2016 IEEE International Symposium on Multimedia

Smart Stadium for Smarter Living: Enriching the Fan Experience

Sethuraman Panchanathan, Shayok Chakraborty, Troy McDaniel, Matt Bunch
Center for Cognitive Ubiquitous Computing (CUbiC)
Arizona State University
Tempe, AZ
{panch, shayok.chakraborty, troy.mcdaniel}@asu.edu

Noel O'Connor, Suzanne Little, Kevin McGuinness, Mark Marsden
Insight Centre for Data Analytics
Dublin City University
Dublin, Ireland
{noel.oconnor, suzanne.little}@dcu.ie

***Abstract*—Rapid urbanization has led to more people residing in cities than ever before, and projections estimate that 64% of the global population will be urban by 2050. Cities are beginning to explore Smart City initiatives to reduce expenses and complexities while increasing efficiency and quality of life for its citizens. To achieve this goal, advances in technology and policies are needed together with rethinking traditional solutions to transportation, safety, sustainability, among other priority areas. We propose the use of a Smart Stadium as a 'living laboratory' to identify, deploy and test Internet of Things technologies and Smart City solutions in an environment small enough to practically trial but large enough to evaluate effectiveness and scalability. The Smart Stadium for Smarter Living initiative brings together Arizona State University, Dublin City University, Intel Corporation, Gaelic Athletic Association, Sun Devil Stadium and Croke Park to explore smart environment solutions.**

Keywords-Internet of Things; smart environments; smart stadium; smart cities; crowd behavior analytics; object counting

I. INTRODUCTION

Within the last few decades, our planet has undergone rapid urbanization. In 2014, 54% of the global population was urban, and by 2050, 64% of the global population will be urban [1] with this percentage projected to steadily rise. To handle population and city growth, cities are seeking ways to reduce complexity and expenses while improving efficiency and quality of life for its citizens. Cities that perform well flourish by creating wealth and increasing productivity, enabling a pathway to growth and success [2]. Such a transformation relies upon advancements in both technology and policies for cities to reimagine their traditional approaches to sustainability, communication, transportation, citizen safety, security, and citizen engagement toward better urban development and improved quality of living for citizens. Such cities are now termed 'Smart Cities'. To identify, develop and evaluate Internet of Things (IoT) and Communications technologies and applications that have the potential to scale and be useful within a smart city, we propose the use of a Smart Stadium as a living laboratory that is small enough to practically trial IoT technologies while large enough to verify their effectiveness and capability to scale.

Smart Stadium for Smarter Living is an initiative that joins Arizona State University (ASU), Dublin City University (DCU), Intel Corporation, Gaelic Athletic Association (GAA), Sun Devil Stadium in Tempe, AZ, and Croke Park in Dublin, Ireland. The aim of the initiative is to investigate the usefulness, effectiveness and scalability of IoT technologies and applications for smart cities through their deployment and testing within world class smart stadia testbeds: Croke Park and Sun Devil Stadium. The projects of the initiative may be divided between efforts focused on the fan (i.e., stadium attendee) or the stadium. Projects utilize a variety of deployed sensors, including video cameras and microphones, to address issues of fan engagement, crowd management, event logistics, stadium management, pitch maintenance, and environmental monitoring. Given the sheer scope of this project, we limit discussions in this paper to projects focused on the fan.

A fan's experience includes not only his or her interactions and activities within the stadium, but the entire 'journey' around attending an event at the stadium. This journey may include preparation for attending an upcoming event, travel to and from the stadium, presence on social media, and of course activities within the stadium as well as related events before/during/after the main event. This paper presents three fan-focused projects targeting priority areas of convenience, safety and engagement: (1) Real-time access to information about wait times across restrooms and concession stands within the stadium; (2) Monitoring of crowd movement and activity; and (3) Interactive games-within-a-game.

The rest of this paper is organized as follows: Section 2 presents project overviews for the three projects. Section 3 presents implementation details for each project. Section 4 presents preliminary results for each project. Section 5 provide final thoughts and proposes directions for future work.

