

Copyright © 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# Detection of Dynamic Background due to Swaying Movements from Motion Features

Duc-Son Pham\*, *Member, IEEE*, Ognjen Arandjelović, and Svetha Venkatesh, *Senior Member, IEEE*

**Abstract**—Dynamically changing background (‘dynamic background’) still presents a great challenge to many motion-based video surveillance systems. In the context of event detection, it is a major source of false alarms. There is a strong need from the security industry either to detect and suppress these false alarms, or dampen the effects of background changes, so as to increase the sensitivity to meaningful events of interest. In this paper, we restrict our focus to one of the most common causes of dynamic background changes: that of swaying tree branches and their shadows under windy conditions. Considering the ultimate goal in a video analytics pipeline, we formulate a new dynamic background detection problem as a signal processing alternative to the previously described but unreliable computer vision-based approaches. Within this new framework, we directly reduce the number of false alarms by testing if the detected events are due to characteristic background motions. In addition, we introduce a new dataset suitable for the evaluation of dynamic background detection. It consists of real-world events detected by a commercial surveillance system from two static surveillance cameras. The research question we address is whether dynamic background can be detected reliably and efficiently using simple motion features and in the presence of similar but meaningful events such as loitering. Inspired by the tree aerodynamics theory, we propose a novel method named local variation persistence (LVP), that captures the key characteristics of swaying motions. The method is posed as a convex optimization problem whose variable is the local variation. We derive a computationally efficient algorithm for solving the optimization problem, the solution of which is then used to form a powerful detection statistic. On our newly collected dataset, we demonstrate that the proposed LVP achieves excellent detection results and outperforms the best alternative adapted from existing art in the dynamic background literature.

**Index Terms**—dynamic background, detection algorithms, shadow, motion-based analysis, convex optimization, ADMM, sparsity learning, ROC, mixture of Gaussians

## I. INTRODUCTION

A dynamically changing background (or just ‘dynamic background’ for short) in a video presents a major challenge

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

Manuscript received March 17, 2014; revised August 11, 2014 and November 9, 2014; accepted November 30, 2014; date of current version December 1, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nikolaos Boulgouris.

D.-S. Pham is with Department of Computing, Curtin University, PO Box U1987, WA 6845 (email: [dspham@ieee.org](mailto:dspham@ieee.org)).

O. Arandjelović and S. Venkatesh are with the School of Information Technology, Deakin University, Geelong, VIC 3125, Australia (email: [svetha.venkatesh@deakin.edu.au](mailto:svetha.venkatesh@deakin.edu.au)).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>

Digital Object Identifier 10.110/TIP.2014.XXXXXX

to numerous machine vision tasks, confounding the extraction of useful information about moving objects of interest [31]. Surveillance is one of the areas with the most urgent need for methods which can improve the robustness of video analytics in the presence of this challenge [12]. The literature in the last decade has seen a considerable progress in the modeling of dynamic background and the extraction of foreground objects. The first class of techniques is based on the statistical modeling of temporal data, which started with the seminal work of Stauffer and Grimson [27]. They modeled the variation in background appearance using an underlying mixture of Gaussians (MoG), which can be seen as a simplified description of the multi-modality of video data. The background is described by mixture components that account for major energy contribution in the distribution. A foreground pixel can be detected by computing its likelihood conditional on the identified background components. Since then, MoG has inspired many extensions. Some of these examined more robust representations, such as dynamic texture [8] and local binary patterns [5]. Others sought to model more complex statistical patterns within the data, e.g. by using nonparametric density estimation techniques rather than the parametric MoG approach [11] [25], or by modeling the correlation of nearby pixels [16]. The second class of techniques analyzes the subspace structure of the background [2], some of which may include an internal predictor [19], [29]. Finally, more recent methods look at a direct foreground-background discrimination by exploiting both spatial and temporal information. In particular, a number of methods examined the difference between the motions of the background and the foreground [17], [18]. This discrimination is quantified by computing the mutual information between a location of interest and its locally surrounding areas. However, this approach is more suitable for detecting small objects in a known persistent dynamic background. Another work based on machine learning is proposed in [9], where a structure SVM is derived in an incremental setting. Dynamic background is therefore an active research area.

Despite these advances in dynamic background modeling and subtraction, the existing methods are not sufficiently reliable for real-time surveillance. A recent survey [5] compares pixel-level background subtraction techniques (including those formulated specifically for dealing with a dynamic background) using semi-synthetic data. Dynamic background sets they looked at included moving tree branches and changing traffic lights. All existing methods failed to reach performance which would be satisfactory in practice. Another interesting finding of this work was that simpler methods, such as MoG, actually performed better, though they are still far from being

sufficiently reliable for practical use. Lastly, Brutzer *et al.* highlighted a specific challenge posed by shadows [5]. All examined methods struggled to identify weak shadows and frequently misclassified strong shadows as foreground objects. Thus, dynamic background subtraction is still an unsolved problem in machine vision.

In this paper, we suggest a new perspective for addressing the problem posed by a dynamic background, and propose an effective solution. This work is motivated from our exposure to video security through our industry partner. One of the common problems with motion-based video analytics for static cameras is the sudden change in the background statistics. When that happens, non-robust motion-based anomaly detection methods may experience an increase in false alarms. This is because the underlying algorithm, much like all work in the published literature, implicitly assumes a static or a slowly changing background. Thus, the outstanding research challenge is to suppress these false alarms while having minimum computational or structural changes to the existing analytics infrastructure.

To address this, we suggest a new paradigm, which is best appreciated when considering the whole video analytics processing chain. Here, instead of tackling the dynamic background problem at the front end like what the existing methods attempt to do (as described previously, with limited success and significant added computational cost), we approach the problem of detecting background changes at the back end, thereby filtering out false alarms. There are important practical reasons for this alternative proposal:

- *Scarcity of Background Changes:* In many practical situations faced by our industry partner, background changes are neither *permanent* nor of a *fixed* type as assumed in most previous works. Rather, they appear rarely and intermittently. Moreover, there may be a combination of different factors simultaneously in a scene, such as the moving branches and their long cast shadows, reflections in puddles, cobweb build-up in front of cameras under low-light and wind, and hostile weather conditions such as rain or snow. In these cases, it is often impossible to obtain perfect data for the modeling of the background. In addition, this also suggests that continuously running existing dynamic background subtraction algorithms, which are computationally expensive and often unreliable, would create an unnecessary computational burden.
- *Surveillance End-Goal:* An important goal of surveillance video analytics is to control the number of false alarms while maintaining a good detection rate of meaningful events. In the situation faced by our industry partner, the motion-based surveillance analytics raise reasonable alarms when there is no dynamic background. A large number of false alarms are typically raised only when there is a sudden onset of background changes due to extreme conditions. Thus, discarding false alarms due to rare and prominent dynamic background changes would be more sensible in practice.

To make the existing video surveillance analytics more

robust against infrequent dynamic background interference, we formulate a *new* dynamic background detection problem. Here, we focus on the dynamic background with repetitive motions, which often cause unwanted alarms in motion-based analytics. Given motion features from video which correspond to an event, it is desirable to detect if prominent background changes are present in any part of the scene during that event. This is useful for determining whether the raised alarms are purely due to a sudden onset of background changes. Here, we restrict our attention to mostly static scenes and motion-based analytics. By addressing the detection problem in the event post-processing stage, the overall computational cost can be reduced as detection is only needed when an alarm is raised and that dynamic background occurs infrequently and intermittently. In addition, we directly target the end goal of a video analytic system, thus leaving front-end processing blocks unchanged.

The first contribution in this work is an introduction of a new problem that we view as a practical alternative to reducing false alarms in motion-based analytics due to the presence of a dynamic background. To the best of our knowledge, no previous work directly solves a similar problem. Here, the focus is not on the foreground, but the presence of dynamic background. Additionally, it is an event-based post-processing approach rather than the setting used in the corpus of previous work on background modeling. Instead of working with raw pixel data, we address the detection problem when only motion features are available as an input.

Our second contribution is a new dynamic background detection dataset from real-world sites and a novel detection method. The research question is whether we can characterize the signature of dynamic background with repetitive motions and differentiate it against other potentially similar patterns of interest such as loitering. The key finding, which is the most important contribution in this work, is that temporal persistence of local variation seems to be an important clue for recognizing repetitive motion of the dynamic background. This is fundamentally different from most previous works on dynamic background, wherein the temporal dimension is either ignored or treated lightly. The following two papers, though solving different problems, are an exception. One is a work on object tracking from an underwater camera [1] which exploits optical distortion statistics. The other is on object tracking from moving camera under turbulence condition [21] which exploits an effective theory called particle advection. Unfortunately, due to an entirely different physical modeling, it is impossible to apply these works to the swaying motions in this work. Here, we propose a novel method specifically for the detection of swaying motions of trees and their shadows, which we term local variation persistence (LVP). The method first extracts the temporal variation through a robust convex optimization, which is solved by a computationally efficient algorithm. Then a novel detection statistic is formed as the normalized  $\ell_1$ -norm of the local variation vector. On the new evaluation data, we demonstrate that our approach provides a considerable detection gain over an adaptation of the well-known Gaussian mixture model by Stauffer and Grimson in the dynamic background literature.

