see Section 2). There is the potential for a great deal more data to be extracted from these diverse input streams, leading to increased situational awareness.

From an image processing standpoint, there is significant overlap between security and operational analytics in terms of the image processing tasks that are required. Figure 2 shows how a variety of video analytics can be built on a common set of underlying algorithms. Many object tracking algorithms (Tang et al., 2015; Ben Shitrit et al., 2014; **?**; Denman et al., 2010; Ottlik and Nagel, 2008; Zhao and Nevatia, 2004) rely on motion segmentation and/or optical flow, as well object detection. Similarly, other algorithms such as super-resolution (Lin et al., 2005, 2007; Fookes et al., 2012) or crowd counting algorithms (Ryan et al., 2010; Chan et al., 2009) also make use of techniques such as motion segmentation and optical flow, as well as additional cues such as edges. Further, many biometric acquisition systems require the person and their face/gait to first be detected which can be achieved using a wide variety of techniques including object detection, motion or colour segmentation. It should also be noted that at a lower level still there are large overlaps as many of these techniques also make use of gradient information (for example, the motion segmentation of Denman et al. (2009b), optical flow of Black and Anandan (1993); Zach et al. (2007), feature information of Lakemond et al. (2009), object detection of Dalal and Triggs (2005), and colour segmentation of Rother et al. (2004) all utilise image gradients).

These algorithmic tasks can also inform one another, for example, soft biometrics can be extracted once a person has been tracked for a sufficient time period (Denman et al., 2011), or a detected event can trigger an alarm and initiate object tracking.

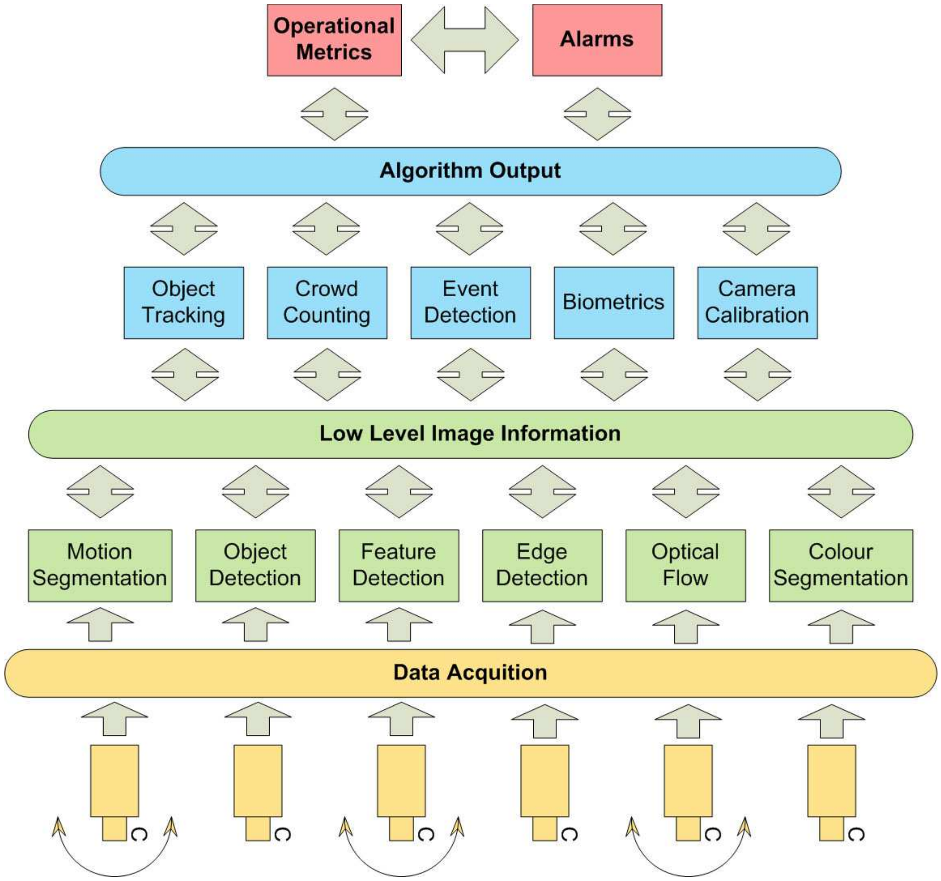


Figure 2: Many video analytics build on a common set of low-level computer vision algorithms. For example object tracking and crowd counting can both make use of motion segmentation, while in both object tracking and biometric extraction object detection is required.

At a higher level, the outputs from these algorithms can be used to generate measures and alarms, once again with significant overlap. For example the output from the crowd counting can provide an operational measure of the number of people in an area, but could also trigger an alarm based on overcrowding. Similarly, object tracking and soft biometrics could be used to track a person of interest through the environment, or monitor the trajectories of all patrons, and record the times taken to move between different areas.

Within this section we will examine two sets of techniques that are principally operational, but can also reinforce the security of an environment:

1. Techniques to count crowds and individuals;
2. Techniques to observe people and their movements.

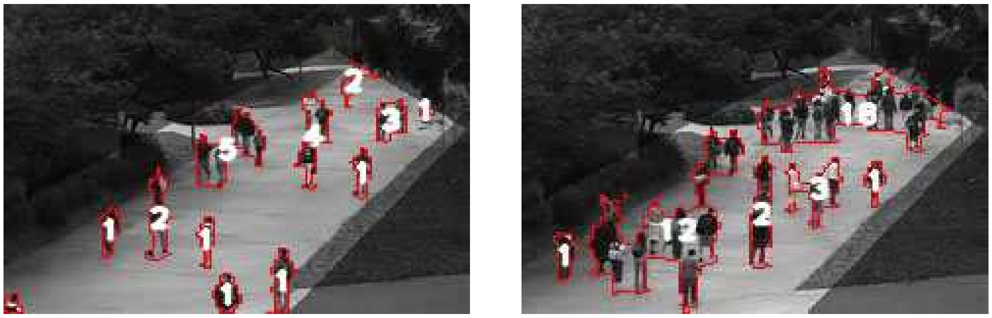
In the following two subsections, we present an overview of both sets of techniques, and demonstrate how they can be applied in a transportation hub environment.

3.2.1. Counting Crowds and Individuals

Counting the number of people in a scene provides an important operational measure. The number of people in a space can inform operational aspects such as staff allocations, and also detect abnormal events such as over or under crowding. Current commercial systems typically require specially placed cameras (see Sensormatic (2012); OTOT (2012)) which while potentially highly accurate, does incur an additional cost in infrastructure. The majority of approaches proposed in the research space have sought to use arbitrarily placed cameras, enabling existing surveillance infrastructure to be utilised.

The majority of crowd counting approaches have used holistic features to estimate the crowd size (Chan et al., 2009; Kong et al., 2006; Davies et al., 1995; Marana et al., 1997; Zhang et al., 2015c,a). Such an approach provides a count for the whole frame, but does not provide any information on the distribution of people within the frame. Furthermore, due to the variations possible in crowd size, distribution and appearance, very large training sets (in the order of hundreds (Kong et al., 2006) or thousands (Chan et al., 2009) of frames) are required; although recently the work of Zhang et al. (2015c) has sought to reduce the training requirements through the use of label distribution learning.

Approaches that use local features (Zhang and Zhang, 2014; Bondi et al., 2014; Ryan et al., 2014, 2010, 2009; Lempitsky and Zisserman, 2010; Kilambi et al., 2008) can overcome many of the limitations of holistic approaches. The approaches of (Zhang and Zhang, 2014; Bondi et al., 2014) use a counting by detection approach, where a learned detector for a a region such as the head and shoulders is used to locate and count all people in the scene. While this works well for uncluttered environments, it does not perform well in dense crowds. Other approaches measure statistics for individual motion regions. Ryan et al. (2010, 2009) extracts a set of features from each motion region such as size, and an edge orientation histogram. Through regression an



(a) Sparse groups in the UCSD data set.

(b) Larger groups in the UCSD data set.

Figure 3: Screenshots of the crowd counting algorithm proposed by Ryan et al. (2012).

estimate of the size of each group is determined. Lempitsky and Zisserman (2010) follow a similar approach, however, a feature vector is extracted for each pixel rather than each motion region. Such approaches can be trained from significantly less training data, with techniques such as Ryan et al. (2010) able to be trained from as few as 10 frames.

We compare results from both local and holistic systems by evaluating them on the UCSD benchmark data set. This data set was introduced by Chan et al. (2008) and contains 2000 annotated frames of pedestrian traffic moving in two directions along a walkway. The video is distributed at a down sampled resolution of 238 × 158 pixels and 10 fps, grayscale. An example frame is shown in Figure 3.

To assess the accuracy of these systems, the testing protocol of Chan et al. (2008) is adhered to. Following this protocol, frames 601-1400 of the UCSD data set are set aside for training, while the remaining 1200 frames are used for testing. In keeping with the training protocol of Lempitsky and Zisserman (2010), a subset of the training data was selected: in MATLAB notation, the ten frames 640:80:1360 were used. Additional subsets, 610:80:1330 and 670:80:1390, were used to give a more representative picture of the performance of local systems using just ten frames.

These results are tabulated in Table 1. The mean absolute error of the holistic systems, Kong et al. (2006) and Chan et al. (2008), lie between 1.92 and 2.47. The local approaches proposed by Lempitsky and Zisserman (2010) and Ryan et al. (2012) outperform the holistic methods significantly, with a mean absolute error ranging from 1.28 to 2.02; supporting the recent

System	Training subset	Error	MSE
Kong, linear	all	1.92	5.60
Kong, neural network	all	2.47 ± 0.41	9.53 ± 3.01
Chan, away+towards	all	1.95	5.75
Chan, all	all	1.95	6.06
Lempitsky	605:5:1400	1.70	-
	640:80:1360	2.02	-
Ryan	610:80:1330	1.72	4.50
	640:80:1360	1.28	2.74
	670:80:1390	1.45	3.39

Table 1: Testing results on the UCSD data set. Frames 601-1400 were set aside for training, and frames 1-600 and 1401-2000 were used for testing. Mean and standard deviation are reported for the neural network based on five runs.

findings of Ryan et al. (2015) which clearly demonstrated the superior performance offered by local approaches. Screen shots from one system (Ryan et al. (2012)) during operation are shown in Figure 3. Blob perimeters are drawn in red and the group size estimates are written on the centroid of each blob, rounded to the nearest integer. In most cases the group estimate is correct to within 1 of the ground truth. An advantage of the local features based approach is that the system can provide a crowding estimate not just for the holistic level, but for the regions occupied by each group within the image. This could be used by a system to detect abnormal crowd distribution patterns or local overcrowding situations, even when the holistic crowd size is within normal ranges.