II. ENRICHING FAN EXPERIENCES: PROJECT OVERVIEWS

A. *Wait Time and Queue Estimation*

The objective of this project is to enrich the fan experience by providing access to wait times at restrooms and concession stands via a mobile app. Such a technology

978-1-5090-4571-6/16 $31.00 © 2016 IEEE
DOI 10.1109/ISM.2016.128

will allow fans to maximize their time watching and enjoying a game rather than waiting in long lines during the course of a game. We adopt a computer vision based approach to count the number of people in a queue. We assume the presence of cameras in strategic locations in the vicinity of restrooms and concession stands; the video feed from these cameras is analyzed to accurately estimate the count of people in the queues. Once the count is obtained, wait times can be obtained from the average service time per person. In this paper, we present preliminary results of vision-based object counting on several publicly available datasets. Besides the prediction accuracy, we also analyze the average prediction time per frame and its ability to operate reliably in challenging, unconstrained environments.

Counting the number of objects in images is a problem of paramount practical importance and arises in myriads of real-world applications including crowd behavior monitoring, security and surveillance, estimating the number of cells in microscopic images and counting the number of trees from the aerial image of a forest. Unsupervised counting algorithms relying on self-similarities [3] or motion similarities [4] are limited in their counting accuracy, which has paved the way for supervised learning approaches to object counting. Among the supervised approaches, ***counting by detection*** algorithms employ object detectors to localize individual object instances within an image and derive the count from the localizations [5][6]. While these approaches achieve higher accuracy compared to unsupervised methods, these techniques need to solve object detection, which is a challenging computer vision problem, especially for overlapping instances. ***Counting by regression*** algorithms avoid solving the hard detection problem and attempt to learn a direct mapping from some global image feature to the number of objects. Neural networks have been extensively used to learn such mappings [7][8]. These approaches, however, discard any information about the spatial locations of the objects in images and use only its 1-dimensional statistics (total number) for learning. They also have the implicit assumption that the density is roughly uniform regardless of the location where the feature is computed. This is largely invalid in most real-world scenarios due to changes in viewpoint and in crowd density. The drawbacks of global feature regression can be overcome by relaxing this assumption. Lempitsky and Zisserman [9] proposed to divide an image into cells and perform regression individually for each cell. Chen et al. [10] recently postulated that information sharing among regions should lead to a robust and improved counting performance and proposed a single multi-output model for joint localized crowd counting based on ridge regression. ***Counting by segmentation*** can be regarded as a hybrid of counting-by-detection and counting-by-regression. They segment the objects into separate clusters and learn regression models from the global properties of each cluster to the number of objects in it [11][12].

The learning based object counting algorithm proposed by Lempitsky and Zisserman [9] has been empirically shown to depict commendable performance under challenging real-world settings. The framework can accept any domain specific visual features; the feature extraction and inference steps are computationally very efficient, which makes it a promising candidate for applications involving real-time processing or dealing with large amounts of visual data. We therefore use this framework for our people counting application.

B. *Crowd Understanding*

Stadiums are locations where thousands of people will gather for events ranging from competitions to concerts to conferences. Understanding and predicting how this crowd of people moves around the stadium is useful for both security and logistics, and can help stadium management to: ensure that fans have a safe and enjoyable experience; manage and optimize staffing levels; and better deploy support staff when abnormal (anomalous) events occur. Both Sun Devil Stadium (56,200) and Croke Park (82,300) have been designed to ensure people can move safely and efficiently, but the Smart Stadium project aims to exploit visual and other sensor data to analyze how crowds move both historically and in real-time.

The crowd understanding project uses existing CCTV camera footage from Croke Park to measure crowd features and determine the relative occupancy of an area ("crowd density") and track the motion of the crowd. Long-term, the system aims to learn a "steady state" of what normal crowd movement and density patterns look like and therefore be able to determine when crowds don't behave according to expected patterns and alert support staff.

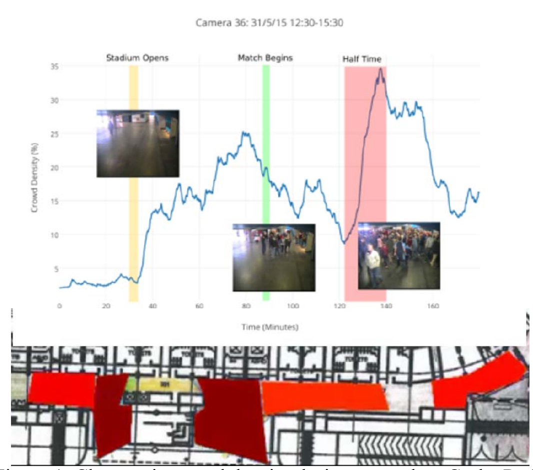

Figure 1. Changes in crowd density during a match at Croke Park.