The remainder of this paper is organized as follows. In Section II, we describe the proposed problem setting in detail and highlight how it differs from the setting adopted by previous works. Section III describes an adaptation of the classic Stauffer & Grimson background modeling to the problem setting described in Section II. Then we develop the new detection algorithm that is based on the temporal persistence of the local variation in Section III. Section IV describes the data used for evaluation. Section V presents experimental results. Finally, a summary of the paper and the key conclusions are given in Section VI.

The data set and the code implementing the methods discussed in this work are made publicly available at <https://sites.google.com/site/dspham>.

## II. PROBLEM STATEMENT

We next describe the specific settings imposed in this work and the definitions of some key technical terms. Visual illustrations can be found in Section IV. The overall scenario is video surveillance of a scene by a static camera. Here, we consider the front-end setting where each frame is divided by a grid of size  $20 \times 20$  into  $P = 400$  equal non-overlapping rectangular blocks  $p = 1, 2, \dots, P$ . Each block corresponds to a spatial region of the scene, which we call a *cell*. Instead of raw pixel values, the available data input is a vector/matrix of motion quantities computed and collected from these  $P$  cells, which we call *motion features*. Though motion features can be any suitable measure of motion statistics (in a sense which is further discussed in Section III), the specific *motion feature*  $x_i$  in each of the  $P$  cells considered in this work is the number of optical flow vectors (i.e., *motion count*) whose magnitude is greater than a noise threshold<sup>1</sup>. The choice of this specific motion feature is largely imposed by the operational framework of our industry partner as optic flow vectors can be computed efficiently and reliably through recent advances in GPU hardware and are more suitable for complex scenes and non-tracking approaches [24].

We define *events* as the alarms generated by a surveillance analytics system. Each event is a collection of consecutive video frames with a start time and an end time. The duration of an event typically ranges from half a minute to several minutes. Given the above front-end processing, each event is represented by a number of corresponding motion count vectors. Each entry in a motion count vector corresponds to a cell, and the motion observed in this cell may be due to foreground, short-term static or dynamic background. In this work, we restrict our attention to swaying tree branches and their shadows. Unless otherwise explicitly specified, the term *dynamic background* refers only to this special type, which is common in many surveillance applications. The goal is to detect if dynamic background is present in these  $P$  cells in a sense that it lasts for a total of more than 50% of the event duration. This allows for some foreground or period of inactivity (purely static background).

<sup>1</sup>A motion feature is not directional for the current setting, but it is possible to extend to the directional case by quantizing the optical flow vectors.

## III. DYNAMIC BACKGROUND DETECTION

As mentioned previously, there are no existing methods in the literature on dynamic background that can be applied directly to the problem we address in this paper. In the proposed setting, properties of the repetitive motions are not known in advance (in fact the motion statistics of swaying tree branches and their shadows may vary considerably over time as well). Thus, it is not possible to formulate the problem as supervised learning (binary classification) simply because of the unavailability of training data. Thus, a wide range of methods, such as those described in [9] [17], [18], that require annotated training, cannot be used. The only class of existing techniques that can be adapted to the proposed setting uses adaptively learnt statistical models. By comparing the underlying assumption of the dynamic background model and the actual data, a conclusion can be made about whether the data behaves like dynamic background.

In this section, we first present an adaptation of the classic Stauffer and Grimson's MoG background modeling [27] for this detection problem. The original MoG method fits a mixture of Gaussian distributions to the statistics of the observed data. Despite being simple, a recent independent survey [5] demonstrated that this method outperforms all other existing dynamic background methods. Its key advantage lies in the ability to detect and extract the dynamic background component from the overall background model of a scene automatically. This is difficult to achieve with nonparametric approaches, such as the computationally expensive kernel density estimation [25].

We propose a novel detection method that is based on the analysis of background dynamics. This analysis further supports the intuition that dynamic background exhibits strong temporal persistence, which is extracted by using a novel formulation. We then derive an efficient algorithm to solve this formulation in order to perform dynamic background detection.

### A. Mixture of Gaussians

In the seminal work of Stauffer and Grimson on background modeling in the context of tracking [27], appearance is modeled using a mixture of Gaussians. The focus of their work is on background modeling with the aim of robust foreground detection. Dynamic background detection as such was not performed explicitly.

To adapt the method of Stauffer and Grimson to the problem considered in this paper, we propose the following. First, we obtain the parameters of a mixture of Gaussians, including: the number of mixture components  $K$ , the associated means and standard deviations  $\mu_k$  and  $\sigma_k$  and their weights  $w_k$ ,  $k = 1, \dots, K$ . Then, the mixture components are sorted in a decreasing order of  $w/\sigma$  as originally suggested. Next, we separate the background from the foreground by selecting the first  $K_{bg}$  components such that their total energy contribution exceeds a specified threshold  $T$

$$\sum_{k=1}^{K_{bg}} w_k \geq T. \quad (1)$$



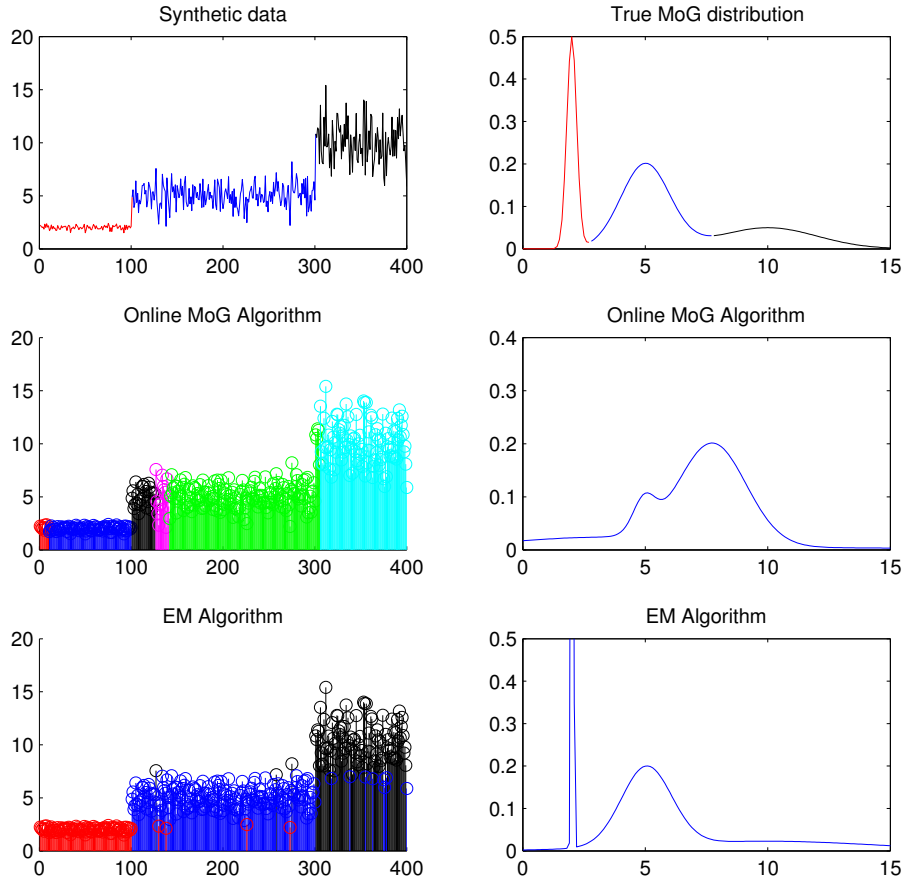


Fig. 1. Synthetic experiment for MoG estimation

Here,  $T = 0.7$  is an average value for background modeling used in previous works [14], [15]. Generally, a higher  $T$  allows a richer multimodal modeling of background that contains repetitive motion [27]. Next, we remove the Gaussian component which corresponds to the static elements of the background as the one with the smallest standard deviation  $\sigma_k \leq \varepsilon$  for some threshold  $\varepsilon$ . In some cases, the background is purely dynamic, and hence a static background component may not exist. The threshold  $\varepsilon$  is data dependent. Once the static background component is removed, the next step is to compute a score for the dynamic background components. The desired criterion is that if the variation is large, the background should appear more dynamic. To achieve this goal, we use negative differential entropy on the remaining background mixture components as the score  $S = -\int p(x) \log(p(x)) dx$ , where

$$p(x) = \sum_{k=1}^{K_{bg}-1} w_k \mathcal{N}(x|\mu_k, \sigma_k). \quad (2)$$

The entropy for mixture of Gaussians does not have an analytical form, and thus it is computed numerically.

For the estimation of the MoG model, an online algorithm was derived to construct the mixture [27]. However, our

experience over intensive numerical studies demonstrate that the online algorithm does not have sufficient accuracy to obtain a reliable MoG model. As our detection problem is formulated as the basis for post-event analysis, we propose to use the expectation-maximization (EM) algorithm to estimate the MoG model in a batch setting. To select the optimal value of  $K$  that avoids over-fitting, we track how the likelihood value varies as  $K$  increases. If the relative decrease in the likelihood as  $K$  is incremented falls below a specific threshold  $\delta$  (in this work we used  $\delta = 0.1$ ) then we stop at that value of  $K$ , which we declare optimal.