For a transport hub which potentially contains hundreds of cameras, having to annotate hundreds of frames for each camera is not practical. Local techniques offer a solution to this problem in that they can be implemented in a view invariant manner (Ryan et al., 2014, 2012; Fu et al., 2014; Zhang et al., 2015a). This means that a model can be trained on a set of standard data and be deployed across multiple cameras without any further training, ensuring improved utility within a real world environment.

Scene invariance can be achieved in a variety of ways. Fu et al. (2014) achieves scene invariance through the use of overhead mounted depth cameras, which although successful, requires the installation of specialist hardware. Zhang et al. (2015a) uses deep-learning to learn how to count holisticly in unseen images. While this is potentially a very powerful approach, holistic

Test Set	Mean abs. error	Mean square error
PETS 2009, View 1	1.65	3.91
PETS 2009, View 2	1.23	3.31
PETS 2006, View 3	0.34	0.39
PETS 2006, View 4	0.79	1.15
QUT, Camera A	0.92	1.56
QUT, Camera B	2.06	9.37
QUT, Camera C	1.22	2.42
All tests	1.17 ± 0.57	3.16 ± 3.00

Table 2: Scene invariant testing results on seven data sets with camera calibration. When testing each viewpoint, the system is trained on the six other viewpoints.

approaches are limited by only providing a single count for the entire frame, and ignoring the distribution of people in the scene. The approaches of (Ryan et al., 2014, 2012) use camera calibration to normalise features such that they become scale invariant. This approach has the added advantage that by using camera calibration, the approach is inherently capable of counting across multiple cameras, and we evaluate this approach here.

Seven benchmark data sets with camera calibration have been identified for this purpose: PETS 2009 (Views 1 and 2), PETS 2006 (Views 3 and 4) and the QUT data set introduced by Ryan et al. (2011b). Camera calibration is used to normalise features between viewpoints in order to account for differences in camera position and orientation with respect to the objects in the scene. In each experiment one viewpoint was withheld for testing, and the remaining six viewpoints were used for training. Ten frames from each training viewpoint were selected, so that a total of sixty training frames were used to train the system in each experiment. Testing was then performed on the remaining viewpoint.

Results for these experiments are tabulated in Table 2. Across all experiments, weighted equally, the mean absolute error was 1.17±0.57.

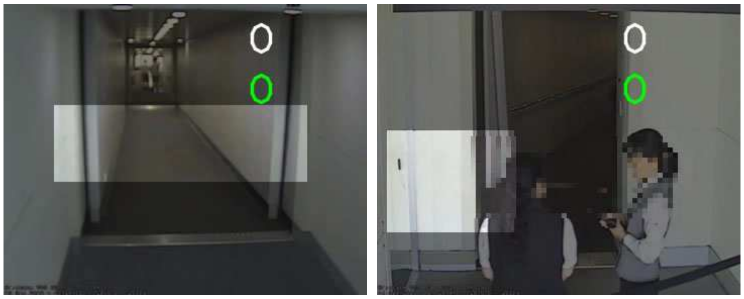
This local feature approach has a number of advantages, most notably that it enables multi-camera crowd counting as outlined in (Ryan et al., 2014), and that it allows for a true 'turn-key' solution. By removing the requirement to train for a specific camera view this approach can be easily deployed across a large camera network without compromising performance, whilst also being able to make use of existing infrastructure.

An alternate counting problem is that of counting the number of people moving past a point, such as the number entering a queue, or entering a shop. This task is quite different from crowd counting, in that it concerns itself with the number of people passing a specific point in space over time, rather than the total number of people in a space at any given instant.

Several approaches have sought to use used overhead cameras (Terada et al., 1999; Kim et al., 2002; Chen, 2003; Chen et al., 2006; Velipasalar et al., 2006; Barandiaran et al., 2008; Albiol et al., 2009), from which people can be easily located and counted through motion segmentation. However, a solution such as this is not appropriate for the vast majority of existing CCTV infrastructure. Kim et al. (2008) proposed the concept of the 'virtual gate' for counting crowds past a point. Kim et al. (2008) uses a single line in the image, and observed optical flow perpendicular to line over time. The observed flow is integrated and scaled by a learned coefficient to obtain a count. Similar approaches have been proposed by Ma and Chan (2013), who introduced a fixed-length sliding temporal window, generating a larger set of samples to train a Bayesian Poisson regression model; and Mukherjee et al. (2014) who used the concept of the influx and outflux count from a region of interest to count people as they passed through a region by tracking pixels on boundary of the ROI. A region based approach proposed by Denman et al. (2015a) detects feature points corresponding to pedestrians and accumulates them as they pass through the virtual gate. During each window of time, the number of people who entered the gate is estimated. A small evaluation of Denman et al. (2015a) is shown below. Two 'virtual gates' are trained to measure the disembarkation rate from an aircraft. Approximately 3 minutes of data is used to train each gate, during which approximately 50 people pass through the region of interest. Figure 4 shows the region of interests for each of the gates.

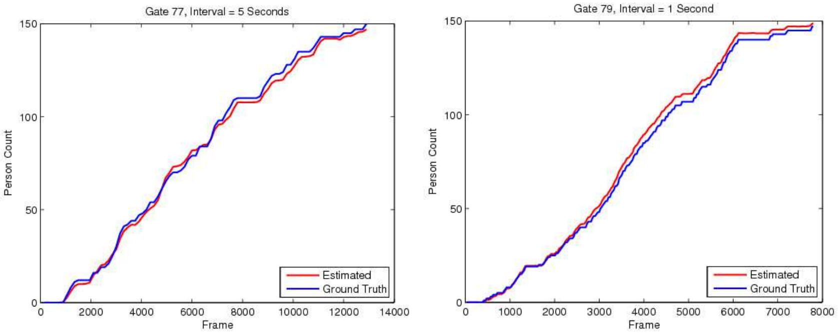
As can be seen from Figure 5, the approach of Denman et al. (2015a) is able to accurately estimate the number of people passing through the gate over time. A sample of the output from the algorithm of Denman et al. (2015a) is shown in Figure 6, and it can be seen that the technique is able to obtain an accurate estimate, despite the crowding and occlusions present.

While techniques such as the 'virtual gate' are highly useful for monitoring entryways in retail spaces, they can also be combined with crowd counting techniques in a complementary way. Such a combination could allow the total people entering and exiting an area, as well as the overall number within the area, to be counted with the results from the two processes combined to achieve a more robust measure. A similar combination could be used to



(a) Gate 77 (b) Gate 79

Figure 4: Configuration of two virtual gates. The non-shaded region is the 'gate' over which people are counted.



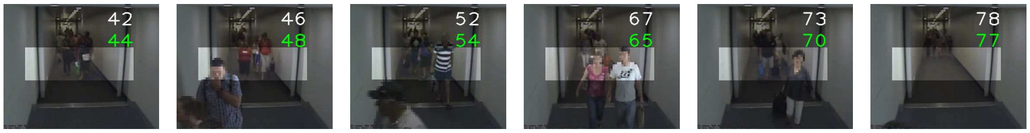
(a) Gate 77 (b) Gate 79

Figure 5: Performance of the Virtual Gate for Two Sequences.

determine queue measures, as shown in Figure 7.

Within a queue, there are three parameters than can be measured: the queue size (Q), the arrival rate (A), and the service rate (S). Measuring any two of these allows the third to be estimated. We demonstrate how a queue can be monitored using footage obtained from an international airport. For this evaluation, Q is measured using the crowd counting algorithm of Ryan et al. (2012), A is estimated using the virtual gate presented in Denman et al. (2015a), and S is estimated from these two measurements. Two video sequences are used in this evaluation:

- **Sequence A**, of length 12 minutes, containing 35 passenger arrivals, 27 passengers serviced and 26-32 people in the queue.
- **Sequence B**, of length 20 minutes, containing 37 passenger arrivals,



(a) t = 0s (b) t = 15s (c) t = 30s (d) t = 45s (e) t = 60s (f) t = 75s

Figure 6: An example of correct operation of the virtual gate, for gate 77 with an interval of 5 seconds. The green number indicates the estimated count, and the white number is the ground truth.

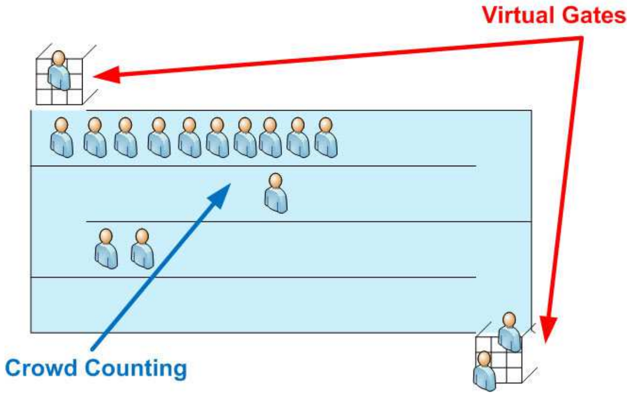


Figure 7: Combining crowd counting and virtual gates to measure queue parameters.

57 passengers serviced and 19-38 people in the queue.

These sequences feature queues located at the check-in counter. They contain both human and non-human objects, making crowd counting and crowd flow monitoring particularly challenging. For each experiment, two tests were run: the system was trained on Sequence A and tested on Sequence B, and then vice versa.

In order to evaluate performance of the crowd counting component, the mean absolute error per frame was used. The performance of the crowd counting algorithm is tabulated in Table 3. The mean absolute error across both sequences was 2.80, and the mean relative error was 9.74%. These values indicate an acceptable level of error for queue length estimates.