Figure 1 shows an illustration of crowd density calculated from five CCTV cameras covering a concession area at Croke Park during a busy match. The peaks and troughs as patrons come through the concession area before the match starts and return at half-time can be clearly identified. The heatmap below is taken from a video animation that displays the changing crowd density estimation over time.

C. Victory Cheer

Victory Cheer is an interactive game-within-a-game for fans to enhance their involvement at events, increase camaraderie, and improve their overall fan experience. The game involves stadium sections competing against each other by cheering the loudest for their team, e.g., the ASU Sun Devils. Fans will be able to see which section is winning by watching the stadium's Jumbotron. Once the game ends, a push notification is sent to the winners within the winning section.

III. ENRICHING FAN EXPERIENCES: IMPLEMENTATION

A. Wait Time and Queue Estimation

The algorithm assumes the presence of dotted annotation images, where a single dot is put on each object instance in each image. This is depicted in Figure 2. Dotting or pointing is a natural way to count objects. The dots provide useful information about the spatial distribution of objects in an image, which can be exploited to derive a more robust estimate of the count.

Figure 2. A dotted annotation image.

Given an image ***I***, the algorithm computes a density function ***F*** as a real function of pixels in the image. From the density function ***F***, the number of objects in the image is estimated by integrating ***F*** over the entire image ***I***. Further, integrating over an image sub-region ***S*** yields an estimate of the count of objects in that sub-region.

The algorithm assumes that each pixel p is represented by a feature vector x_p and models the density function as a linear transformation of x_p: $F(p) = w^T x_p$. Given a set of training images, the weight vector ***w*** is learned using a regularized risk framework so that the density estimates of the training images matches the ground truth density obtained from the user annotations.

Counting Posed as a Learning Problem: Formally, we are given a set of ***N*** training images $I_1, I_2, \ldots I_N$. Each training image I_i is annotated with a set of 2D points $P_i = \{P_1, P_2, \ldots P_{C(i)}\}$, where ***C(i)*** is the total number of objects annotated by the user. For training image I_i, the ground truth density function is defined to be a kernel density estimate based on the provided points. A normalized 2D Gaussian kernel was used to estimate the density in this work:

$$F_0^i(p) = \sum_{P \in P_i} N(p; P, \sigma^2 \mathbf{1}_{2\times 2}), \forall p \in I_i \quad (1)$$

Here, ***p*** denotes a pixel and ***N()*** denotes a normalized 2D Gaussian kernel evaluated at ***p***, with the mean at the user placed dot ***P*** and an isotropic covariance matrix with a small standard deviation (typically, a few pixels). Given a set of training images together with their ground truth densities, we aim to learn a linear transformation of the feature representation that approximates the density function at each pixel:

$$F_i(p|w) = w^T x_p^i, \ \forall p \in I_i \quad (2)$$

where ***w*** is the weight vector that needs to be learned from the training data and $F_i(.|w)$ is the estimate of the density function for a particular value of ***w***. The regularized risk framework selects ***w*** so that it minimizes the sum of the mismatches between the ground truth and the estimated density functions under regularization:

$$w = argmin_w\{w^T w + \lambda \sum_{i=1}^{N} D(F_i^0(.), F_i(.|w))\} \quad (3)$$

Here, λ is a weight parameter controlling the relative importance of the two terms and is the only hyper-parameter in the framework. Once the optimal weight vector is learned from the training data, the system can produce the density estimate of an unknown image by a simple linear weighting of the feature vector computed in each pixel, as in Equation (2). The problem therefore reduces to selecting an appropriate loss function ***D*** and computing the optimal ***w*** under that loss, as depicted in Equation (3).

The MESA distance: The distance ***D*** measures the mismatch between the ground truth and estimated densities and has a significant effect on the performance of the framework. Lempitsky and Zisserman [9] suggested a metric called the ***MESA*** (Maximum Excess over SubArrays) distance to compute the mismatch. Given an image ***I***, the MESA distance between two functions ***F1(p)*** and ***F2(p)*** on the pixel grid is defined as the maximum absolute difference between sums of ***F1(p)*** and ***F2(p)*** over all box sub-arrays in ***I***:

$$D_{MESA}(F_1, F_2) = \max_{B \in \mathbf{B}} |\sum_{p \in B} F_1(p) - \sum_{p \in B} F_2(p)| \quad (4)$$

Here, $\boldsymbol{B}$ is the set of all box sub-arrays of $\boldsymbol{I}$. The MESA distance is a metric and has a number of desirable properties. It is robust to the additive local perturbations such as independent noise or high frequency signals as long as the integrals of these perturbations over large regions are close to 0. Thus, it does not matter how the ground truth density is defined locally, as long as the integrals of the ground truth densities over larger regions reflect the count correctly. Further, the MESA distance can be computed exactly via an efficient combinatorial algorithm (maximum sub-arrays).