Fig. 1 shows a synthetic example that demonstrates the superior performance (both in terms of higher accuracy and consistency) of the proposed algorithm over the online algorithm. In this synthetic example, we generate random samples from a Gaussian mixture with three components  $\mathcal{N}(2, 0.2)$ ,  $\mathcal{N}(5, 1)$ , and  $\mathcal{N}(10, 0.2)$ . Their weights are 0.25, 0.5, and 0.25 respectively. The top right subplot shows the actual mixture distribution, whilst the top left subplot shows 400 random samples we generate. To make it even easier for Stauffer and Grimson's online EM method, we have sorted and ordered the samples according to the mixture component they come from. The middle left and right subplots show

how the online EM algorithm performs. Here, it is observed that it overestimates the number of components (as shown by the number of distinct colors on the middle left subplot) and consequently the estimated distribution is considerably different from the true mixture distribution. On the other hand, the batch EM approach produces a consistent estimate of the mixture order and more accurate mixture distribution (as shown on the bottom subplots).

### B. Local Variation Persistence

1) *Overview of Tree Aerodynamics:* The forestry research literature suggests that the response of an isolated tree under an action of wind flow can be modeled by a second-order differential equation, which depends on both internal factors (mass, damping, and stiffness) and external factors (aerodynamic dragging force and constant gravitational force) [20], [26]. The characteristic movement of a tree under an external impulse largely depends on the tree itself and is described as a damped oscillation. The oscillation comprises a wide range of frequencies. Theoretical predictions and actual measurements show that the mechanical transfer function decays at a slope of about  $-4/3$  [20], with local spectral peaks corresponding to the vibration modes of the trees, also known as eigen-frequencies [20], [26]. Because of this steep slope, the oscillation is dominated by the principal eigen-frequency.

Further information can be deduced from tree aerodynamic research. Obviously, the onset and the magnitude of the swaying motion depend on the dragging force, which is dependent on the airflow characteristics and the actual tree-wind interaction. Wind velocity is likely to have large, intermittent fluctuations, and it is established in [20] that the movements of the upper parts of a tree correlate positively to wind velocity in general. A sudden onset of a large wind flow, known as wind gust, will excite large tree movements, though there could be a lag due to previous motion. Within a longer temporal scope, it is expected that the tree movement will have intermittent driving pulses. The duration between the pulses may vary widely, but it is generally much larger than the fundamental oscillation of the tree mentioned above. Between large driving pulses, the swaying motion is a damped natural oscillation trending to rest in the absence of wind.

2) *Proposed Method:* Inspired by the previous analysis, it is reasonable to assume that the tree swaying motion at a *particular location* is dominated by its characteristic motion over a short time scale during which the damping does not change significantly. In other words, the normalized short-time motion can be approximated by the transfer function in the frequency domain and it is purely dependent on the tree characteristics.

The tree aerodynamic theory above governs the behavior of tree segment displacements. To make use of the theory when only motion counts are available, we need to assume that the mapping from the tree movement to the feature domain is locally monotonic. In other words, for a small area of interest, the observed motion counts have large values if the actual tree movement is large and vice versa. In practice, this may only hold approximately as the actual mapping may vary between

locations. Thus, the short-term spectral structure of the motion counts is likely to be temporally persistent for the whole swaying duration. Depending on how the mapping function converts an actual movement to the feature domain, the spectral pattern of the motion counts may be different to the spectral structure of the actual characteristic tree movement. Nevertheless, the crucial aspect of our argument is that persistence in short-term spectral structure is the key signature of the motion counts for a fixed mapping function at a particular location.

To test our argument empirically, we analyze the motion counts of a cell of the shadow of a swaying tree from CAM2 dataset (see Section IV for details) as shown in Fig. 2. The top subplot shows values of the motion counts over the entire event, which is entirely dominated by the swaying shadow. Using the 32-point short-time Fourier transform, we analyze the instant spectrogram of the motion data in the second image subplot. Here, it shows a rich frequency content. The large values at the bottom of the plot correspond to low frequencies, while the lighter color of higher frequencies implies a smaller value. Each column in this subplot corresponds to a temporal sliding analysis window. The lowest frequency corresponds to the average power of the motion signal over the window which may vary due to the varying power of the wind at different times. To account for this variation, we normalize the spectrum by the lowest frequency and plot this normalized version from all windows in the third subplot of Fig. 2. Interestingly, we observe that there is an overall stable structure in the normalized spectrum over all windows, with very small variations. This is consistent with the arguments above. Finally, we plot the sum of the normalized short-time spectrum in the bottom subplot of Fig. 2. This serves as an overall summary of the spectrum. Clearly, over the entire event, this summary statistic varies little, mainly between 1 and 2. This suggests that a persistence measure over the summary statistic would lead to a powerful detector. For completeness, we also analyze a loitering event and show the results in Fig. 3. Clearly, this loitering event exhibits large variation in the normalized short-time spectrum, and thus the summary statistic does not appear to be strongly persistent.

Based on the above arguments, we now detail a detection method, which we call local variation persistence (LVP). Note that we detect at the cell level and do not exploit any spatial correlation between cells. This is because, unlike pixel-based approaches, the entire dynamic background might be confined to a cell. Thus, it is reasonable to ignore correlation between cells to simplify the computation.

Denote as  $x_i, i = 1, 2, \dots$ , the motion counts of the event from one cell. We first compute time-series features of the motion counts signals by using sliding windows with a lag  $\Delta$  and a length  $L$ . The data points from the  $j$ -th window are  $W_j = \{x_{(j-1)\Delta+1}, \dots, x_{(j-1)\Delta+L-1}\}$ . Denote as  $y_j = F(W_j, L)$  a feature extraction for the window  $W_j$  at length  $L$ . Here, the feature extractor  $F$  computes an augmented normalized spectrum of the window  $W_j$ . Suppose that  $[S_1, S_2, \dots, S_n]$  is the spectrum at normalized frequencies  $f_1, f_2, \dots, f_n$ , where  $f_1$  corresponds to the lowest frequency.

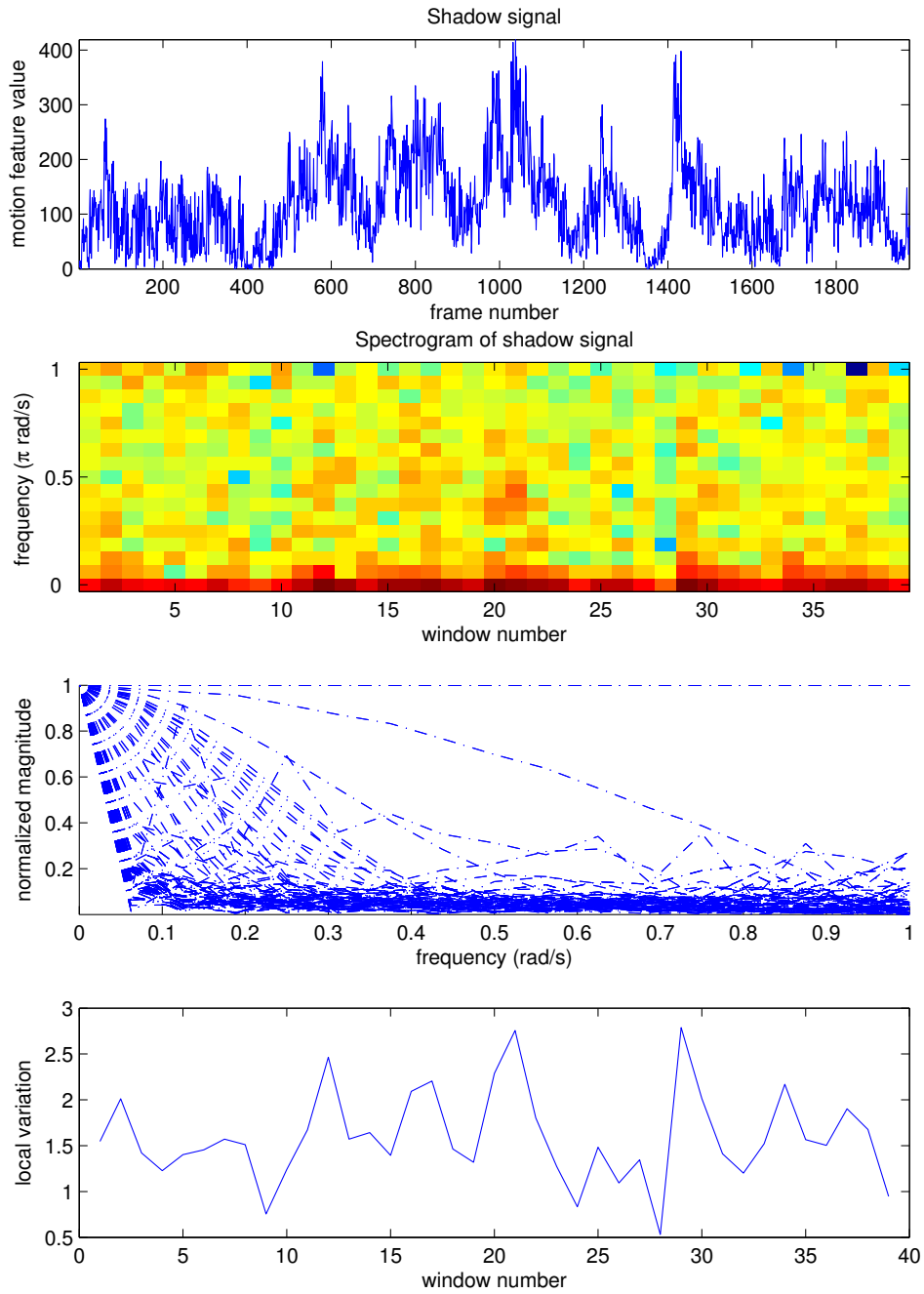


Fig. 2. Local variation analysis of motion counts of a cell observing a purely swaying movement from a shadow of a tree branch.