The virtual gate algorithm was evaluated by monitoring a sequence of arrivals at the end of a queue. Time windows of length 30 seconds were considered, and the virtual gate module was evaluated using the mean absolute error per 30 second window. Across both sequences, a mean error per

Sequence	**Absolute error**	**Percent error**	**Square error (MSE)**
Sequence A	2.59	8.95%	8.48
Sequence B	3.01	10.53%	11.35
Average	2.80	9.74%	9.91

Table 3: Performance of crowd counting on monitoring queue length.

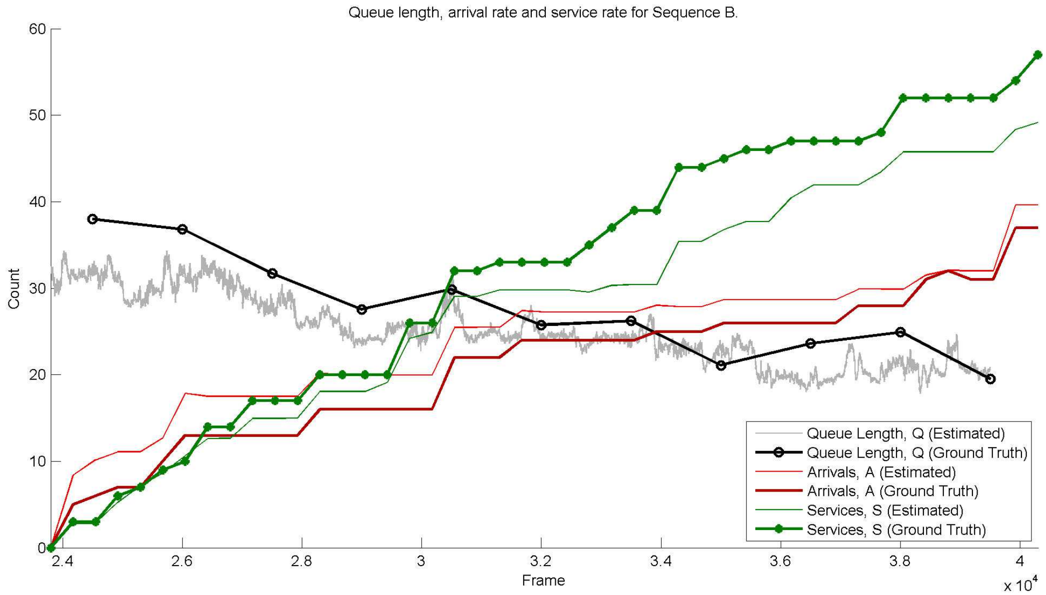


Figure 8: Queue length, arrival rate and service rate for Sequence B. The queue length decreases when the service rate exceeds the arrival rate.

window of 0.394 people was obtained.

Finally, all three parameters (Q,A,S) are plotted in Figure 8 for Sequence B. Estimates are plotted alongside the ground truth for visual comparison. This plot also illustrates the relationship between the three parameters: the size of the queue decreases when service rate exceeds the arrival rate, for example. Although the combined techniques are frequently in error by a few people, overall trends (i.e. growth) are clearly visible and correctly aligned to the ground truth.

As noted in (Denman et al., 2015a), the relationship between multiple virtual gates and crowd sizes could also be exploited outside of a queuing scenario to measure building utilisation, or even track passenger movements across multiple locations. The use of GPR in both the demonstrated crowd counting and virtual gate methods also offers the possibility to model multiple gates and even crowd size jointly by learning the relationship between the different sites (Osborne et al., 2012).

3.2.2. Observing People and their Behaviours

While the techniques outlined in Section 3.2.1 outline how crowds can be observed, there also exists a need to understand how an individual behaves. Questions such as 'How long does it take a person to get from A to B?' and 'What paths do people take through an environment?' can't be answered by simply observing the crowd.

Soft biometrics (Reid et al., 2014; Tome et al., 2014; Dantcheva et al., 2011; Denman et al., 2009a; Jain et al., 2004) are an emerging technology that allow people to be coarsely identified through surveillance footage. Soft biometrics use traits such as height, weight/build, skin and hair colour, and clothing colour to identify people, albeit not uniquely. Using these models, it is possible to repeatedly detect specific individuals, and capture time-based metrics for these individuals as they move through the environment.

In many respects, soft biometrics are similar to the techniques used in person re-identification in multi-camera networks. However many re-identification techniques focus (at least partially) on texture based features (Farenzena et al., 2010; Bak et al., 2010) which are not ideal as soft biometrics as they are less likely to be view invariant. In an unconstrained surveillance situation, it is important that a level of view invariance (tolerance to different view points, i.e. front on, side on) is present, as people can look vastly different from different viewing angles. Also, some modalities can only be acquired from certain angles (i.e. depending on how the subject is dressed, it may be difficult to capture skin colour when observing the subject from behind).

Soft biometrics do not offer the same level of discriminablity found in traditional biometrics (i.e. face (Anantharajah et al., 2014; El Shafey et al., 2013; Thanh et al., 2012), voice (Kanagasundaram et al., 2015; Vogt and Sridharan, 2008), iris (Nguyen et al., 2013; Daugman, 2004)), and gait (Martín-Félez and Xiang, 2014; Sivapalan et al., 2013)) . In a crowded space where there are hundreds or possibly thousands of people present, there are likely to be a large number of people who share a very similar appearance (i.e. at an airport there are likely to be a large number of people in dark suits) making accurate matching across views challenging and error prone. However, recent work (Denman et al., 2011) has shown that it is possible to use soft biometrics to determine what the ‘average’ appearance is, and thus locate people who look unusual and can be more accurately re-detected. This information can then be exploited to provide additional operational information. Figure 9 illustrates this process.

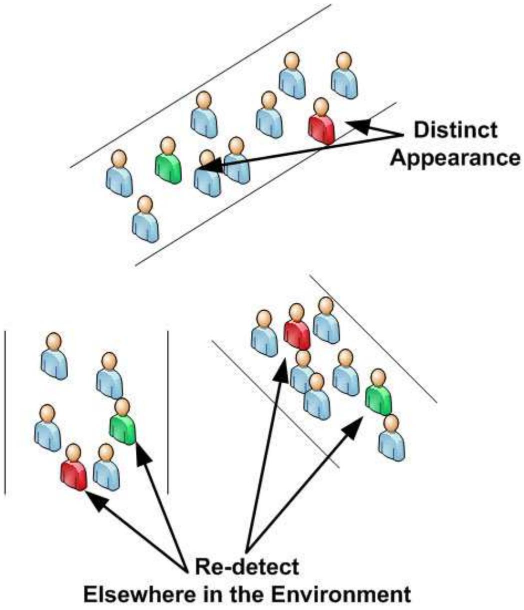


Figure 9: Detecting and re-detecting distinct looking people in an environment.

Through this continuous re-detection of a subset of people, parameters such as the time taken to get between points as well as trajectories through an environment can be continuously estimated. By incorporating additional soft biometrics such as gender and age, it also becomes possible to measure demographics.

We demonstrate how soft biometrics can be used to estimate dwell times on data captured at the security screening point at an international airport. The virtual gate of Denman et al. (2015a) is used to monitor entrances and exits to the queue, and people are detected and extracted as they enter and exit the queues. The soft biometric models of Denman et al. (2011) are built to represent the people that are detected as they pass through the virtual gate, and these models are compared to a global average, allowing people that look distinct (and are thus easier to match) to be identified. These ‘distinct’ people are added to a watch list, and are re-detected as they exit the system, allowing a dwell time for the person to be estimated. It should be noted that it is not possible to extract every person who passes through the virtual gate, as people may be occluded (i.e. two people entering together), which prevents accurate segmentation and localisation of the individuals.

The security screening point used in this evaluation is monitored by four cameras (two at the entrances and two at the exits), and a diagram of the proposed system is shown in Figure 10. This approach is able to estimate the entry and exit rates from the screening area, the total number of people waiting, and the average dwell time for people in the area.

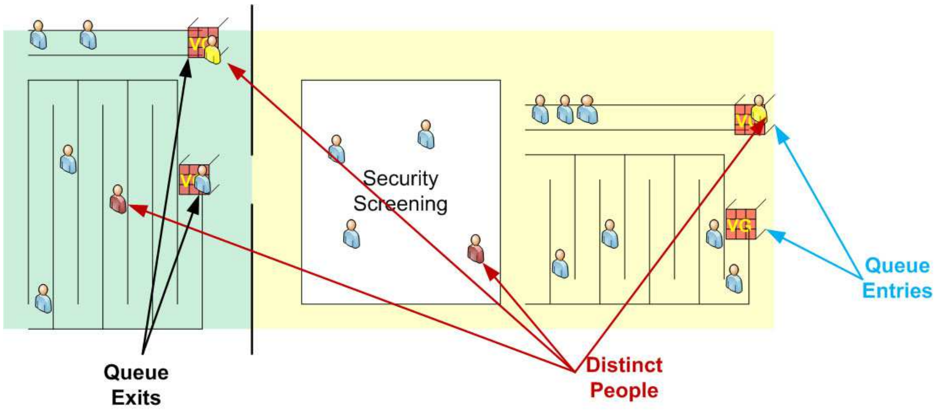
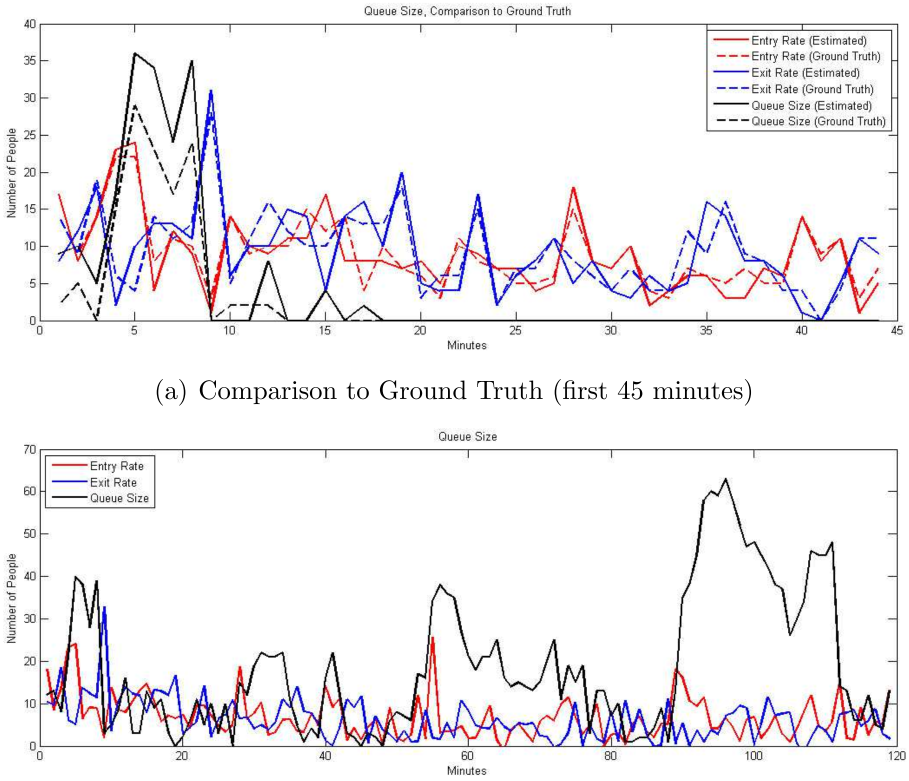