Solving the Optimization Problem: Plugging the definition of the MESA distance into Equation (3), we note that the learning problem can be rewritten as a convex quadratic program:

$$\min_{w,\xi_1,\ldots\xi_N} w^T w + \lambda \sum_{i=1}^{N} \xi_i$$

$$s.t.\ \ \xi_i \geq \sum_{p \in B} (F_i^0(p) - w^T x_p^i),$$

$$\xi_i \geq \sum_{p \in B} (w^T x_p^i - F_i^0(p)), \forall i, \forall B \in \mathbf{B}_i \tag{5}$$

Here ξ_i are the auxiliary slack variables (one for each training image) and $\boldsymbol{B_i}$ is the set of all sub-arrays in image $\boldsymbol{i}$. At optimum, the optimal vector $\boldsymbol{w^*}$ is the solution of Equation (3) while the slack variables equal the MESA distances. The number of linear constraints is combinatorial and so a custom QPsolver cannot be applied directly. However, the cutting plane procedure finds the close approximation to the globally optimal $\boldsymbol{w}$ after a small number of iterations. Please refer [9] for more details.

B. *Crowd Understanding*

The developed technique for crowd behavior anomaly detection uses a set of efficiently computed, easily interpretable, scene-level holistic features [13]. This low-dimensional descriptor combines two features from the literature: crowd collectiveness [14] and crowd conflict [15], with two newly developed features: mean motion speed and a unique formulation of crowd density [13]. Crowd collectiveness can be defined as the degree to which individuals in a scene move in unison. Crowd conflict refers to the level of friction/interaction observed between people. Crowd density is the level of congestion observed across a scene at a given instant while mean motion speed is the mean magnitude of all local motion vectors in the current frame. Each of these features capture a distinct aspect of crowd behavior.

Our holistic features are extracted for each frame in a given video sequence using the following steps. Firstly, the scene foreground is segmented using the Gaussian-mixture based method of KaewTraKulPong and Bowden [16] before interest points are tracked using a KLT tracker [17]. These local trajectories or tracklets are then analyzed to calculate 4 holistic features for each frame. This high-level descriptor of crowd behavior can be computed in real-time (30+ frames per second) even on commodity hardware (e.g., an Intel i5 CPU).

Anomalous crowd behavior then needs to be detected using this descriptor. There are two situations to consider when we train a machine learning model to perform this task. When only *normal* behavior training data is available, we use a Gaussian Mixture Model (GMM) for outlier detection. When both *normal* and *abnormal* training data is available, we use a Support Vector Machine (SVM) for binary classification.

C. *Victory Cheer*

Victory Cheer uses custom built sensor packs, distributed in sections of the ASU Sun Devil Stadium, which include GPS, altimeter, temperature sensor, piezo vibration sensor, Bluetooth low energy and a decibel meter. These sensors are connected to an Intel Edison board, leveraging a lightweight Internet protocol (MQTT), in conjunction with a user agent, to stream data to a gateway device where data is generically processed, packaged, and streamed to the cloud. Using a Pub/Sub Service, the data is exposed to the world IE including the ASU Mobile App and GameDay Applications. The process of streaming a large volume of data and processing the data on the fly was a major challenge. After disabling devices from retrying to send lost packet information, to not hold up the high speed data flow, data could be processed at a sufficient rate of 125ms (1/8th of a second) per complete data set (44 sensor packs).

IV. ENRICHING FAN EXPERIENCES: RESULTS

A. *Wait Time and Queue Estimation*

To depict the generalizability of the object counting algorithm, we validated it on three challenging publicly available datasets from different application domains (sample images are shown in Figure 3):

1. The bacterial cells dataset: This dataset contains 200 synthetic images, emulating microscopic views of the colonies of bacterial cell [18]. Such synthetic images are highly realistic and simulate effects such as cell overlaps, shape variability, out-of-focus blur, vignetting, etc.

2. The UCSD pedestrians dataset: This is a video dataset, containing 2000 frames from a camera overviewing a busy pedestrian street [11]. The dataset also contains the dotted ground truth annotations for these frames, the position of the ground plane and the region of interest where the count should be performed.