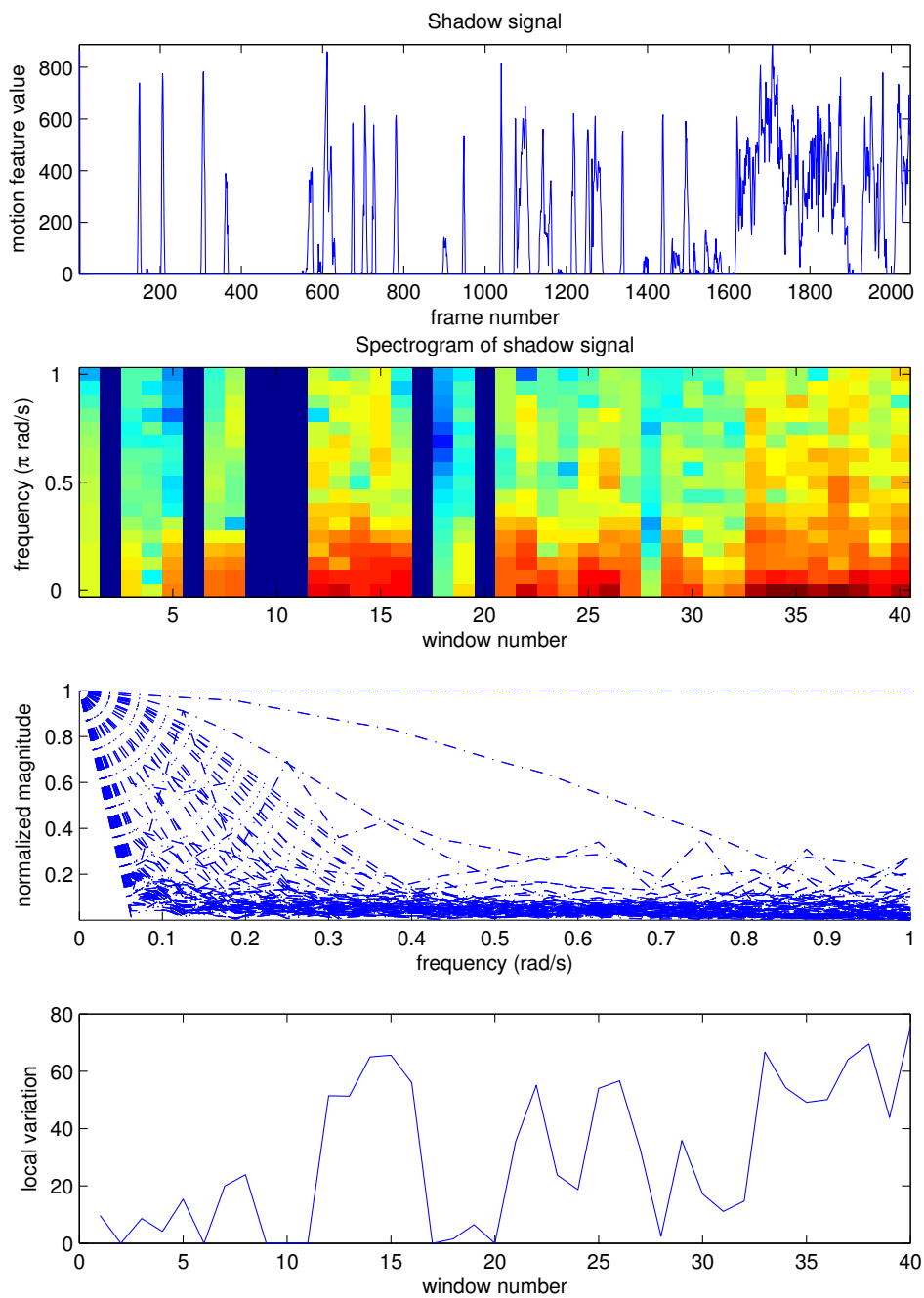


Fig. 3. Local variation analysis of motion counts of a cell observing a loitering movement.



Then, the feature vector for window  $W_j$  is

$$\mathbf{y}_j = \left[ \frac{S_2}{S_1 + \epsilon}, \dots, \frac{S_n}{S_1 + \epsilon} \right]^T. \quad (3)$$

Here,  $\epsilon$  is a small number introduced in order to avoid numerical instability. Then, we collect all feature vectors over the temporal dimension to get a time-series feature matrix

$$\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_n]. \quad (4)$$

The motivation behind the method is to capture the temporal variation as illustrated in the bottom subplots of Figs. 2 and 3 from this feature matrix. In the ideal case of a constant swaying motion, the normalized response (as shown in the third subplot of Figs. 2 and 3) is a constant vector across the window and so is the temporal variation. In practice, there could be slight variations between frames in terms of both the normalized response and temporal variation. In addition, we also need to account for both dense and sparse noise. To describe this, we consider the following rank-1 approximation modeling for the time-series feature matrix

$$\mathbf{Y} = \sigma \mathbf{u} \mathbf{v}^T + \mathbf{E} + \mathbf{N}. \quad (5)$$

Here,  $\mathbf{L} = \sigma \mathbf{u} \mathbf{v}^T$  is the rank-1 approximation of the intrinsic dynamic background data. In this modeling,  $\mathbf{u}$  represents the spectral structure that is common over the entire event,  $\mathbf{v} \succcurlyeq \mathbf{0}$  is a vector which denotes the temporal variation, and  $\sigma > 0$  is a normalizing constant to reduce ambiguity together with the unitary constraints  $\|\mathbf{u}\|_2 = 1$ ,  $\|\mathbf{v}\|_2 = 1$ .  $\mathbf{E}$  and  $\mathbf{N}$  respectively represent the outliers (sparse) and noise (dense) in the feature due to extraction or modeling.

Our modeling is inspired by a recent seminal work known as robust principal component analysis (RPCA) [7], in particular its stable version [30]. It has seen related applications in background subtraction for video surveillance [3]. In the image processing context, extensions of RPCA to exploit the spatio-temporal constraints, such as local sparseness [28] and foreground local coherence [13], have been observed. Note that RPCA-based variants operating at the pixel level in video surveillance are far from being practical due to their high computational cost [7]. In contrast, our proposed method operates on motion counts and thus it is more practically feasible. The fundamental difference is that RPCA seeks a general ‘‘low-rank’’ structure, while we seek a specific rank-1 structure. Under the modeling described by (5) and in the absence of noise and corruption, the **intrinsic** background  $\mathbf{L} = \sigma \mathbf{u} \mathbf{v}^T$  is a rank-1 matrix. Thus, the rank-1 constraint imposed by the proposed method is more suitable for the problem. This will also lead to major algorithmic differences which we will demonstrate subsequently.

We propose to extract the temporal variation  $\mathbf{v}$  using the following formulation

$$\begin{aligned} \arg \min_{\sigma \geq 0, \mathbf{u}, \mathbf{v}, \mathbf{E}} \quad & \|\mathbf{Y} - \sigma \mathbf{u} \mathbf{v}^T - \mathbf{E}\|_F^2 + \lambda \|\mathbf{E}\|_1 \quad (6) \\ \text{s.t.} \quad & \mathbf{u}^T \mathbf{u} = 1 \\ & \mathbf{v}^T \mathbf{v} = 1. \end{aligned}$$

This problem is very challenging as it involves a nonlinear regularization term and unitary constraints. To solve it, we follow

the framework of alternative direction method of multipliers (ADMM) [4], which has recently been found to be powerful in many practical applications [23]. To convert the problem to a solvable form, we rewrite the formulation by introducing auxiliary variables

$$\begin{aligned} \arg \min_{\sigma \geq 0, \mathbf{u}, \mathbf{v}, \mathbf{E}} \quad & \|\mathbf{Y} - \sigma \mathbf{u} \mathbf{v}^T - \mathbf{E}\|_F^2 + \lambda \|\mathbf{E}\|_1 \quad (7) \\ \text{s.t.} \quad & \mathbf{u}^T \mathbf{p} = 1, \mathbf{p} = \mathbf{u} \\ & \mathbf{v}^T \mathbf{q} = 1, \mathbf{q} = \mathbf{v}. \end{aligned}$$