Figure 10: Diagram of the System - Virtual gates are placed on the entrances to the security screening queues (yellow zone), and the entrances to the customs outbound processing queues (green zone). Distinct people are detected as they enter and are matched as they exit to determine overall dwell time.

The system is evaluated on a two hour sequence, and separate footage (approximately 30 minutes per camera) is used to train the virtual gates and average soft biometric models. The entry and exit rates, and overall queue size for the sequence (as well as a comparison to hand annotated ground truth for a portion of the sequence) are shown in Figure 11. The mean absolute error between the estimate and the ground truth (the error is calculated based on the counts in each minute) for the entry rates and exit rates are 1.59 and 2.34 people. The mean absolute error for the queue size, is 1.52. It can be seen that the observed errors are small, and the trends observed in the ground truth are reflected in the system estimates.

Rates at which counted people are extracted, as well as the number of unique subjects, are shown in Table 4.



(a) Comparison to Ground Truth (first 45 minutes)

(b) Whole Sequence (2 Hours)

Figure 11: Evaluation of the Combined System. A comparison to hand annotated ground truth for the first 45 minutes of the sequence is shown in (a), and the output across the entire system is shown in (b).

From the unique people detected, 54 are matched and an average dwell time of 3 minutes and 43 seconds is estimated. Figure 12 shows a plot of the different dwell times detected, as well the likelihoods for each match. A robust average is used to estimate the dwell time, so that those people that are matched with a higher probability are given greater weight. Of the 54 matched people, 14 are matched correctly, 18 are matched incorrectly, and the accuracy of the remaining 22 matches cannot be determined from the video footage due to the low resolution nature of the footage, and changes in pose between the entrances and exits. Despite the matching performance, the estimated dwell time of 3:43 is very close to the ground truth dwell time

Gate	Count	# Detected (%)	# Distinct (%)
Before North	197	110 (55.8%)	57 (52.8%)
Before South	630	435 (69.0%)	99 (22.8%)
After North	173	133 (76.9%)	N/A
After South	580	499 (86.0%)	N/A

Table 4: Number of people counted at each gate, and the rate at which they are detected, and are classified as distinct.

of 3:54, indicating that this approach can obtain accurate estimates for dwell times, despite difficult conditions.

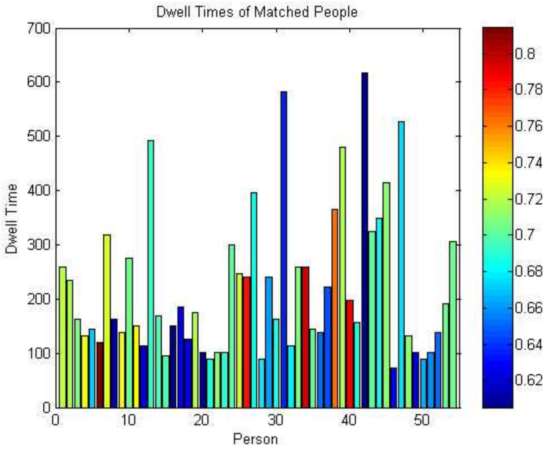


Figure 12: Dwell times for matched people, and the likelihoods of each match. The colour of the bar indicates the likelihood of the match.

It is important to note however that performance could be further improved by using better person re-detection techniques as a basis. The method used in the proposed approach has advantages in that it is computationally simple and lends itself well to locating unique people, however alternate person re-detection methods such as (Bak et al., 2015; Bazzani et al., 2013) are likely to yield further improvements.

Similar approaches to that shown can be used to monitor dwell times across a multi-step process, or record coarse trajectories for a subset of people. Soft biometrics are also a potentially powerful tool for security, as there exists the potential to use them to track a person through a crowded, disjoint camera network, and locate a specific person of interest given a previous sighting.

3.3. Aggregation Layer

Traditionally, sub-systems such as CCTV, building access and fire alarms have been both physically and logically separated. When an alarm is triggered by one sub-system, the onus is on the operators to check other systems, to verify the alarm and/or get additional information. Recent command and control packages are beginning to facilitate the integration of these sub-systems, through features such as enabling relevant CCTV feeds to be displayed when an alarm is triggered. However, this integration is still limited.

There are two possible ways to integrate these signals:

1. Through rules that reconfigure or initialise monitoring activities upon a specific signal or event;
2. Through machine learning techniques that can detect events and anomalies across multiple modalities.

The first of these approaches is, to a large extent, already possible. However, video analytics technologies can offer greater functionality. For example, when access is denied to a security door, rather than simply switching to a camera covering the doorway and allowing the operator to take action, a single object tracking approach (see Section 2.1) could be engaged to track the person, and during tracking biometric information (i.e. face, gait) could be acquired to determine the identity of the person who was denied access.

Machine learning techniques can also be used to integrate multiple signals to detect events across a diverse range of inputs. Probabilistic topic models (Blei et al., 2010, 2003) have been shown to be capable of combining data sources such as images and key-words describing the images to classify unknown images (Wang et al., 2009b), detect multi-agent events (Wang et al., 2009a) (i.e. traffic events), or actions in video Umakanthan et al. (2014). Such techniques could also be used to combine diverse signals such as video, audio and building access to detect events and anomalies. Similarly, Gaussian Process Regression can be used to combine multiple related signals (Osborne et al., 2012) to obtain improved estimates by learning and exploiting the relationship between the signals. Other recent research (Wu et al., 2014) has shown how surveillance outputs can be combined with 'static' forms of data, such as staff or flight schedules to monitor and predict performance of airport processes.

Techniques such as (Wu et al., 2014) can also support partial data, allowing comparisons to be made to historic data, and data from other sites within the broader network (i.e. incorporate data from other transport hubs). Integration such as this can allow abnormalities to be detected through comparisons to previous conditions, and allow for prediction of future behaviour and conditions, enabling improved forward planning and staff scheduling. Analytics could also be developed that operate on historic data, looking for differences or anomalies in performance of the system, or for anomalies across multiple sites.

3.4. Enterprise-Wide Reporting and Monitoring

The integration of systems enables a single point of management for multiple facilities (e.g. the integrated solution employed at Stockholm Arlanda and Stockholm Bromma Airports (Adrem et al., 2007); or the integration required by multiple stakeholders to fully utilise common use and self service technologies Rostworowski (2012)). However, such integration results in an increase in complexity and the amount of data being delivered to operators.

Like the acquisition of data, which can come from a diverse set of sensors, the output of any analysis can be disseminated to an equally diverse set of devices. These include: video walls and command centres, personal computers/laptops, tablet PCs, and mobile phones or PDAs.

The rise of Internet capable tablet PCs and mobile phones provides a new avenue for distributing such data. Rather than simply relying on verbal communication between the command centre and staff on-site, staff in the field can interact with the data directly. Alternate interfaces to those used in a command centre and video wall are required for such devices, however, this is achievable and interfaces can be tailored to the role of the person so that they only see assets and alarms in their area of responsibility.

Mobile devices also offer the advantage of allowing on-site staff to access and enter information directly, rather than relaying information to an operator at the command centre. A staff member could potentially pull up additional information as they need it, enter further details as they are uncovered, or initiate an event. For example, a staff member could report a suspicious person and enter a description (i.e. 1.85m tall, brown hair, fair skin, wearing a red shirt and jeans) through their phone, from which a soft biometric query could be executed to locate the person in the system. Other staff who are nearby the suspect could then be alerted to their location.

4. Case Study: The Future for Automatic Surveillance in Airports

In this section, we assess the modern airport as a case study, and detail how the automatic surveillance framework presented in Section 3 could be perceived to provide a secure, responsive and efficient airport environment. In Section 4.1 we discuss the continued support for traditional airport and aircraft security, and highlight the emergence of capabilities which can support timely and effective incident response management. In Section 4.2 we look at new capabilities for monitoring airport operations which will impact on both airport and enterprise reporting of Key Performance Indicators (KPI) into the future.

With a move away from security-only applications of video analytics, the concepts and applications discussed in these sections should provide insight to the design and installation of CCTV camera networks in airports into the future so that they are able to fully harness the potential that has been described in this paper.

4.1. Security and Incident Response Management

While the potential for organisation-wide performance monitoring is an important factor in the innovative use of CCTV systems, added capability enhancement is not limited to this area. Incident response capability and more broadly safety management within airport contexts can also be enhanced by the integration of automatic surveillance frameworks with other systems whose functionality might be seen as unrelated to conventional usage of CCTV (video-analytic) usage. For example if this capability is combined with existing security system components such as Identity Management (access control), emergency evacuation systems and spatial building plans, significant expansion of incident response and management capabilities are possible. An important element of such an emergent capability in the visualisation of human activity in relation to 'as-built' space is enhanced situational awareness of the extent of a loss of function or damage (Boehm et al., 2005), or in extreme contexts, a security crisis.

The scale of consequences from incidents in such complex settings is often difficult to anticipate. Furthermore, because of the potential for the rapid flow of impacts throughout the many public and private spaces of an airport, management may be unlikely to face single incidents but rather a series of incidents within and across functional areas, often appearing concurrently. The potential for this type of cascading effect may lead to decisions to evacuate parts or all of an airport. If both land and airside spaces in an airport are involved, the consequences of escalation of a response to such a level will be significant. The ability to create a 'common operating picture' of security incidents at a team level is an important factor in response and recovery. Such ability is aided greatly by sophisticated use of CCTV and video analytics.

An outcome of this integration of video analytics with emergency recognition and response mechanisms is a nascent ability to rapidly assess the nature of an incident and to enhance evacuation where needed. This is an important issue in instances where conventional alarm systems such as for smoke or fire are activated as these systems alone cannot differentiate between emergency incidents caused by accident or from human intent (Boswell and Gwynne, 2007).

While they may be considered by some to be cost prohibitive, higher-end video analytic systems, if used in a dual function context of business-as-usual and business-not-as-usual can transform the agility of airport operations in respect to both security and incident response.

Benefits would include:

- Enhancement of early threat recognition via rapid situational awareness;

- Protection of critical infrastructure(s) if used in combination with comprehensive building design information detailing the position and nature of at-risk building components;

- Support of protocols for the rapid escalation of response by capability and capacity.

A Managed Response capability aided by the integration of in-situ CCTV camera networks and video analytics, as described here, allows the combination of functions such as the estimation of crowd density and numbers within specific airport locations as well as determination of best routes for evacuation and location and directions of movement of self-evacuees within the complex spaces of a modern airport. The use of 'virtual gate' crowd counting capabilities enabled by the video analytic functionality is an obvious aid in safe evacuation. Additionally, when re-occupancy of airport is delayed due to post-incident forensic assessments, use of video data can assist in damage estimation and planning for remediation and repairs.

4.2. Real-Time Monitoring of Terminal Status and KPI Measurement

In addition to the ability to monitor the security and safety levels in the airport as described in the previous section, the advance of surveillance systems has the ability to allow for more sophisticated operational monitoring.

In terms of operations monitoring, airport operators typically have very good data for airside operations (in particular aircraft movements), but have limited – often aggregated – data at the level of passenger flows on an hourly basis (de Neufville and Odoni, 2003). With the advancements in video analytic technology and the supporting framework, airport operators (and other airport stakeholders such as government agencies, airlines, retail owners, ground handlers etc) will be able to collect data with greater fidelity in order to assist with operations management. Such data can also be used as a gauge in the planning and design phase, complementing (and potentially improving) peak-hour analysis.

In particular, such video analytics will provide the capability to perform real-time monitoring of a wide range of indicators which are linked to internal and external stakeholder key performance indicators, including widely established Level Of Service (LOS) metrics. Typical LOS metrics which are used in airports are related to space and time. Examples include the space provided for passengers in different facilities (e.g. check-in or baggage reclaim), and in passageways. These metrics often have associated LOS standards which are defined by industry-wide bodies such as the International Air Transportation Association (IATA), or by individual airports. These standards are based on the premise that the greater the amount of space passengers have, the higher the perceived level of service (de Neufville and Odoni, 2003).

Other commonly used levels of service are based on time (de Neufville and Odoni, 2003; Correia and Wirasinghe, 2004). Examples include waiting times in queues, service times for an individual process (i.e. check-in), or aggregated service times such as the total time to be processed through all landside areas (i.e. check-in, security and immigration). These metrics are also linked with the levels of service standards for space, since passengers only occupy space for a finite length of time (de Neufville and Odoni, 2003).

For both space and time levels of service, there can easily be a discrepancy between the LOS used during the planning and design phase, and the actual LOS during operations. Through the emerging capabilities of the Intelligence Layer (refer to Fig. 1), there is now an ability to understand the real-time relationship between designed LOS and actual LOS. In particular, space-defined LOS can be obtained through crowd counting analytics (Section 3.2.1) and knowledge of the physical area of the space. Alternatively, waiting times in queues can be calculated using queue-based metrics such as measurement of arrivals and departures to queues through "virtual gates". These types of analytics provide the ability to determine average wait times, but variations in times could also be captured with the addition of other analytics techniques such as soft biometrics.

Crowd counting and space utilisation analytics can also be used to identify bottlenecks and cross-flows (i.e. restrictions on passenger flows) which arise during operation, and were not perceived in the design of the space. The analytics have the ability to highlight where difficulties may be arising, and the data warehousing capability provides a means by which the root cause of the problem can be identified.

There are other less-formal levels of service which can be captured through the use of video analytics and the proposed framework. For example, from the review of LOS by Correia and Wirasinghe (2004) and the discussion presented by Fodness and Murray (2007) around perceptions of airport service quality, other LOS metrics which are not formally measured may now be easily captured. Metrics such as the availability of seating and baggage carts can be obtained through object detection and tracking, whilst advanced behaviour monitoring could be used to provide indications of passenger orientation, in particular identifying passengers who seem 'lost'. Soft biometrics could also be used to capture passenger walking distances.

Having this real-time capability for monitoring the performance of the facility can be used to provide accurate information back to the passenger (through the visualisation module of the proposed framework) in relation to travel times and expected delays. Making this type of information available to passengers may help to lessen the stress of navigating the airport in a timely manner. Some airports already provide this type of information, either through static walking times, or through delay indicators.

The advancement of video analytics has advantages for other airport stakeholders aside from the operator themselves. For instance, government agencies (who run immigration controls) will have the ability to monitor their service rates in order to ensure they are meeting their required KPIs (as is the case in Australia). Airlines will have the ability to monitor their service rates at check-in, and to monitor passenger boarding and deboarding which are key elements to ensuring fast turn-around times of aircraft (Fricke and Schultz, 2009).

The enterprise-wide reporting capabilities that have been included in the proposed surveillance framework is important for benchmarking operations across multiple facilities. Whilst this includes airport operators who manage multiple airports (e.g. Queensland Airports Limited who operate Gold Coast, Townsville and Mt Isa airports, or NT Airports who operate Darwin, Alice Springs and Tennant Creek airports, both in Australia), this capability is envisaged to be of greater benefit to stakeholders such as security contractors and government agencies.

The proposed surveillance framework also provides the ability to report KPIs to a common framework which is measuring airport performance. Given the number of stakeholders present in the airport, it is often difficult to assign accountability of the passenger movement to any one stakeholder. Having the ability to monitor individual queues (e.g. check-in, security, immigration), spaces (e.g. baggage reclaim) and processes (e.g. aircraft boarding/deboarding) enables the development of a true performance framework in which each stakeholder has full responsibility over their own operations, and can identify the effects of their own performance on other areas of the airport, and the overall airport level of service.

The data management aspects of the surveillance framework can also provide a historical snapshot of the airport's operations, which could provide more detailed information to validate the data used to guide airport design. Likewise, the video analytics capabilities can provide more advanced data for airport passenger simulation models which require significantly more data than is required for airside models (de Neufville and Odoni, 2003).

In summary, the introduction of operational analytics has the potential to completely change, and ultimately improve, the way airport operations are managed and reported. The proposed framework is able to support this through an Intelligence Layer which includes new algorithms for tracking passengers, counting crowds in spaces, and monitoring demand and service of queues. The data warehousing aspects enable the collection of additional data to improve future airport planning and design, and also enables stakeholders who operate in multiple airports to benchmark their performance across all facilities.

5. Conclusion

In this paper, we have explored how intelligent surveillance is becoming increasingly focussed on operational analytics, and the implications of this for transport hubs and other large infrastructure. We have outlined two particular areas of operational analytics: crowd counting (including counting everyone in a scene, measuring pedestrian throughput and estimating queue size) and dwell time estimation; and have shown how valuable operational information can be obtained using these techniques. We have also discussed the wider ramifications of these analytics, including how multiple analytics can be combined, and how they can be combined with static information, to provide a richer understanding of the infrastructure's performance.

Despite the promise shown by these techniques and by the approaches outlined in this paper, there are a number of problems that are yet to be overcome. Whilst some techniques presented here such as the scene invariant crowd-counting are 'turn-key' type approaches in that they can be deployed with minimal set-up, other techniques such as the virtual gate require training for each instance. While the annotation and training requirements are not particularly onerous (training data can generally be annotated much faster than real-time, and as little as 30 minutes is sufficient to train a model), this requirement becomes more demanding as the number of cameras scales up. The integration of such diverse sets of data streams across very large sites is still an open problem, and although progress has been made, whether through Bayesian approaches such as (Wu et al., 2014), or large scale data collection such as (Denman et al., 2015a), there is as yet still no complete system to integrate multiple operational analytics with other data sources.

To overcome these limitations and further develop operational analytics, a number of possible avenues exist. The joint modelling of multiple data streams (i.e. counting multiple 'virtual gates' simultaneously) has the potential to improve accuracy by incorporating observations where there is mutual information (Osborne et al., 2012). Similarly, the integration with other sources of data, such as passenger manifests or data from check-in or border security terminals, may provide a way to further improve accuracy by providing a secondary estimation of crowd size. Reducing the need for data annotation and training for all camera views is also highly desirable. While this can be achieved through camera calibration for techniques such as crowd counting, the solution for other problems such as throughput estimation (i.e. the 'virtual gate') is less clear, although using other techniques (such as crowd counting) to effectively boot-strap the system by automatically generating a training set may yield one solution. The recent success achieved by applying deep learning to crowd counting (Zhang et al., 2015a) offers another direction that is also worth exploring.

Finally, there also exist a number of other security analytics that are yet to be exploited for operational tasks. For instance, event recognition is primarily focused on detecting specific or abnormal events in a security context, but is equally applicable to monitoring processes in an operational scenario. By incorporating event detection within a state machine (such as Bayesian network), processes such as security inspections could be monitored and detailed information on time taken could be gathered. Similarly, abnormal events within the process (or even steps being performed out of order) could be detected and personnel could be notified.

Acknowledgments

This research was supported by the Australian Research Council's Linkage Project "Airports of the Future" (LP0990135). The authors also acknowledge the contribution made by the many aviation industry stakeholders also involved in this project.

Adrem, A., Dell'orto, S., Lennerman, A., 2007. Moving away from a traditional airport security set-up to a new integrated security model. Journal of Airport Management 1 (3), 262–267.

Ajiboye, S. O., Birch, P., Chatwin, C., Young, R., 2015. Hierarchical video surveillance architecture: a chassis for video big data analytics and exploration. In: IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics.

Albiol, A., Albiol, A., Silla, J., 2009. Statistical video analysis for crowds counting. In: International Conference on Image Processing. pp. 2569–2572.

Anantharajah, K., Ge, Z., McCool, C., Denman, S., Fookes, C., Corke, P., Tjondronegoro, D., Sridharan, S., 2014. Local inter-session variability modelling for object classification. In: Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on. IEEE, pp. 309–316.

Arroyo, R., Yebes, J. J., Bergasa, L. M., Daza, I. G., Almazán, J., 2015. Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls. Expert Systems with Applications.

Ashford, N., Coutu, P., Beasley, J., 2013. Airport operations, 3rd Edition. McGraw- Hill Ryerson Ltd., 300 Water Street, Whitby, ON L1N 9B6 Canada.

Auvinet, E., Grossmann, E., Rougier, C., Dahmane, M., Meunier, J., 2006. Left-luggage detection using homographies and simple heuristics. In: IEEE International Workshop on PETS, New York, June 18, 2006. New York, pp. 51–58.

Bak, S., Corvee, E., Brémond, F., Thonnat, M., 2010. Person re-identification using haar-based and DCD-based signature. In: Advanced Video and Signal Based Surveillance (AVSS). p. 1–8.

Bak, S., Martins, F., Bremond, F., 2015. Person re-identification by pose priors. In: IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics.

Bandini, S., Federici, M. L., Manzoni, S., 2007. A qualitative evaluation of technologies and techniques for data collection on pedestrians and crowded situations. In: Proceedings of the 2007 Summer Computation Simulation Conference. Society for Computer Simulation International, San Diego, CA, USA, pp. 1057–1064.

Barandiaran, J., Murguia, B., Boto, F., may 2008. Real-time people counting using multiple lines. pp. 159 –162.

Bazzani, L., Cristani, M., Murino, V., 2013. Symmetry-driven accumulation of local features for human characterization and re-identification. Computer Vision and Image Understanding 117 (2), 130–144.

Bazzani, L., Zanotto, M., Cristani, M., Murino, V., 2015. Joint individual-group modeling for tracking. Pattern Analysis and Machine Intelligence, IEEE Transactions on 37 (4), 746–759.

Bedagkar-Gala, A., Shah, S. K., 2014. Gait-assisted person re-identification in wide area surveillance. In: Computer Vision-ACCV 2014 Workshops. Springer, pp. 633–649.

Bellotto, N., Sommerlade, E., Benfold, B., Bibby, C., Reid, I., Roth, D., Fernandez, C., Gool, L. V., Gonzalez, J., 2009. A distributed camera system for multi-resolution surveillance. In: International Conference on Distributed Smart Cameras. pp. 1–8.

Ben Shitrit, H., Berclaz, J., Fleuret, F., Fua, P., 2014. Multi-commodity network flow for tracking multiple people. Pattern Analysis and Machine Intelligence, IEEE Transactions on 36 (8), 1614–1627.

Berclaz, J., Fleuret, F., Turetken, E., Fua, P., Sep. 2011. Multiple object tracking using k-shortest paths optimization. IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (9), 1806–1819.

Black, M., Anandan, P., 1993. A framework for the robust estimation of optical flow. In: Fourth International Conference on Computer Vision. pp. 231 – 236.

Blei, D., Carin, L., Dunson, D., nov. 2010. Probabilistic topic models. IEEE Signal Processing Magazine 27 (6), 55 –65.

Blei, D. M., Ng, A. Y., Jordan, M. I., March 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.

Boehm, K., Roth, V., Kelley, J., 2005. Enhancing situation awareness in real time geospatial visualization. In: Americas Conference on Information Systems (AMCIS 2005). p. Paper 234.

Bondi, E., Seidenari, L., Bagdanov, A. D., Del Bimbo, A., 2014. Real-time people counting from depth imagery of crowded environments. In: Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on. IEEE, pp. 337–342.

Boswell, D., Gwynne, S., July/August 2007. Air, fire and ice: Fire & security challenges unique to airports. Fire & Security Today, 30–37.

Brulin, M., Nicolas, H., Maillet, C., 2010. Video surveillance traffic analysis using scene geometry. In: Pacific Rim Symposium on Image and Video Technologies. Singapore.

Cancela, B., Hospedales, T. M., Gong, S., 2014. Open-world person re-identification by multi-label assignment inference. British Machine Vision Association, BMVA.

Chan, A., Liang, Z.-S., Vasconcelos, N., June 2008. Privacy preserving crowd monitoring: Counting people without people models or tracking. IEEE Conference on Computer Vision and Pattern Recognition, 1–7.

Chan, A. B., Morrow, M., Vasconcelos, N., 2009. Analysis of crowded scenes using holistic properties. In: Performance Evaluation of Tracking and Surveillance Workshop. Miami, Florida, pp. 101–108.

Chang, R.-I., Wang, T.-C., Wang, C.-H., Liu, J.-C., Ho, J.-M., 2012. Effective distributed service architecture for ubiquitous video surveillance. Information Systems Frontiers 14 (3), 499–515.

Chen, T.-H., oct. 2003. An automatic bi-directional passing-people counting method based on color image processing. In: International Carnahan Conference on Security Technology. pp. 200 – 207.

Chen, T.-H., Chen, T.-Y., Chen, Z.-X., June 2006. An intelligent people-flow counting method for passing through a gate. In: IEEE Conference on Robotics, Automation and Mechatronics. pp. 1–6.

Correia, A., Wirasinghe, S., 2004. Evaluating level of service at airport passenger terminals. Transportation Research Record: Journal of the Transportation Research Board (1888), 1–6.

Dalal, N., Triggs, B., June 2005. Histograms of oriented gradients for human detection. In: Schmid, C., Soatto, S., Tomasi, C. (Eds.), International Conference on Computer Vision & Pattern Recognition. Vol. 2. INRIA Rhône-Alpes, ZIRST-655, av. de l'Europe, Montbonnot-38334, pp. 886–893.

Danelljan, M., Khan, F. S., Felsberg, M., van de Weijer, J., 2014. Adaptive color attributes for real-time visual tracking. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, pp. 1090–1097.

D'Angelo, A., Dugelay, J.-L., 30 2010-june 2 2010. Color based soft biometry for hooligans detection. In: IEEE International Symposium on Circuits and Systems (ISCAS). pp. 1691 –1694.

Dantcheva, A., Velardo, C., DAngelo, A., Dugelay, J.-L., 2011. Bag of soft biometrics for person identification: New trends and challenges. Multimedia Tools and Applications 51 (2), 739–777.

Daugman, J., 2004. How iris recognition works. IEEE Transactions on Circuits and Systems for Video Technology 14, 21–30.

Davies, A., Yin, J. H., Velastin, S., Feb 1995. Crowd monitoring using image processing. Electronics & Communication Engineering Journal 7 (1), 37–47.

de Neufville, R., Odoni, A., 2003. Airport systems planning design and management. McGraw-Hill, Inc.

Deng, L., Yu, D., 2014. Deep learning: methods and applications. Foundations and Trends in Signal Processing 7 (3–4), 197–387.

Denman, S., Bialkowski, A., Fookes, C., Sridharan, S., Aug. 2011. Determining Operational Measures from Multi-Camera Surveillance Systems using Soft Biometrics. In: 8th IEEE International Conference on Advanced Video and Signal-Based Surveillancei (AVSS). p. 6.

Denman, S., Chandran, V., Sridharan, S., 2006a. A multi-class tracker using a scalable condensation filter. In: Advanced Video and Signal Based Surveillance. Sydney.

Denman, S., Fookes, C., Bialkowski, A., Sridharan, S., 2009a. Soft-Biometrics: unconstrained authentication in a surveillance environment. Digital Image Computing: Techniques and Applications, 196–203.

Denman, S., Fookes, C., Cook, J., Davoren, C., Mamic, A., Farquharson, G., Chen, D., Chen, B., Sridharan, S., 2006b. Multi-view intelligent vehicle surveillance system. In: Advanced Video and Signal Based Surveillance. Sydney.

Denman, S., Fookes, C., Ryan, D., Sridharan, S., 2015a. Large scale monitoring of crowds and building utilisation: A new database and distributed approach. In: 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE.

Denman, S., Fookes, C., Sridharan, S., 2009b. Improved simultaneous computation of motion detection and optical flow for object tracking. In: Digital Image Computing: Techniques and Applications. Melbourne, Australia.

Denman, S., Fookes, C., Sridharan, S., 2010. Group segmentation during object tracking using optical flow discontinuities. In: The 4th Pacific-Rim Symposium on Image and Video Technology. Singapore.

Denman, S., Halstead, M., Bialkowski, A., Fookes, C., Sridharan, S., December 2012. Can you describe him for me? a technique for semantic person search in video. In: Digital Image Computing: Techniques and Applications.

Denman, S., Halstead, M., Fookes, C., Sridharan, S., 2015b. Searching for people using semantic soft biometric descriptions. Pattern Recognition Letters.

Denman, S., Sridharan, S., Chandran, V., 2007. Abandoned object detection using multi-layer motion detection. In: International Conference on Signal Processing and Communication Systems (ICSPCS). Vol. 1. Gold Coast, QLD, pp. 439–448.

El Shafey, L., McCool, C., Wallace, R., Marcel, S., 2013. A scalable formulation of probabilistic linear discriminant analysis: Applied to face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35 (7), 1788–1794.

Erhan, D., Szegedy, C., Toshev, A., Anguelov, D., 2014. Scalable object detection using deep neural networks. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, pp. 2155–2162.

Eshel, R., Moses, Y., 2008. Homography based multiple camera detection and tracking of people in a dense crowd. In: Computer Vision and Pattern Recognition. pp. 1–8.

Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M., 2010. Person re-identification by symmetry-driven accumulation of local features. In: Computer Vision and Pattern Recognition (CVPR). pp. 2360–2367.

Felsberg, M., 2013. Enhanced distribution field tracking using channel representations. In: International Conference on Computer Vision Workshops. IEEE, pp. 121–128.

Fodness, D., Murray, B., 2007. Passenger's expectation of airport service quality. Journal of Services Marketing 21 (7), 492–506.

Fookes, C., Denman, S., Lakemond, R., Ryan, R., Sridharan, S., Piccardi, M., 2010. Semi-supervised intelligent surveillance system for secure environments. In: IEEE International Symposium on Industrial Electronics. pp. 2815–2820.

Fookes, C., Lin, F., Chandran, V., Sridharan, S., 2012. Evaluation of image resolution and super-resolution on face recognition performance. Journal of Visual Communication and Image Representation 23 (1), 75 – 93.

Fricke, H., Schultz, M., 2009. Delay impacts onto turnaround performance: Optimal time buffering for minimizing delay propagation. In: Proceedings of the USA/Europe Air Traffic Management Research and Development Seminar. pp. 1–10.

Fu, H., Ma, H., Xiao, H., 2014. Scene-adaptive accurate and fast vertical crowd counting via joint using depth and color information. Multimedia Tools and Applications 73 (1), 273–289.

Galoogahi, H. K., 2010. Tracking groups of people in presence of occlusion. In: Pacific Rim Symposium on Image and Video Technologies. Singapore.

Gray, D., Brennan, S., Tao, H., 2007. Evaluating appearance models for recognition, acquisition and tracking. In: IEEE International Workshop on PETS.

Halstead, M., Denman, S., Sridharan, S., Fookes, C., 2014. Locating people in video from semantic descriptions: A new database and approach. In: 22nd International Conference on Pattern Recognition (ICPR). IEEE, pp. 4501–4506.

Hare, S., Saffari, A., Torr, P. H., 2011. Struck: Structured output tracking with kernels. In: Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, pp. 263–270.

Haritaoglu, I., Harwood, D., Davis, L., 2000. W4: real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8), 809 – 830.

Haritaoglu, I., Harwood, D., Davis, L. S., 1999. Hydra: Multiple people detection and tracking using silhouettes. In: Second IEEE Workshop on Visual Surveillance. IEEE Computer Society, Washington, DC, USA, p. 6.

Henriques, J. F., Caseiro, R., Martins, P., Batista, J., 2015. High-speed tracking with kernelized correlation filters. Pattern Analysis and Machine Intelligence, IEEE Transactions on 37 (3), 583–596.

Herrero, E., Orrite, C., Senar, J., 2003. Detected motion classification with a double-background and a neighborhood-based difference. Pattern Recognition Letters 24, 2079–2092.

Hospedales, T., Li, J., Gong, S., Xiang, T., dec. 2011. Identifying rare and subtle behaviors: A weakly supervised joint topic model. IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (12), 2451 –2464.

Huang, C., Wu, B., Nevatia, R., 2008. Robust object tracking by hierarchical association of detection responses. In: Proceedings of the 10th European Conference on Computer Vision: Part II. ECCV '08. Springer-Verlag, Berlin, Heidelberg, pp. 788–801.

Huang, L., Barth, M., jun. 2010. Real-time multi-vehicle tracking based on feature detection and color probability model. In: IEEE Intelligent Vehicles Symposium (IV). pp. 981 –986.

Jain, A. K., Dass, S. C., Nandakumar, K., Jul. 2004. Soft biometric traits for personal recognition systems. Lecture notes in computer science, 731–738.

Kanagasundaram, A., Dean, D., Sridharan, S., 2015. Improving out-domain plda speaker verification using unsupervised inter-dataset variability compensation approach. In: Proceedings of 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015). IEEE, pp. 4654–4658.

Kenk, V. S., Kovačič, S., Kristan, M., Hajdinjak, M., Perš, J., et al., 2015. Visual re-identification across large, distributed camera networks. Image and Vision Computing 34, 11–26.

Kilambi, P., Ribnick, E., Joshi, A. J., Masoud, O., Papanikolopoulos, N., 2008. Estimating pedestrian counts in groups. Computer Vision and Image Understanding 110 (1), 43 – 59.

Kim, B.-S., Lee, G.-G., Yoon, J.-Y., Kim, J.-J., Kim, W.-Y., 2008. A method of counting pedestrians in crowded scenes. In: ICIC 08. Springer-Verlag, Berlin, Heidelberg, pp. 1117–1126.

Kim, J.-W., Choi, K.-S., Choi, B.-D., Ko, S.-J., 2002. Real-time vision-based people counting system for the security door. In: International Technical Conference on Circuits/Systems Computers and Communications. pp. 1416–1419.

Kong, D., Gray, D., Tao, H., 0-0 2006. A viewpoint invariant approach for crowd counting. In: International Conference on Pattern Recognition. Vol. 3. pp. 1187 –1190.

Kooij, J. F., Englebienne, G., Gavrila, D. M., 2015. Identifying multiple objects from their appearance in inaccurate detections. Computer Vision and Image Understanding 136, 103–116.

Krahnstoever, N., Tu, P., Sebastian, T., Perera, A., Collins, R., 2006. Multiview detection and tracking of travelers and luggage in mass transit environments. In: IEEE International Workshop on PETS, New York, June 18, 2006. New York, pp. 67–74.

Lakemond, R., Fookes, C., Sridharan, S., 2009. Afne adaptation of local image features using the hessian matrix. In: IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 496–501.

Leibe, B., Schindler, K., Cornelis, N., Van Gool, L., Oct. 2008. Coupled object detection and tracking from static cameras and moving vehicles. IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (10), 1683–1698.

Lempitsky, V., Zisserman, A., 2010. Learning to count objects in images. In: Advances in Neural Information Processing Systems.

Li, Y., Wu, Z., Karanam, S., Radke, R. J., 2014. Real-world re-identification in an airport camera network. In: Proceedings of the International Conference on Distributed Smart Cameras. ACM, p. 35.

Lin, F., Fookes, C., Chandran, V., Sridharan, S., 2005. Investigation into optical flow super-resolution for surveillance applications. In: APRS Workshop on Digital Image Computing. p. 7378.

Lin, F., Fookes, C., Chandran, V., Sridharan, S., 2007. Super-resolved faces for improved face recognition from surveillance video. In: Lecture Notes in Computer Science. p. 1 10.

Liu, B., Yang, L., Huang, J., Meer, P., Gong, L., Kulikowski, C., 2010. Robust and fast collaborative tracking with two stage sparse optimization. In: Computer Vision–ECCV 2010. Springer, pp. 624–637.

López-Méndez, A., Monay, F., Odobez, J.-M., 2014. Exploiting scene cues for dropped object detection. In: 9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. No. EPFL-CONF-192504.

Lu, W., Tan, Y.-P., 2001. A color histogram based people tracking system. In: 2001 IEEE International Symposium on Circuits and Systems. Vol. 2. pp. 137 – 140.

Luo, P., Tian, Y., Wang, X., Tang, X., 2014. Switchable deep network for pedestrian detection. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, pp. 899–906.

Ma, Z., Chan, A. B., 2013. Crossing the line: Crowd counting by integer programming with local features. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, pp. 2539–2546.

Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N., jun. 2010. Anomaly detection in crowded scenes. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1975 –1981.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., McClosky, D., 2014. The stanford corenlp natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. pp. 55–60.

Marana, A., Velastin, S., Costa, L., Lotufo, R., Mar 1997. Estimation of crowd density using image processing. Image Processing for Security Applications (Digest No.: 1997/074), IEE Colloquium on, 11/1–11/8.

Martín-Félez, R., Xiang, T., 2014. Uncooperative gait recognition by learning to rank. Pattern Recognition 47 (12), 3793–3806.

Medeiros, C., Costa, J., Fernandes, C., 2011. Passive uhf rfid tag for airport suitcase tracking and identification. IEEE Antennas and Wireless Propagation Letters 10, 123 –126.

Mehran, R., Oyama, A., Shah, M., jun. 2009. Abnormal crowd behavior detection using social force model. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 935 –942.

Mukherjee, S., Gil, S., Ray, N., 2014. Unique people count from monocular videos. The Visual Computer, 1–13.

Nallaivarothayan, H., Fookes, C., Denman, S., Sridharan, S., 2014. An mrf based abnormal event detection approach using motion and appearance features. In: 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, pp. 343–348.

Nam, Y., 2015. Real-time abandoned and stolen object detection based on spatio-temporal features in crowded scenes. Multimedia Tools and Applications, 1–26.

Nazare, A. C., dos Santos, C. E., Ferreira, R., Robson Schwartz, W., 2014. Smart surveillance framework: A versatile tool for video analysis. In: Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on. IEEE, pp. 753–760.

Nguyen, K., Fookes, C., Sridharan, S., Denman, S., 2013. Feature-domain super-resolution for iris recognition. Computer Vision and Image Understanding 117 (10), 1526–1535.

Okuma, K., Taleghani, A., Freitas, N. d., Little, J., Lowe, D., 2004. A boosted particle filter: Multitarget detection and tracking. In: 8th European Conference on Computer Vision (ECCV). Vol. 1. Prague, Czech Republic, pp. 28–39.

Osborne, M. A., Roberts, S. J., Rogers, A., Jennings, N. R., 2012. Real-time information processing of environmental sensor network data using bayesian gaussian processes. ACM Transactions on Sensor Networks 9 (1), 1.

OTOT, 2012. Overhead Thermal Sensor For People Counting. Http://www.otot.ws/en/products/overhead-thermal-sensor-for-people-counting.

Ottlik, A., Nagel, H.-H., 2008. Initialization of model-based vehicle tracking in video sequences of inner-city intersections. International Journal of Computer Vision 80, 211–225.

Park, U., Jain, A., Kitahara, I., Kogure, K., Hagita, N., 0-0 2006. Vise: Visual search engine using multiple networked cameras. In: International Conference on Pattern Recognition. Vol. 3. pp. 1204 –1207.

Patil, P., Kokil, A., 2015. Wifipi-tracking at mass events. In: Pervasive Computing (ICPC), 2015 International Conference on. IEEE, pp. 1–4.

Pirsiavash, H., Ramanan, D., Fowlkes, C. C., 2011. Globally-optimal greedy algorithms for tracking a variable number of objects. In: IEEE Conference on Computer Vision and Pattern Recognition. CVPR '11. IEEE Computer Society, Washington, DC, USA, pp. 1201–1208.

Reid, D., Nixon, M. S., Stevenage, S. V., et al., 2014. Soft biometrics; human identification using comparative descriptions. Pattern Analysis and Machine Intelligence, IEEE Transactions on 36 (6), 1216–1228.

Roshtkhari, M. J., Levine, M. D., 2013. Online dominant and anomalous behavior detection in videos. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2611–2618.

Rostworowski, A., 2012. Developing the intelligent airport. Journal of Airport Management 6 (3), 202–206.

Rother, C., Kolmogorov, V., Blake, A., 2004. "grabcut" - interactive foreground extraction using iterated graph cuts. In: International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH). pp. 309 – 314.

Ryan, D., Denman, S., Fookes, C., Sridharan, S., December 2009. Crowd counting using multiple local features. In: Digital Image Computing: Techniques and Applications. pp. 81 –88.

Ryan, D., Denman, S., Fookes, C., Sridharan, S., 2010. Crowd counting using group tracking and local features. In: 7th IEEE International Conference on Advanced Video and Signal-Based Surveillance. Boston, USA.

Ryan, D., Denman, S., Fookes, C., Sridharan, S., August 2011a. Textures of optical flow for real-time anomaly detection in crowds. In: IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS).

Ryan, D., Denman, S., Fookes, C., Sridharan, S., 2014. Scene invariant multi camera crowd counting. Pattern Recognition Letters 44, 98–112.

Ryan, D., Denman, S., Sridharan, S., Fookes, C., 2011b. Scene invariant crowd counting. In: Digital Image Computing: Technqiues and Applications.

Ryan, D., Denman, S., Sridharan, S., Fookes, C., 2012. Scene invariant crowd counting and crowd occupancy analysis. In: Video Analytics for Business Intelligence. Springer-Verlag, pp. 161–198.

Ryan, D., Denman, S., Sridharan, S., Fookes, C., 2015. An evaluation of crowd counting methods, features and regression models. Computer Vision and Image Understanding 130, 1–17.

Sacchi, C., Regazzoni, C., 2000. A distributed surveillance system for detection of abandoned objects in unmanned railway environments. Vehicular Technology, IEEE Transactions on 49 (5), 2013–2026.

Satta, R., Fumera, G., Roli, F., 12/10/2012 2012. A general method for appearance-based people search based on textual queries. In: First International ECCV Workshop on Re-Identification (ReID 2012). Florence, Italy.

Satta, R., Pala, F., Fumera, G., Roli, F., 2014. People search with textual queries about clothing appearance attributes. In: Person Re-Identification. Springer, pp. 371–389.

Sensormatic, 2012. Sensormatic - Store Business Intelligence - Clarity? II Overhead People Counting. http://www.sensormatic.com/Products/StoreBusinessIntelligence2/ SmartEAS /ClarityII_Overhead_People_Counting.aspx.

Sevilla-Lara, L., Learned-Miller, E., 2012. Distribution fields for tracking. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1910–1917.

Shen, X., Chen, W., Lu, M., 2008. Wireless sensor networks for resources tracking at building construction sites. Tsinghua Science and Technology 13, Supplement 1 (0), 78 – 83.

Singh, A., Sawan, S., Hanmandlu, M., Madasu, V., Lovell, B., sept. 2009. An abandoned object detection system based on dual background segmentation. In: Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 352 –357.

Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C., 2013. Histogram of weighted local directions for gait recognition. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on. IEEE, pp. 125–130.

Stringa, E., Regazzoni, C., 2000. Real-time video-shot detection for scene surveillance applications. Image Processing, IEEE Transactions on 9 (1), 69–79.

Tamersoy, B., Aggarwal, J. K., 2009. Robust vehicle detection for tracking in highway surveillance videos using unsupervised learning. In: Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE Computer Society, Washington, DC, USA, pp. 529–534.

Tang, S., Andres, B., Andriluka, M., Schiele, B., 2015. Subgraph decomposition for multi-target tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5033–5041.

Terada, K., Yoshida, D., Oe, S., Yamaguchi, J., 1999. A method of counting the passing people by using the stereo images. In: International Conference on Image Processing. Vol. 2. pp. 338–342 vol.2.

Thanh, K. N., Sridharan, S., Fookes, C., Denman, S., 2012. Feature-domain super-resolution framework for gabor-based face and iris recognition. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2642–2649.

Thirde, D., Li, L., Ferryman, J., 2006. An overview of the pets 2006 dataset. In: Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. pp. 47–50.

Ting, Z., Zhang, X., Yuanxin, O., sept. 2006. A framework of networked rfid system supporting location tracking. In: 2006 2nd IEEE/IFIP International Conference in Central Asia on Internet. pp. 1 –4.

Tome, P., Fierrez, J., Vera-Rodriguez, R., Nixon, M. S., 2014. Soft biometrics and their application in person recognition at a distance. Information Forensics and Security, IEEE Transactions on 9 (3), 464–475.

Tsai, J., Rathi, S., Keikintveld, C., Ordonez, F., Tambe, M., May 2009. IRIS - a tool for strategic security allocation in transportation networks. In: Proc. of the 8th Int. Conf. on Autonomous and Multiagent Systems. International Foundation for Autonomous Agents and Multiagent Systems, Budapest, HUN, pp. 37–44.

Umakanthan, S., Denman, S., Fookes, C., Sridharan, S., 2014. Supervised latent dirichlet allocation models for efficient activity representation. In: Digital lmage Computing: Techniques and Applications (DlCTA), 2014 International Conference on. IEEE, pp. 1–6.

Vaquero, D., Feris, R., Tran, D., Brown, L., Hampapur, A., Turk, M., dec. 2009. Attribute-based people search in surveillance environments. In: 2009 Workshop on Applications of Computer Vision (WACV). pp. 1 –8.

Velipasalar, S., Tian, Y.-L., Hampapur, A., July 2006. Automatic counting of interacting people by using a single uncalibrated camera. In: International Conference on Multimedia and Expo. pp. 1265–1268.

Vermaak, J., Doucet, A., Perez, P., 2003. Maintaining multi-modality through mixture tracking. In: ICCV. Vol. 2. Nice, France, pp. 1110–1116.

Versichele, M., Neutens, T., Delafontaine, M., de Weghe, N. V., 2012. The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the ghent festivities. Applied Geography 32 (2), 208 – 220.

Vogt, R., Sridharan, S., Jan. 2008. Explicit modelling of session variability for speaker verification. Comput. Speech Lang. 22 (1), 17–38.

Wang, X., Ma, X., Grimson, W. E. L., March 2009a. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. IEEE Trans. Pattern Anal. Mach. Intell. 31, 539–555.

Wang, X., Ma, X., Grimson, W. E. L., March 2009b. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models. IEEE Trans. Pattern Anal. Mach. Intell. 31, 539–555.

Wells, H., Allard, T., Wilson, P., December 2006. Crime and CCTV in Australia: Understanding the relationship. Online Resource.

Welsh, B. C., Farrington, D. P., 2009. Public area CCTV and crime prevention: An updated systematic review and meta-analysis. Justice Quarterly 26 (4), 716–745.

Wen, J., Gong, H., Zhang, X., Hu, W., nov. 2009. Generative model for abandoned object detection. In: 2009 16th IEEE International Conference on Image Processing (ICIP). pp. 853 –856.

Woo, S., Jeong, S., Mok, E., Xia, L., Choi, C., Pyeon, M., Heo, J., 2011. Application of wifi-based indoor positioning system for labor tracking at construction sites: A case study in guangzhou mtr. Automation in Construction 20 (1), 3 – 13.

Wu, P. P.-Y., Pitchforth, J., Mengersen, K., 2014. A hybrid queue-based bayesian network framework for passenger facilitation modelling. Transportation Research Part C: Emerging Technologies 46, 247–260.

Xu, J., Denman, S., Fookes, C., Sridharan, S., September 2012a. Activity analysis in complicated scenes using dft coefficients of particle trajectories. In: IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS).

Xu, J., Denman, S., Fookes, C., Sridharan, S., September 2012b. Unusual scene detection using distributed behaviour model and sparse representation. In: IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS).

Xu, J., Denman, S., Fookes, C. B., Sridharan, S., 2015. Detecting rare events using kullback-leibler divergence. In: The 40th IEEE International Conference on Acoustics, Speech and Signal Processing.

Yang, C., Duraiswami, R., Davis, L., 2005. Fast multiple object tracking via a hierarchical particle filter. In: Tenth IEEE International Conference on Computer Vision. pp. 212 – 219.

Ye, G., Liao, W., Dong, J., Zeng, D., Zhong, H., 2015. A surveillance video index and browsing system based on object flags and video synopsis. In: MultiMedia Modeling. Springer, pp. 311–314.

Zach, C., Pock, T., Bischof, H., 2007. A duality based approach for realtime tv-l1 optical flow. In: Pattern Recognition (Proc. DAGM). Heidelberg, Germany, pp. 214–223.

Zhang, C., Li, H., Wang, X., Yang, X., 2015a. Cross-scene crowd counting via deep convolutional neural networks. In: Proc. CVPR.

Zhang, T., Liu, S., Xu, C., Yan, S., Ghanem, B., Ahuja, N., Yang, M.-H., 2015b. Structural sparse tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 150–158.

Zhang, T., Ouyang, Y., He, Y., Apr. 2008. Traceable air baggage handling system based on rfid tags in the airport. J. Theor. Appl. Electron. Commer. Res. 3 (1), 106–115.

Zhang, X., Zhang, L., 2014. Real time crowd counting with human detection and human tracking. In: Neural Information Processing. Springer, pp. 1–8.

Zhang, Z., Wang, M., Geng, X., 2015c. Crowd counting in public video surveillance by label distribution learning. Neurocomputing.

Zhao, T., Nevatia, R., 2004. Tracking multiple humans in complex situations. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (9), 1208–1221.