3. The mall dataset: This dataset contains 2000 video frames collected using a publicly accessible webcam for crowd counting and profiling research [19]. It also contains dotted ground truth annotations of the head positions of every pedestrian in all the frames.

(a) Cells (b) UCSD Pedestrians (c) Mall

Figure 3. Sample images from the three datasets.

The features used were based on dense SIFT descriptors [20] computed at each pixel with fixed SIFT frame radius and fixed orientation. Each dataset was randomly split into a training set and a test set (of equal size). The models were trained on the training set; the performance was evaluated on the test set in terms of the mean absolute error (MAE). We also noted the approximate prediction time for each test image and the total time taken to train the models. The algorithm was implemented in MATLAB R2014a on a desktop (running Windows 7) with 16GB RAM and 3.60 GHz Intel Core Processor. We studied the performances of two types of regularized models: L1 regularization and Tikhonov regularization. The results are reported in Table 1.

Table 1: Performance of the object counting algorithm.

Dataset	Train Size	Test Size	Image Resolution	MAE (L1)	MAE (Tikhonov)	Test Time per Image	Training Time
Cells	100	100	256 X 256	7.73±5.59	6.36±4.36	1.19 secs	1 min 45 secs
UCSD	1000	1000	238 X 158	6.17±5.14	4.66±4.20	0.71 sec	19 mins 34 secs
Mall	1000	1000	640 X 480	3.42±2.56	3.48±2.44	5.54 secs	70 mins 38 secs

We note that both models consistently depict low error values across all three datasets (the Tikhonov regularized model marginally outperforms the L1 regularized model for the Cells and UCSD datasets). This corroborates the tremendous promise and potential of the algorithm for real-world object counting applications. The training time increases with the resolution of the image and the size of the training set. However, in our application of wait time estimation, the training will be performed offline and hence, the training time is not a major concern. The test time (prediction time) per image primarily depends on the image resolution. We note that in our application, real-time performance is not a crucial requirement; from a practical point of view, it suffices to convey the wait time at a particular restroom/concession stand every minute (or 30 secs) so that fans can select the best option at a given time. In summary, the proposed object counting algorithm is a very promising candidate to estimate the wait times in restrooms and concession stands in a smart stadium application.

A visual demonstration of the performance of the algorithm can be downloaded using the following link: https://www.dropbox.com/sh/1ssndoz1w11fny0/AAAvlWXp8awAJHweE_HKcmM2a?dl=0

The demo shows the predictions made by the two models (M1 and M2) on each frame of two test videos, together with the ground truth values.

B. Crowd Understanding

The proposed method for crowd behavior anomaly detection is evaluated on two distinct datasets: i) the UMN dataset [21], and ii) the violent-flows dataset [15].

The UMN dataset contains 11 sequences filmed in 3 different locations. Each sequence begins with a period of normal passive crowd behavior before a panic event/anomaly occurs towards the end. Only normal behavior training data is provided, meaning our GMM-based outlier detection approach is used. The goal of this task is to detect the anomalous frames. Our technique achieves very competitive performance on this dataset, with an AUC (Area under the curve) of 0.92 reported [13].

The violent-flows dataset contains 246 clips of violent (abnormal) and non-violent behavior. Performance is evaluated using a 5-fold cross validation with anomaly detection performed at the clip level. As both normal and abnormal training data is provided, our SVM-based binary classification approach is used. A state-of-the-art accuracy of 85.53±0.17% is reported on the violent-flows dataset [13]. Greater than real-time processing speed (40 frames per second) is achieved on commodity hardware (an Intel i5 CPU) for all experiments on the UMN and violent-flows datasets [13].

C. Victory Cheer

The results of the victory cheer project can be broken into three main result categories: 1) the mobile application, 2) the installation of the Intel hardware and transmittal of data to the IBM cloud, and 3) the Jumbotron application.

The mobile application has been completed and deployed to the appropriate app stores. The application was successful in terms of meeting its required technical specifications; however, there was minimal fan engagement. It is believed that this was either due to individuals not updating their applications, connectivity issues, or simply that they did not have the application installed on their mobile devices. Intel hardware was installed and pushed sound data to the affiliated Intel gateway devices where the data was aggregated and streamed over MQTT to the IBM cloud.

There are a total of 44 sensor packs, and 4 lower bowl mics. The lower bowl mics are used to collect sound data from the sections without Intel sensor packs. Together, the sensors allow for complete sound data collection from the lower bowl. Figure 4 depicts the Intel sensor box.