Under ADMM, the augmented Lagrangian in the scaled form can be written as (see [4, Sec. 3.1.1])

$$\begin{aligned} \mathcal{L} = \quad & \|\mathbf{Y} - \sigma \mathbf{u} \mathbf{v}^T - \mathbf{E}\|_F^2 + \lambda \|\mathbf{E}\|_1 \\ & + \eta_1 (\mathbf{u}^T \mathbf{p} - 1 + a_1)^2 + \eta_2 \|\mathbf{p} - \mathbf{u} + \mathbf{a}_2\|_2^2 \\ & + \eta_3 (\mathbf{v}^T \mathbf{q} - 1 + a_3)^2 + \eta_4 \|\mathbf{q} - \mathbf{v} + \mathbf{a}_4\|_2^2. \quad (8) \end{aligned}$$

Here,  $a_1, \mathbf{a}_2, a_3, \mathbf{a}_4$  are the scaled dual variables of the Lagrangian and  $\eta_1, \eta_2, \eta_3, \eta_4$  are the augmented Lagrangian parameters. According to ADMM theory, the augmented Lagrangian is minimized by iteratively solving for one variable at the time while fixing the others. Denote as  $k$  the current iteration number. First, the step for updating  $\sigma$  is found from solving

$$\sigma^{k+1} = \arg \min_{\sigma \geq 0} \|\mathbf{Y} - \sigma \mathbf{L}^k - \mathbf{E}^k\|_F^2, \quad \mathbf{L}^k = \mathbf{u}^k (\mathbf{v}^k)^T, \quad (9)$$

which yields

$$\sigma^{k+1} = \frac{\text{tr}((\mathbf{Y} - \mathbf{E}^k)^T \mathbf{L}^k)}{\|\mathbf{L}^k\|_F^2}. \quad (10)$$

Next, we derive the update step for  $\mathbf{u}$  by minimizing the augmented Lagrangian with respect to  $\mathbf{u}$

$$\begin{aligned} \mathbf{u}^{k+1} = \quad & \arg \min_{\mathbf{u}} \|\mathbf{Y} - \mathbf{E}^k - \sigma^{k+1} \mathbf{u} (\mathbf{v}^k)^T\|_F^2 \\ & + \eta_1 (\mathbf{u}^T \mathbf{p}^k - 1 + a_1^k)^2 \\ & + \eta_2 \|\mathbf{p}^k - \mathbf{u} + \mathbf{a}_2^k\|_2^2. \quad (11) \end{aligned}$$

Straightforward algebra leads to

$$\begin{aligned} \mathbf{u}^{k+1} = \quad & \mathbf{Q}_u^{-1} (\sigma^{k+1} (\mathbf{Y} - \mathbf{E}^k) \mathbf{v}^k + \eta_1 \mathbf{p}^k (1 - a_1^k) \\ & + \eta_2 (\mathbf{p}^k + \mathbf{a}_2^k)), \quad (12) \end{aligned}$$

where  $\mathbf{Q}_u = ((\sigma^{k+1})^2 \|\mathbf{v}^k\|_2^2 + 1) \mathbf{I} + \eta_1 \mathbf{p}^k (\mathbf{p}^k)^T$ . It is noted that due to the matrix inversion lemma, the inversion of  $\mathbf{Q}$  can be computed efficiently as follows

$$\mathbf{Q}_u^{-1} = \rho^{-1} \mathbf{I} - \gamma \mathbf{p}^k (\mathbf{p}^k)^T, \quad (13)$$

here  $\rho = (\sigma^{k+1})^2 \|\mathbf{v}^k\|_2^2 + 1$  and  $\gamma = \rho^{-2} (\eta_1^{-1} + \rho^{-1} (\mathbf{p}^k)^T \mathbf{p}^k)^{-1}$ .

Next, we derive the update for  $\mathbf{p}$  by solving the following problem

$$\begin{aligned} \mathbf{p}^{k+1} = \quad & \arg \min_{\mathbf{p}} \eta_1 ((\mathbf{u}^{k+1})^T \mathbf{p} - 1 + a_1^k)^2 \\ & + \eta_2 \|\mathbf{p} - \mathbf{u}^{k+1} + \mathbf{a}_2^k\|_2^2, \quad (14) \end{aligned}$$

which is

$$\mathbf{p}^{k+1} = \mathbf{Q}_p^{-1} (\eta_1 (1 - a_1^k) \mathbf{u}^{k+1} + \eta_2 (\mathbf{u}^{k+1} - \mathbf{a}_2^k)), \quad (15)$$

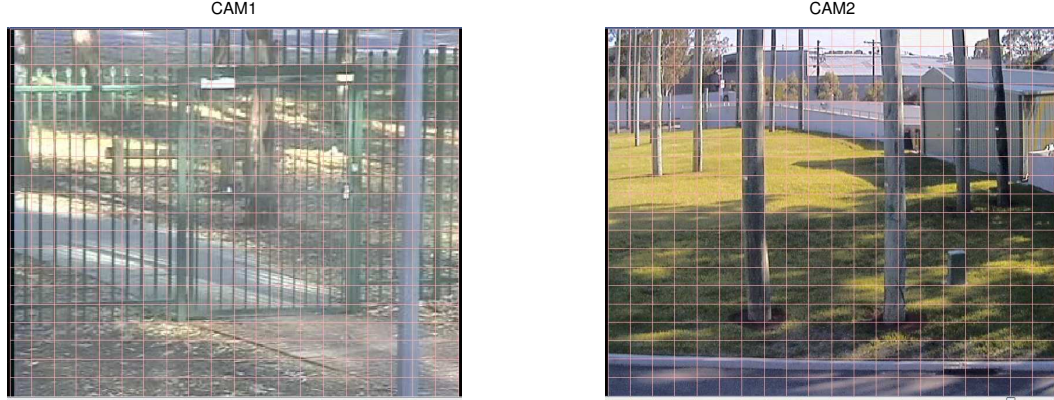


Fig. 4. Screen-shots of the two cameras in the datasets. CAM1 (left) monitors a walk path to an organization, which is surrounded by large trees. CAM2 (right) monitors an asset site.

where  $\mathbf{Q}_p = \eta_1 \mathbf{u}^{k+1} (\mathbf{u}^{k+1})^T + \eta_2 \mathbf{I}$  and it can be efficiently inverted via the inversion lemma.

The update steps for  $\mathbf{v}$  and  $\mathbf{q}$  are also derived similarly,

$$\begin{aligned} \mathbf{v}^{k+1} &= \mathbf{Q}_v^{-1} (\sigma^{k+1} (\mathbf{Y} - \mathbf{E}^k)^T \mathbf{u}^{k+1} + \eta_3 \mathbf{q}^k (1 - a_3^k) \\ &\quad + \eta_4 (\mathbf{q}^k + \mathbf{a}_4^k)), \\ \mathbf{q}^{k+1} &= \mathbf{Q}_q^{-1} (\eta_3 (1 - a_3^k) \mathbf{v}^{k+1} + \eta_4 (\mathbf{v}^{k+1} - \mathbf{a}_4^k)), \end{aligned} \quad (16)$$

where

$$\begin{aligned} \mathbf{Q}_v &= ((\sigma^{k+1})^2 \|\mathbf{u}^k\|_2^2 + \eta_4) \mathbf{I} + \eta_3 \mathbf{q}^k (\mathbf{q}^k)^T \\ \mathbf{Q}_q &= \eta_3 \mathbf{v}^{k+1} (\mathbf{v}^{k+1})^T + \eta_4 \mathbf{I}. \end{aligned} \quad (17)$$

Next, the update step for  $\mathbf{E}$  is

$$\mathbf{E}^k = \arg \min_{\mathbf{E}} \|\mathbf{Y} - \sigma^{k+1} \mathbf{L}^{k+1} - \mathbf{E}\|_F^2 + \lambda \|\mathbf{E}\|_1, \quad (18)$$

which gives

$$E_{ij} = \mathcal{S}_{\lambda/2}(Y_{ij} - \sigma^{k+1} L_{ij}^{k+1}), \quad (19)$$

where  $\mathcal{S}_\tau(x) = \max(0, |x| - \tau) \text{sign}(x)$  is the soft-thresholding shrinkage operator.

Finally, all the dual variables are updated as follows

$$a_1^{k+1} = a_1^k + (\mathbf{u}^{k+1})^T \mathbf{p}^{k+1} - 1, \quad (20)$$

$$\mathbf{a}_2^{k+1} = \mathbf{a}_2^k + \mathbf{p}^{k+1} - \mathbf{u}^{k+1}, \quad (21)$$

$$a_3^{k+1} = a_3^k + (\mathbf{v}^{k+1})^T \mathbf{q}^{k+1} - 1, \quad (22)$$

$$\mathbf{a}_4^{k+1} = \mathbf{a}_4^k + \mathbf{q}^{k+1} - \mathbf{v}^{k+1}. \quad (23)$$

The iteration is terminated when the following primal and dual residuals are smaller than specified thresholds (for more detail, see [4, Sec. 3.2])

$$r_1^k = (\mathbf{u}^k)^T \mathbf{p}^k - 1, \quad (24)$$

$$\mathbf{r}_2 = \mathbf{p}^k - \mathbf{u}^k, \quad (25)$$

$$r_3^k = (\mathbf{v}^k)^T \mathbf{q}^k - 1, \quad (26)$$

$$\mathbf{r}_4^k = \mathbf{q}^k - \mathbf{v}^k, \quad (27)$$

$$\mathbf{s}_1^k = \eta_2 (\mathbf{p}^k - \mathbf{p}^{k-1}), \quad (28)$$

$$\mathbf{s}_2^k = \eta_2 (\mathbf{q}^k - \mathbf{q}^{k-1}). \quad (29)$$

Once the temporal variation  $\mathbf{v}$  is found, we compute a dynamic background score. As we aim to seek persistence over the temporal dimension, this score is proposed as follows

$$S_{\text{LVP}} = \frac{\|\mathbf{v}\|_1}{\sqrt{M}}. \quad (30)$$

Here, the  $\ell_1$  norm is defined as the sum of absolutes  $\|\mathbf{v}\|_1 = \sum_i |v_i|$  and  $M$  is the dimension of  $\mathbf{v}$ . A dynamic background event is raised when the decision statistic is above a specified threshold  $S_{\text{LVP}} \geq S_{th}$ .

We briefly explain the intuition behind this dynamic background statistic. At first, it may appear counter-intuitive as the  $\ell_1$ -norm is often used in the compressed sensing (CS) literature [10] [22] as a sparsity promoting norm. However, CS seeks a minimization of the  $\ell_1$  norm to find sparse patterns. Here, we seek a maximization of the  $\ell_1$  norm to promote *dense* patterns, which are favoured by temporal persistence. In other words, a larger value of the  $\ell_1$  norm indicates a more dynamic background-like event. It is also important to note that as  $\mathbf{v}$  is a singular vector, its  $\ell_2$  norm is  $\|\mathbf{v}\|_2 = 1$ . Thus, it can be shown that  $1 \leq \|\mathbf{v}\|_1 \leq \sqrt{M}$ , where the maximization is attained when all entries in  $\mathbf{v}$  have a magnitude of  $\sqrt{1/M}$ . The normalization factor  $\sqrt{M}$  ensures that events of different lengths are comparable under this scheme, so that the decision statistic is within  $(0, 1]$ , which is scale-invariant to the choice of feature units. Due to the normalization factor, for two signals with same non-zeros entries in  $\mathbf{v}$ , the one with fewer ‘silent’ windows will have a higher score, and thus is more likely to contain dynamic background. In practice, this is useful in discriminating other types of events such as loitering, where the signatures in the active windows maybe very similar but are often mixed with non-active windows.

## IV. DATA AND EVALUATION

### A. Data

The data used in this work comprises two sets of events from two real cameras ‘‘CAM1’’ and ‘‘CAM2’’ operating at an

average rate of 5fps. Each set is a collection of events detected by the video surveillance analytics system of our industry partner during a particular day when there was a combined effect of strong sun and windy conditions:

- CAM1 (see the left-hand subplot in Fig. 4): this camera monitors a walk path to a university, which is surrounded by large trees. It has a total of 31 events for the selected day. Students typically walk through the path back and forth to the site. Events such as loitering or swinging the gate are considered of security interest and thus they are not to be suppressed.
- CAM2 (see the right-hand subplot in Fig. 4): this camera monitors an asset site. It has a total of 30 events for the selected day. The scene is dominated by a grassy area with electric poles and large trees surrounding the area. Only authorized workers are supposed to be present during the day, and no one is supposed to be there during the night. There are some trees on the highway in the far field of the camera as well.

Examples of the motion counts and typical events are shown in Figures 5 and 6. The events triggered by the system are entirely based on the same motion counts used in this work. Whilst the detailed implementation is not available due to commercial confidentiality, the general detection mechanism is based on an early work on unsupervised anomaly detection [6]. This system is being used in many sites world-wide.

### B. Evaluation

Detection of potential dynamic background areas is evaluated according to the standard detection evaluation framework, which looks at the trade-off between the detection probability and the probability of false alarms. This relation is known as receiver operating characteristic (ROC), and the area under the ROC curve (AUC) is commonly used to condense this information for the sake of an easy comparison of different methods.

## V. EXPERIMENTAL RESULTS

*Detection Performance.* We compare the proposed LVP detection method with an adaptation of the Stauffer and Grimson's MoG, using the dataset described in Section IV. Since there is no previous work directly relevant to the specific setting of the new problem described here, this is the only comparison that we could make. Also note that the adapted MoG uses raw motion counts directly as an input due to the specific setting in this work. To extract time-series features, we set the window length to 1 second and the number of short-time FFT points to 32. We then obtain the dynamic background scores over all cells of all events in the dataset. The threshold is varied to detect the dynamic background in each cell. The relation between the false alarm and detection rates is shown in the ROC plot in Fig. 7. In the region where the false alarm rate is lower than 1%, LVP has a similar detection rate to the MoG-based approach, both only detecting 20% of the cells that are considered to be a part of the dynamic background. However, the advantage of LVP becomes evident when the false alarm rate is above 1%. For example, at the 5% false

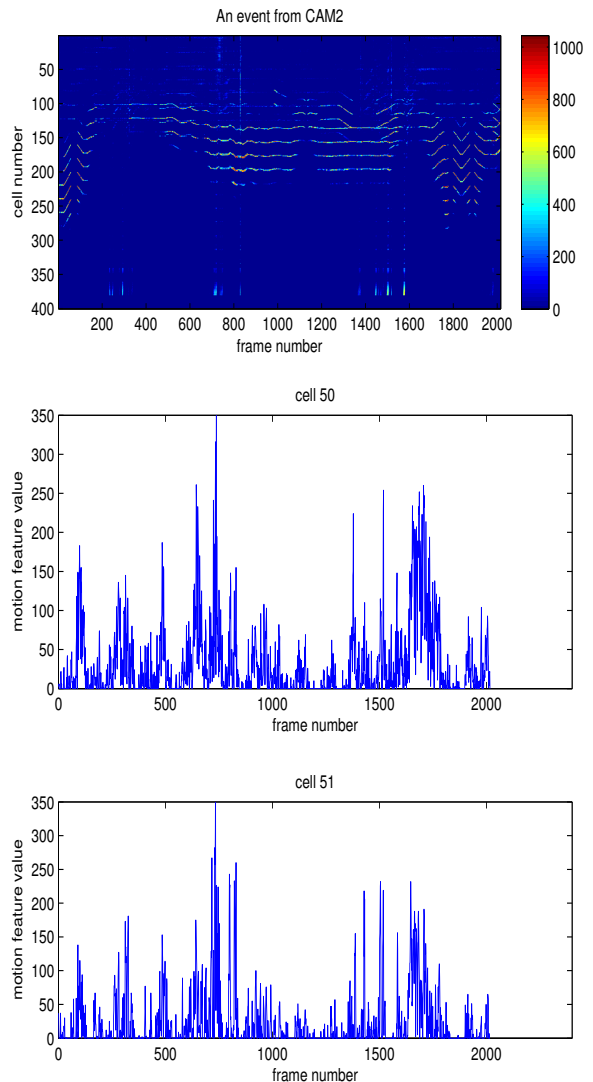


Fig. 5. Example of an event from CAM2. At the top is an image plot of the motion counts for the event. During this event, there was a person walking in the scene, which is considered normal. There were only two cells (50 and 51) that exhibit strong shadow characteristics. The actual time-series values of these two cells are further shown on the bottom pair of subplots. From the top subplot and its heat-map legend, it is observed that the average values of the shadow cells are much smaller than those corresponding to the motion of the person. This highlights the challenge when detecting this particular type of moving shadow. It also suggests that feature spatial intensity is an unreliable cue. The premise of our work is that effective discrimination between the two types of event can be achieved by better exploiting the spatio-temporal characteristics of shadows.

alarm rate, LVP detects 70% of the dynamic background cells correctly. Overall, LVP achieves the ROC AUC value of 0.95, considerably above that of MoG with an AUC score of only 0.76.

*Computational Complexity.* The MoG method is underlain by the EM algorithm which is used to compute the mixture components, while the main computational engine of the proposed LVP method is the ADMM algorithm. As both the EM and the ADMM algorithms are iterative in nature, it is not possible to compare their performance in a fully controlled fashion. In an attempt to make a fair comparison, for both algorithms we set the same termination criterion and measure the actual running time measured for each event in the dataset. Specifically, we terminated iterations when the relative change in the solution falls below  $10^{-3}$ , which we found to be a

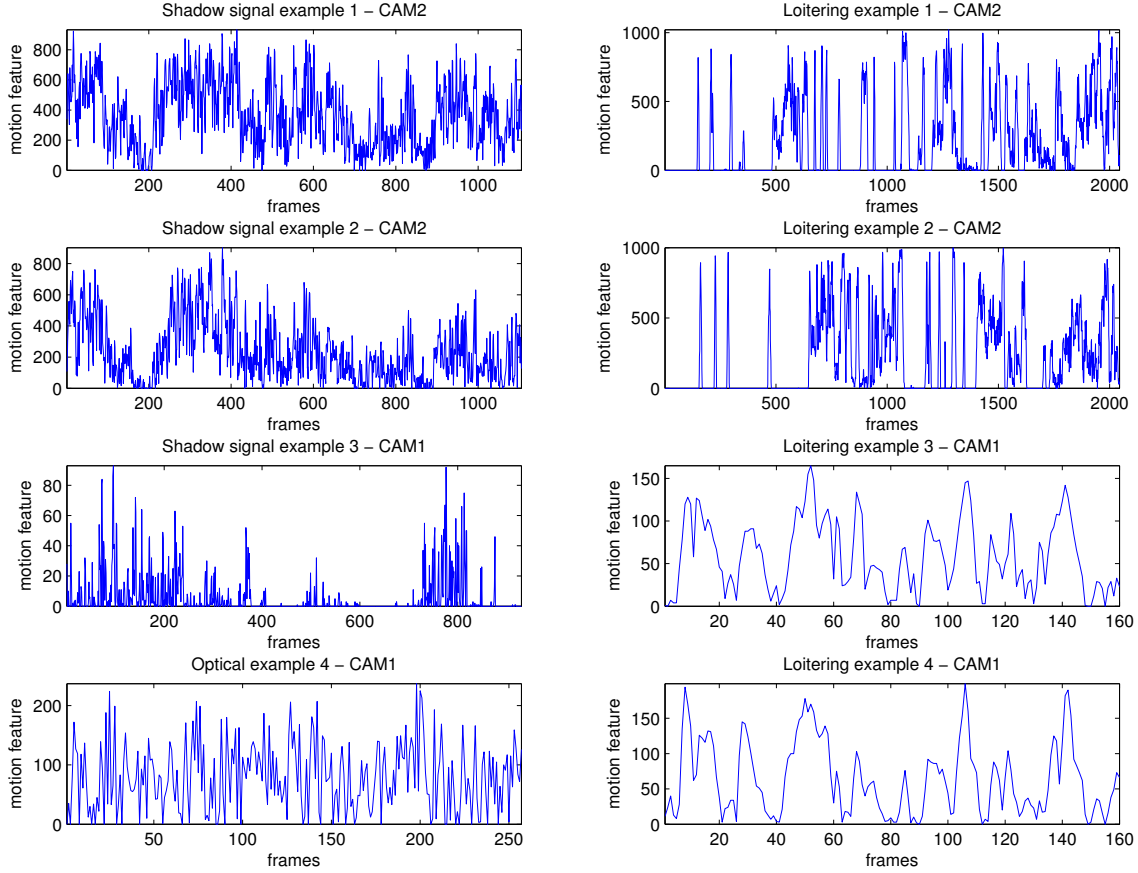


Fig. 6. Example of the motion data. The four subplots on the left-hand side are examples of the dynamic background to be detected. Here, the signals do vary between cameras and the types of dynamic background. The top 3 subplots correspond to purely strong shadow events, whereas the bottom plot corresponds to an optical distortion on the camera lens due to water condensation which leads to repetitive motion in the video frames. These events are not of security interest and it is desirable to suppress them. The four subplots on the right-hand side are examples of repetitive motions from loitering events, which are of security interest. The motion characteristics of loitering also seems to vary between cameras.

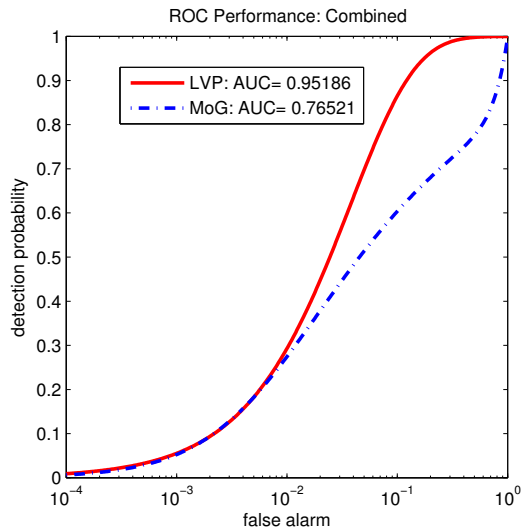


Fig. 7. ROC performance

good trade-off between accuracy and complexity for both algorithms. Fig. 9 shows the measured computation time (in seconds) taken by LVP and MoG. Overall, the LVP method is considerably faster than the MoG based approach. On average, LVP takes only a few seconds to process an event, which makes it readily deployable within the framework of existing surveillance systems.

*Threshold Sensitivity.* Here, we examine the dependence of the false alarm rate on the value of the detection threshold. Using the same parameter values as in the previous experiments, we measured the sensitivity by plotting the required thresholds for LVP needed to achieve different false alarm rates. The threshold false alarm rate curve is shown in Fig. 8. Observe that the threshold varies little over the range where the false alarm rate is lower than 5%. This implies that the proposed method is quite sensitive to threshold setting in this region. If an operating point is to be specified in this region, a careful choice is needed. Thus, this is an important issue for future work.

*Temporal Dynamics.* In the previous experiments, we set



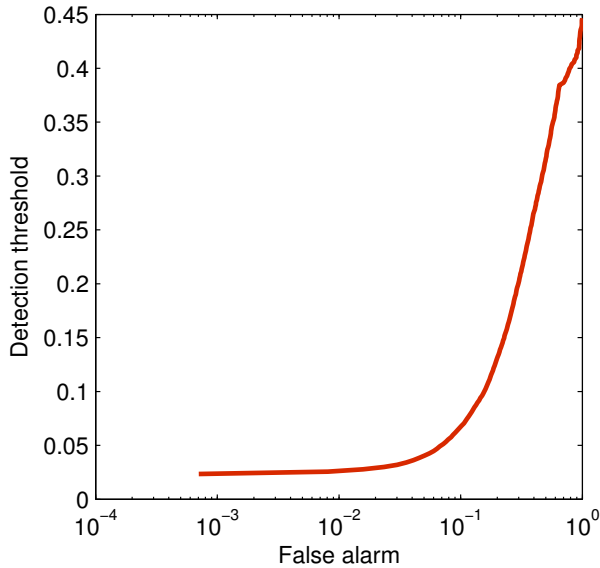


Fig. 8. Threshold sensitivity

the length of the temporal window to 1 second, which is a time scale at which it is reasonable to assume that the observed motion is dominated by the corresponding characteristic frequency. We now examine in further detail the effect of varying the window length. From theory, it is expected that as the time window increases in duration, a greater amount of damping and more external factors affect the observed motion. We varied the window length from 1 second to 20 seconds and examined the corresponding AUC. The results are summarized in Fig. 10. As the plot shows, the detection performance worsens as the window length is increased, with an approximately linear decrease in the AUC. Therefore it appears advantageous to keep the analysis window short. We also observe that shortening the window length beyond 1 second does not improve the detection results any further and thus we omit plotting the ROC curves for these cases for the sake of better clarity.

*Discussion.* The results in the above studies suggest that the proposed method is promising in reducing false alarms in motion-based surveillance analytics. Note that how detection results by the proposed method be used for final decision in a commercial surveillance analytics is not discussed in this work, as it obviously depends on the actual implementation in the subsequent alarm reporting stage, which resolves results from multiple detectors and other high-level semantic rules. The results also suggest that future work should focus on the low false alarm region where the proposed method currently lacks a clear gain over MoG. This could be due to some loitering movements being difficult to be distinguished from swaying motions. One possible direction, which is suggested by an anonymous reviewer of this work, is to extend local variation persistence with an analysis of the high-frequency contents.

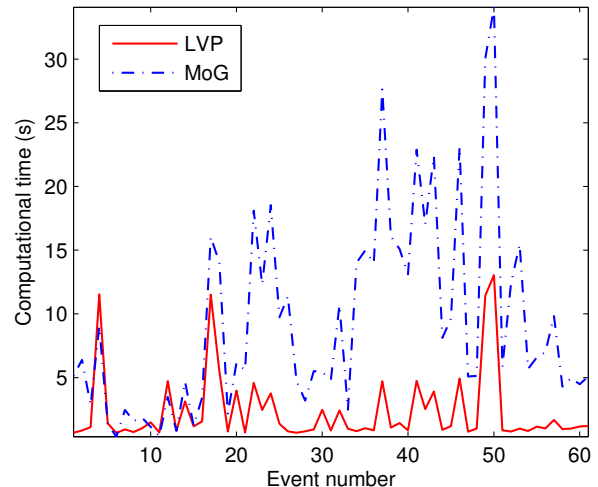


Fig. 9. Computational time

## VI. CONCLUSION

In this paper, we introduced a novel method that addresses the challenge posed by background changes to surveillance systems aimed at detecting salient foreground events. It was argued that in many practical surveillance applications, it is more effective and efficient to detect background motion at the back-end of the system pipeline, rather than at the front-end using the traditional background-foreground separation approach, which is complex and unreliable. Based on this key idea, we formulated a new problem paradigm, derived an adaptation of the leading existing method to this setting, and proposed a fundamentally novel approach based on a theoretical analysis of wind-induced oscillations and specifically the persistence of local variation which we derive as the cornerstone discriminative factor. We made a further contribution by showing how the ADMM algorithm can be used to solve an optimization problem with unitary constraints. Lastly, in this work we also introduced a new evaluation dataset, acquired in the real world using an operational commercial surveillance system. Using this data set we conducted a comprehensive experimental evaluation and demonstrated that the proposed method outperforms the current state-of-the-art on a number of different performance measures.

## REFERENCES

- [1] M. Alterman, Y. Y. Schechner, P. Perona, and J. Shamir, "Detecting motion through dynamic refraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 245–251, 2013.
- [2] T. Bouwmans, "Subspace learning for background modeling: A survey," *Recent Patent On Computer Science*, vol. 2, no. 3, pp. 223–234, 2009.
- [3] T. Bouwmans and E. H. Zahzah, "Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance," *Computer Vision and Image Understanding*, vol. 122, pp. 22–34, 2014.
- [4] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [5] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. CVPR*. IEEE, 2011, pp. 1937–1944.



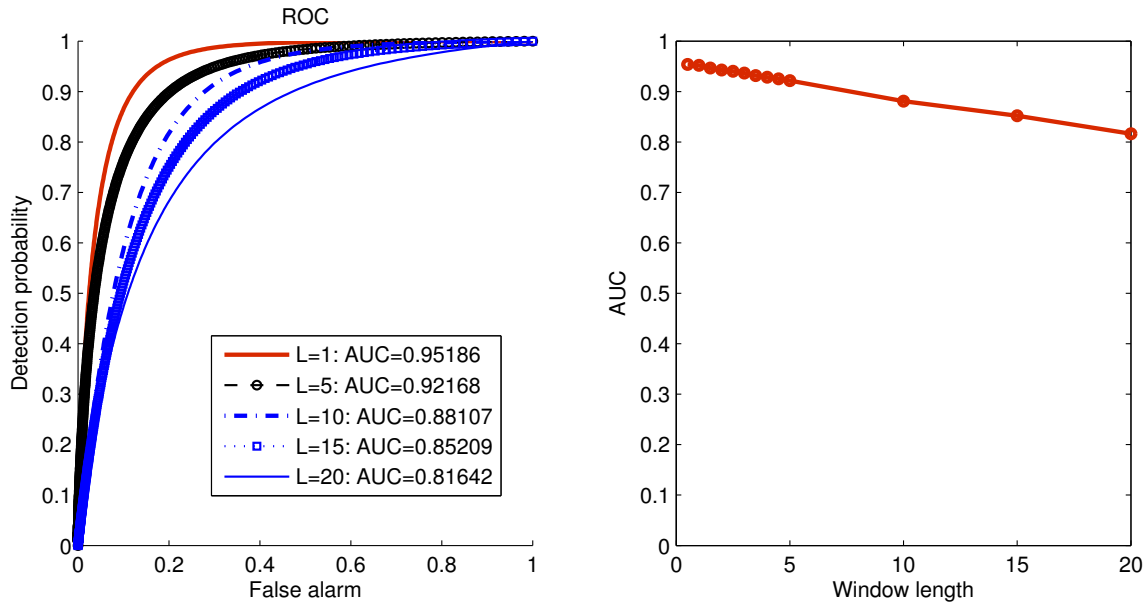


Fig. 10. Temporal dynamics

- [6] S. Budhaditya, D.-S. Pham, M. Lazarescu, and S. Venkatesh, "Effective anomaly detection in sensor networks data streams," in *Proc. IEEE International Conference on Data Mining*. IEEE, 2009, pp. 722–727.
- [7] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.
- [8] A. B. Chan, V. Mahadevan, and N. Vasconcelos, "Generalized Stauffer-Grimson background subtraction for dynamic scenes," *Machine Vision and Applications*, vol. 22, no. 5, pp. 751–766, 2011.
- [9] L. Cheng and M. Gong, "Realtime background subtraction from dynamic scenes," in *Proc. CVPR*. IEEE, 2009, pp. 2066–2073.
- [10] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [11] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proc. IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.
- [12] H.-L. Eng, J. Wang, A. S. Wah, and W.-Y. Yau, "Robust human detection within a highly dynamic aquatic environment in real time," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1583–1600, 2006.
- [13] C. Guyon, T. Bouwmans, and E.-H. Zahzah, "Foreground detection via robust low rank matrix decomposition including spatio-temporal constraint," in *Proc. 1st ACCV Workshop on Background Model Challenges*. Springer, 2013, pp. 315–320.
- [14] M. Heikkilä and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 657–662, 2006.
- [15] M. Heikkilä, M. Pietikainen, and J. Heikkilä, "A texture-based method for detecting moving objects," in *Proc. BMVC*, 2004.
- [16] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1459–1472, 2004.
- [17] V. Mahadevan and N. Vasconcelos, "Background subtraction in highly dynamic scenes," in *Proc. CVPR*. IEEE, 2008, pp. 1–6.
- [18] —, "Spatiotemporal saliency in dynamic scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 171–177, 2010.
- [19] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in *Proc. CVPR*. IEEE, 2003, pp. 1305–1312.
- [20] J. R. Moore and D. A. Maguire, "Simulating the dynamic behavior of douglas-fir trees under applied loads by the finite element method," *Tree physiology*, vol. 28, no. 1, pp. 75–83, 2008.
- [21] O. Oreifej, X. Li, and M. Shah, "Simultaneous video stabilization and moving object detection in turbulence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 450–462, 2013.
- [22] D. S. Pham and S. Venkatesh, "Improved image recovery from compressed data contaminated with impulsive noise," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 397–405, 2012.
- [23] —, "Efficient algorithms for robust recovery of images from compressed data," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4724–4737, 2013.
- [24] I. Saleemi, L. Hartung, and M. Shah, "Scene understanding by statistical modeling of motion patterns," in *Proc. CVPR*. IEEE, 2010, pp. 2069–2076.
- [25] Y. Sheikh and M. Shah, "Bayesian object detection in dynamic scenes," in *Proc. CVPR*, vol. 1. IEEE, 2005, pp. 74–79.
- [26] H.-C. Spatz, F. Brüchert, and J. Pfisterer, "Multiple resonance damping or how do trees escape dangerously large oscillations?" *American Journal of Botany*, vol. 94, no. 10, pp. 1603–1611, 2007.
- [27] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, 2000.
- [28] B. Wohlberg, R. Chartrand, and J. Theiler, "Local principal component pursuit for nonlinear datasets," in *Proc. ICASSP*. IEEE, 2012, pp. 3925–3928.
- [29] J. Zhang and D. Ma, "Nonlinear prediction for gaussian mixture image models," *IEEE Trans. Image Process.*, vol. 13, no. 6, pp. 836–847, 2004.
- [30] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, "Stable principal component pursuit," in *Proc. 2010 IEEE International Symposium on Information Theory Proceedings (ISIT)*. IEEE, 2010, pp. 1518–1522.
- [31] Q. Zhu, Z. Song, Y. Xie, and L. Wang, "A novel recursive Bayesian learning-based method for the efficient and accurate segmentation of video with dynamic background," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3865–3876, 2012.



on Signal Processing.

**Duc-Son Pham** received the PhD degree from Curtin University of Technology in 2005. He is currently a Senior Research Fellow with the Institute for Multi-sensor Processing and Content Analysis (IMPCA) at Curtin University, Perth, Western Australia. His current research interests include sparse learning theory, large-scale data mining, and convex optimization with applications to computer vision and image processing. He is a Member of the IEEE. He is a recipient of the Young Author Best Paper Award 2010 for a publication in IEEE Transactions



**Ognjen Arandjelović** graduated top of his class from the Department of Engineering Science at the University of Oxford (M.Eng.). In 2007 he was awarded the Ph.D. degree from the University of Cambridge. After spending 4 years as a Fellow of Trinity College Cambridge, he moved to Swansea University as a Lecturer in Visual Computing. Currently he is a Senior Lecturer at Deakin University. His main research interests are computer vision and pattern recognition, and their applications in various fields of science. He is a Fellow of the Cambridge

Overseas Trust and a winner of multiple best research paper awards.



**Svetha Venkatesh** is Alfred Deakin Professor and Director of Centre for Pattern Recognition and Data Analytics (PRaDA) at Deakin University. Venkatesh was elected a Fellow of the International Association of Pattern Recognition in 2004 for contributions to formulation and extraction of semantics in multimedia data. She is a Fellow of the Australian Academy of Technological Sciences and Engineering. She was on the editorial board of IEEE Transactions on Multimedia and was on the board of ACM Transactions on Multimedia (2008-2011). She is a

program member of several international conferences such as ACM Multimedia. Venkatesh has developed frontier technologies in large scale pattern recognition exemplified through 416 publications. She has won over \$13 million in competitive research funding. Venkatesh and her colleagues have filed 3 full patents. One start-up company, spun out of these patents is Virtual Observer and based on the paradigm shifting methods that leverages mobile cameras to deliver wide area surveillance solutions. The technology won the Runner up in both the WA Inventor of the year (Early stage) and Global Security Challenge (Asia-Pacific) in 2007. A recent spin-out company is iCetana and is based on novel methods to find anomalies in video data. iCetana won the Broadband Innovation Award at the prestigious Tech23 in 2010.