Finally, the Jumbotron application successfully pulled data from the IBM cloud and displayed it in a real-time sound graph on the stadium Jumbotron. Some difficulties arose in the consistency of data being streamed into the app from the IBM cloud. This required collaboration with Intel to smooth out their sample rate and turn off data retries. Once that change occurred, the results were much smoother and faster (from 1/4 second to 1/8 second). The Jumbotron app will continue to be utilized throughout different sporting events, and working toward providing awards to participants that can be redeemed from smart vending machines.

Figure 4. Intel sensor box.

V. Conclusion and Future Work

This paper presented three projects within the Smart Stadium for Smarter Living initiative focused on enriching the stadium attendee experience through improved convenience, safety and engagement. Our preliminary results demonstrate the potential of these technologies for smart city solutions and the usefulness of investigating technologies within smaller testbeds, such as a smart stadium.

As part of future work, we are continuing to explore crowd behavior analytics within the stadium setting to improve safety and security. We are also investigating smart solutions to address issues of traffic and parking that often accompany large stadium events. We also plan to further explore smart stadium solutions focused on the stadium itself targeting priority areas of energy optimization and sustainability.

Acknowledgment

The authors thank Intel Corporation, Arizona State University and Dublin City University for their funding support. This material is partially based on work supported by the Intel Corporation under grant Joint Path Finding (JPF) Proposal: Smart Stadium and Smart Living Research.

References

[1] U. Nations, "World urbanization prospects: The 2014 revision, highlights," Department of Economic and Social Affairs, 2014.

[2] S. Hodgkinson, "Is your city smart enough? Digitally enabled cities and societies will enhance economic, social, and environmental sustainability in the urban century," OVUM report, 2011.

[3] N. Ahuja and S. Todorovic, "Extracting texels in 2.1d natural textures," in 2007 IEEE 11th International Conference on Computer Vision, 2007, pp. 1–8.

[4] V. Rabaud and S. Belongie, "Counting Crowded Moving Objects," in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 2006, vol. 1, pp. 705–711.

[5] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09), 2009, pp. 304–311.

[6] S.-F. Lin, J.-Y. Chen, and H.-X. Chao, "Estimation of number of people in crowded scenes using perspective transformation," IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 31, no. 6, pp. 645–654, Nov. 2001.

[7] D. Kong, D. Gray, and H. Tao, "A viewpoint invariant approach for crowd counting," in 18th International Conference on Pattern Recognition (ICPR'06), 2006, vol. 3, pp. 1187–1190.

[8] D. Kong, D. Gray, and H. Tao, "Counting pedestrians in crowds using viewpoint invariant training," in British Machine Vision Conference (BMVC'05), 2005.

[9] V. Lempitsky and A. Zisserman, "Learning to count objects in images," in Neural Information Processing Systems (NIPS'10), 2010.

[10] K. Chen, C. Loy, S. Gong and T. Xiang, "Feature mining for localized crowd counting", in British Machine Vision Conference (BMVC'12), 2012.

[11] A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08), 2008, pp. 1–7.

[12] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, "Crowd counting using multiple local features", in Digital Image Computing: Techniques and Applications, 2009.

[13] M. Marsden, K. McGuinness, S. Little, and N. E. O'Connor, "Holistic features for real-time crowd behaviour anomaly detection," in IEEE International Conference on Image Processing (ICIP'16), 2016, pp. 918–922.

[14] J. Shao, K. Kang, C. C. Loy, and X. Wang, "Deeply learned attributes for crowded scene understanding," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15), 2015, pp. 4657–4666.

[15] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2012, pp. 1–6.

[16] P. KaewTraKulPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in Video-Based Surveillance Systems, P. Remagnino, G. A. Jones, N. Paragios, and C. S. Regazzoni, Eds. Springer US, pp. 135–144, 2002.

[17] C. Tomasi and T. Kanade, "Detection and tracking of point features," Technical Report CMU-CS-91-132, 1991.

[18] A. Lemmussola, P. Ruusuvuori, J. Selinummi, H. Huttunen, and O. Yli-Harja, "Computational framework for simulating fluorescence microscope images with cell populations", IEEE Transactions on Medical Imaging, vol. 26, no. 7, pp. 1010-1016, Jul. 2007.

[19] K. Chen, S. Gong, T. Xiang, and C. Loy, "Cumulative attribute space for age and crowd density estimation", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR'13), 2013, pp. 2467-2474.

[20] D. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision (IJCV), vol. 60, no. 2, pp. 91-110, 2004.

[21] http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi