

MOTION DETECTION WITH FALSE DISCOVERY RATE CONTROL

J. Mike McHugh[†], Janusz Konrad[†], Venkatesh Saligrama[†], Pierre-Marc Jodoin^{††}, and David Castañón[†]

[†] Boston University, Department of Electrical and Computer Engineering, Boston MA 02215, USA

^{††} Université de Sherbrooke, Département d'informatique, Sherbrooke, Qc, Canada, J1K 2R1

[jmmchugh, jkonrad, srv, dac]@bu.edu, Pierre-Marc.Jodoin@usherbrooke.ca

ABSTRACT

Visual surveillance applications such as object identification, object tracking, and anomaly detection require reliable motion detection as an initial processing step. Such a detection is often accomplished by means of background subtraction which can be as simple as thresholding of intensity difference between movement-free background and current frame. However, more effective background subtraction methods employ probabilistic modeling of the background followed by probability thresholding. In this case, the balance between false positives and false negatives (misses) is controlled by a threshold that needs to be adjusted heuristically depending on object sparsity. In this paper, we propose a different detection method that is based on *false discovery rate* control, a multiple-comparison procedure that applies thresholding in significance-score rather than probability space. The proposed approach allows explicit control of false positives and automatically adapts to object sparsity. The new method offers a qualitative improvement in real scenarios as well as a measurable performance gain over non-adaptive techniques when tested on synthetic sequences.

Index Terms— Motion detection, background subtraction, false discovery rate control.

1. INTRODUCTION

Motion detection entails making a decision at each pixel as to whether it is part of the background (\mathcal{B}) or foreground (\mathcal{F}). This may be accomplished by simple subtraction of background intensities/color (void of moving objects) from the current frame followed by thresholding [1]. Such a simple form of background subtraction does not work well if the background is not truly static (e.g., leaves fluttering in the wind, waves on water surface), and thus methods based on probabilistic background modeling have been developed. Among the most successful ones are those that model the probability density function (PDF) of background pixels using a single Gaussian [2], mixture of Gaussians [3], or non-parametric kernel [4].

Such methods can be interpreted in the context of binary hypothesis testing, expressed as the following likelihood ratio test (LRT):

$$\frac{P_{\mathcal{B}}(I[\mathbf{n}])}{P_{\mathcal{F}}(I[\mathbf{n}])} \underset{\mathcal{F}}{\overset{\mathcal{B}}{\gtrless}} \eta \frac{\pi_{\mathcal{F}}}{\pi_{\mathcal{B}}}, \quad (1)$$

where $P_{\mathcal{B}}(I[\mathbf{n}])$ and $P_{\mathcal{F}}(I[\mathbf{n}])$ are probabilities of observing the intensity value I at spatial location \mathbf{n} given that it is the projection of, respectively, the background scene and a foreground object. $\pi_{\mathcal{B}}$ and $\pi_{\mathcal{F}}$ are prior probabilities of the background and foreground, respectively, biasing the decision. The cost term η permits unequal penalty assigned to the four decision/truth scenarios: \mathcal{B}/\mathcal{B} , \mathcal{B}/\mathcal{F} , \mathcal{F}/\mathcal{F} , and \mathcal{F}/\mathcal{B} . If prior probabilities are fixed and $P_{\mathcal{F}}$ is assumed uniform (no particular knowledge about the foreground), then the LRT simplifies to:

$$P_{\mathcal{B}}(I[\mathbf{n}]) \underset{\mathcal{F}}{\overset{\mathcal{B}}{\gtrless}} \theta, \quad (2)$$

where θ is a constant threshold. Clearly, thresholding occurs in the space of probabilities. Note that although θ is the same for all pixels in a frame, the PDF is, in general, different at each pixel. In order to further improve the robustness of background subtraction, adaptive thresholds [5] have been proposed. A particularly powerful threshold adaptation can be accomplished by means of a prior spatial model (e.g., Markov) applied to \mathcal{B}/\mathcal{F} labels [6].

We propose a new approach to motion detection based on *false discovery rate* control. This approach allows explicit control of false positives and automatically adapts to object sparsity, particularly important in surveillance scenarios of scenes with dynamic complexity.

2. FALSE DISCOVERY RATE CONTROL

A binary hypothesis test results in one of two possible outcomes. This deterministic decision rule maps an observation onto the 2-element decision space, commonly denoted by $\{H_0, H_1\}$ corresponding to the null and positive hypotheses. In the context of background subtraction, this test is being performed at every pixel in an image, thus we have many such binary tests.

The framework of the so-called *multiple comparisons procedure* (MCP) is essentially the same; there are M binary tests each with a unique probability distribution under the null hypothesis. A decision rule maps each of M observations into a decision of either H_0 or H_1 . To facilitate further discussion, we introduce some terminology relevant to MCP. Table 1 illustrates the definitions of random variables which denote *quantities* of trials. That is, each takes an integer value in the range $[0, M]$. In general, U, V, T, and S are unobservable random variables while R is observable (as is the quantity $M - R$, naturally). With these random variables defined, we may now define

Table 1. Definition of random variables in MCP.

	# declared H_0	# declared H_1
# true H_0	U	V
# true H_1	T	S
Total	$M - R$	R

error rates which are of particular importance. The global false positive rate (FPR) and false negative rate (FNR) are defined as follows:

$$FPR = \mathbb{E}\{V\}/M, \quad FNR = \mathbb{E}\{T\}/M$$

where $\mathbb{E}\{\cdot\}$ denotes expectation. Additionally, the total error rate (TER) is simply the sum of these two individual error types ($TER = FPR + FNR$). For our purposes, these quantities suffice to qualify a particular detection strategy and, of course, our general aim is to reduce all error rates. Note, however, that both FPR and FNR are defined as expected values normalized by the total number of comparisons, and thus insensitive to the number of positive declarations. In order to address this, the so-called *false discovery rate* was

proposed in the context of MCP [7]. It is defined as the expected proportion of H_1 declarations which are erroneous:

$$FDR = \mathbb{E}\{V/R\}. \quad (3)$$

When the ground truth is known, the realizations of all quantities in Table 1 are observable. The error rates then can be estimated by averaging such realizations over multiple trials.

In MCP formulations, it is common to perform what is known as *significance testing*; by assigning a significance score (p -score) to each observation, the observations can be classified as either significant or insignificant. The common definition of a p -score, from the Neyman-Pearson statistical testing viewpoint, is *the probability that a less likely outcome than the current observation could occur given a null hypothesis* [8]. Interpreted mathematically, this translates to

$$p = f(x) = \int_{\mathcal{R}_1(x)} P_0(y) dy, \quad \mathcal{R}_1(x) = \{y \mid P_0(y) \leq P_0(x)\}, \quad (4)$$

where $P_0(x)$ is the probability of observing value x for random variable X , given the null hypothesis. Being a deterministic function of the random variable X , the significance score is itself a random variable, denoted \mathcal{P} . The p -score is, in fact, a probability, and therefore it must exist in the interval $[0, 1]$. A crucial property of the significance score defined above is that it is uniformly distributed under the null hypothesis [9]. That is, when $X \sim P_0(x)$, it must follow that $\mathcal{P} \sim \text{Unif}[0, 1]$.

3. PROPOSED METHOD

We now apply the idea of significance testing to the background subtraction problem. Let's return to our original notation where the null hypothesis denotes background detected (\mathcal{B}), and the positive hypothesis denotes a foreground object detected (\mathcal{F}). Note that the definition of the significance score depends only on the PDF of the null hypothesis, $P_{\mathcal{B}}$. This is analogous to the simplified LRT whereby the background PDF is compared to a threshold: $P_{\mathcal{B}} \geq_{\mathcal{F}}^{\mathcal{B}} \theta$. The significance score as defined in (4) has a profound meaning in this context. It is exactly the false alarm probability when the threshold θ is set equal to $P_{\mathcal{B}}(x)$. When $\theta = P_{\mathcal{B}}(x)$, it is clear that \mathcal{R}_1 denotes the region where \mathcal{F} is declared - its complement, the region where \mathcal{B} is declared. The integral (4), therefore, corresponds to:

$$\Pr\{\text{declare } \mathcal{F} \mid \mathcal{B} \text{ true}\} = FPR.$$

Herein lies the power of statistical significance testing. The error rate (FPR) and significance scores are intimately related. In order to control the global FPR to be within some level α , one needs to declare all observations \mathcal{F} for which $p(x) \leq \alpha$.

According to (2), a global threshold θ is applied to every test, regardless of the probability distribution $P_{\mathcal{B}}$. In general, each null hypothesis will have a different PDF. This implies that thresholding the PDFs with a global θ will not have any direct effect on the overall error probability. In MCP terms, this is known as *uncorrected* testing [8]; the error probability is not controlled for each individual test. By testing in the *significance* domain, rather than in the traditional probability domain, tighter control of the error rate is possible.

The transformation between the probability and significance domains is, in general, space variant. This is because the background PDF will have a different shape at each pixel. For an arbitrary PDF, such as a non-parametric density, there is no compact functional form for $p = f(x)$. Absent a closed-form expression, we must revert to calculating the integral in (4) explicitly. For 8-bit grayscale images this requires that we compute *all* 256 values of $P_{\mathcal{B}}$ and approximate the integral *via* the sum of those that fall below the current observation.

An exception to this is when $P_{\mathcal{B}}$ is, for example, distributed normally with mean $B[\mathbf{n}]$ and fixed variance σ^2 throughout the image domain. This single-Gaussian assumption at each pixel is then equivalent to literal background subtraction (it is easy to show that $P_{\mathcal{B}}(I[\mathbf{n}]) \geq_{\mathcal{F}}^{\mathcal{B}} \theta$ is equivalent to $|I[\mathbf{n}] - B[\mathbf{n}]| \geq_{\mathcal{F}}^{\mathcal{B}} \rho$, where ρ is another threshold). In this case, FPR control (thresholding p at α) is equivalent to LRT (2). Note that the background image B is often estimated *via* temporal median or mean.

Let the argument x represent the difference $I[\mathbf{n}] - B[\mathbf{n}]$, i.e., the mean is subtracted from each observation. In the case of the zero-mean Gaussian, the function that defines the significance score, $f(x)$, translates to the area under both tails of the distribution:

$$p = f(x) = 2(1 - \text{CDF}(|x|)) = \text{erfc}\left(x/\sqrt{2\sigma^2}\right) \quad (5)$$

where CDF is the cumulative density function and erfc denotes the complementary error function for the zero-mean Gaussian. Clearly, in this specific case there is a space-invariant relationship between a global threshold, θ , and a global significance, p :

$$\theta = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(\text{erfc}^{-1} p)^2\right), \quad p = \text{erfc}\left(\sqrt{\frac{-1}{2} \ln(2\pi\sigma^2\theta^2)}\right)$$

where erfc^{-1} denotes the inverse complementary error function. In practice, the values of $f(x)$ can be precomputed according to (5) for positive integral values of x and looked up in a list.

Clearly, the significance domain is convenient since controlling the global error rate is straightforward. Recall that controlling the global false positive rate to some level ($FPR = \mathbb{E}\{V\}/M \leq \alpha$), generally requires controlling the individual false alarm thresholds at each trial ($\Pr\{\text{declare } \mathcal{F} \mid \mathcal{B}\} \leq \alpha$). Such types of decision rules can lead to a large number of false detections. The FPR cannot be made arbitrarily small, however, since doing so will increase the false negative rate.

The concept behind the FDR controlling procedure is to control the expected proportion of positive detections that are falsely declared, i.e., $FDR = \mathbb{E}\{V/R\} \leq \alpha$. It is clear that any method which controls the FDR must also control the FPR since $R \leq M$, thus $\mathbb{E}\{V\}/M \leq \mathbb{E}\{V/R\}$. The procedure for controlling the FDR is as follows [7]:

- Assign a p value to each observation.
- Sort all p values in ascending order, p_1, p_2, \dots, p_M
- Find the largest index, i , such that $p_i \leq \frac{i}{M}\alpha$. Call the value corresponding to this index p^* .
- Declare observations p_i significant for all $p_i \leq p^*$.

For the proof of this method, the reader is referred to the original work [7]. A graphical illustration of the power and the adaptive nature of the procedure can be also found in [9].

4. RESULTS

We have applied the above FPR and FDR control to both synthetic and natural video sequences. First, significance scores were determined for each pixel. Then, in case of FPR control, all observations with $p_i \leq \alpha$ were declared as \mathcal{F} . In case of FDR control, p -scores were sorted, the significance threshold, p^* , was determined, and all observations with $p_i \leq p^*$ were declared as \mathcal{F} . In all experiments $\alpha=0.01$ was used.

In a synthetic experiment, three 100-frame sequences were generated, each with the same static background and randomly dispersed foreground (square) objects. White Gaussian noise with fixed variance across the entire image was added in each case. Fig. 1 shows sample results for three different densities of \mathcal{F} (1, 10 and 25 moving objects), whereas Table 2 presents error measurements.

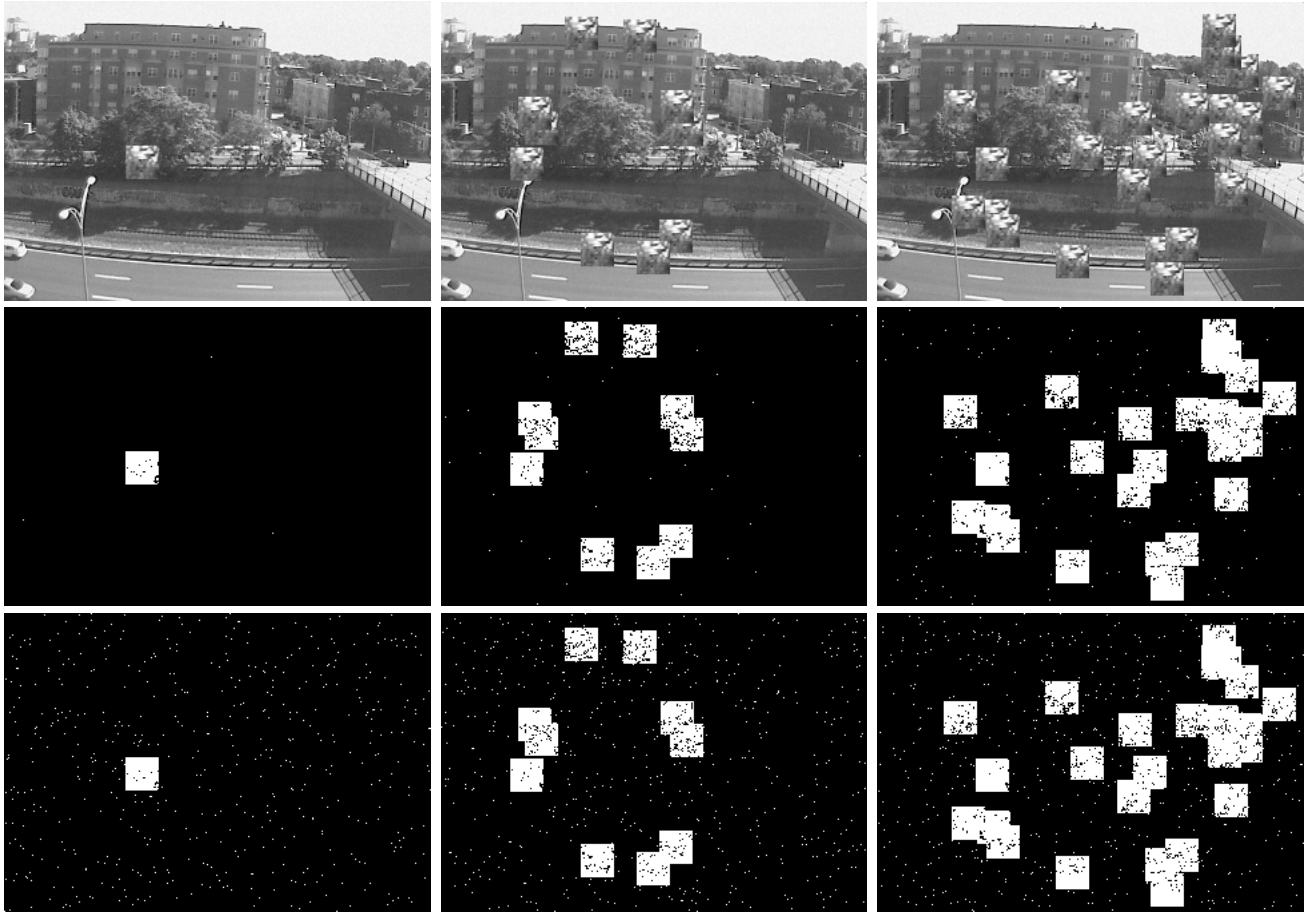


Fig. 1. Detection results for synthetic sequences with low (left), medium (center), and high (right) density of moving objects (\mathcal{F}): original frame (top), FDR control (middle), FPR control (bottom).

Since the background is static, we used a single Gaussian as the PDF model for each background pixel. Clearly, while the false negative rates stay on the same order between FPR and FDR control at each level of object density, the false alarm rate is always lower for the FDR procedure - by as much as two orders of magnitude.

Fig. 2 shows similar results on real sequences. We used a non-parametric background model [4] to determine P_B at each pixel, and for each observation in the current image, we computed p -scores (4). Note that, as expected, the number of false positives is lower when the FDR is controlled instead of the FPR. Also, the FDR-controlled result improves noticeably when the true density of foreground objects is low whereas the FPR-controlled result does not.

Table 2. Detection error rates for synthetic-sequence experiments.

# obj.	controlled FPR		controlled FDR	
	FPR	FNR	FPR	FNR
1	$7.8 \cdot 10^{-3}$	$6.8 \cdot 10^{-4}$	$9.4 \cdot 10^{-5}$	$9.8 \cdot 10^{-4}$
10	$7.3 \cdot 10^{-3}$	$7.0 \cdot 10^{-3}$	$4.7 \cdot 10^{-4}$	$9.1 \cdot 10^{-3}$
25	$6.4 \cdot 10^{-3}$	$1.6 \cdot 10^{-2}$	$1.8 \cdot 10^{-3}$	$1.8 \cdot 10^{-2}$

5. DISCUSSION AND CONCLUSIONS

The presented motion detection with FDR control clearly outperforms FPR control in that it adapts to the density of foreground pixels; the rate of falsely discovered foreground pixels is bounded

and proportional to the number of foreground pixels. This has importance in surveillance of complex scenes, such as airports, train stations, etc., where the number of moving objects changes dramatically in time. While for any particular object density false positives can be controlled by a judicious threshold selection, adjustment of this threshold as a function of object density is non-trivial. This is where the power of FDR lies; we can set a non-arbitrary parameter to achieve a strictly bounded error rate regardless of the complexity of surveyed scene. A more detailed analysis and additional experimental results, e.g., showing threshold dependence on object density, can be found in [9].

Having shown the detection power of the FDR one needs to consider its computational complexity. The transformation from probability to significance domain is simple when the scene is truly static and single Gaussian is sufficient to model the background (analytic transformation (5) applies). When the background PDFs at each pixel are arbitrary, however, transforming between the two domains is not trivial. Then, determining the p -score of a given observation requires explicit evaluation of the integral (4). Once these scores are determined for every pixel, the FDR procedure can be used to determine a significance threshold p^* . Translating this significance threshold back into a probability threshold, θ , is again a non-trivial task. Essentially, one must determine the θ that corresponds to the decision region \mathcal{R}_1 such that

$$p^* > f(\theta) = \int_{\mathcal{R}_1(\theta)} P_0(y) dy.$$

For a non-parametric PDF, this requires a “water-filling” type of

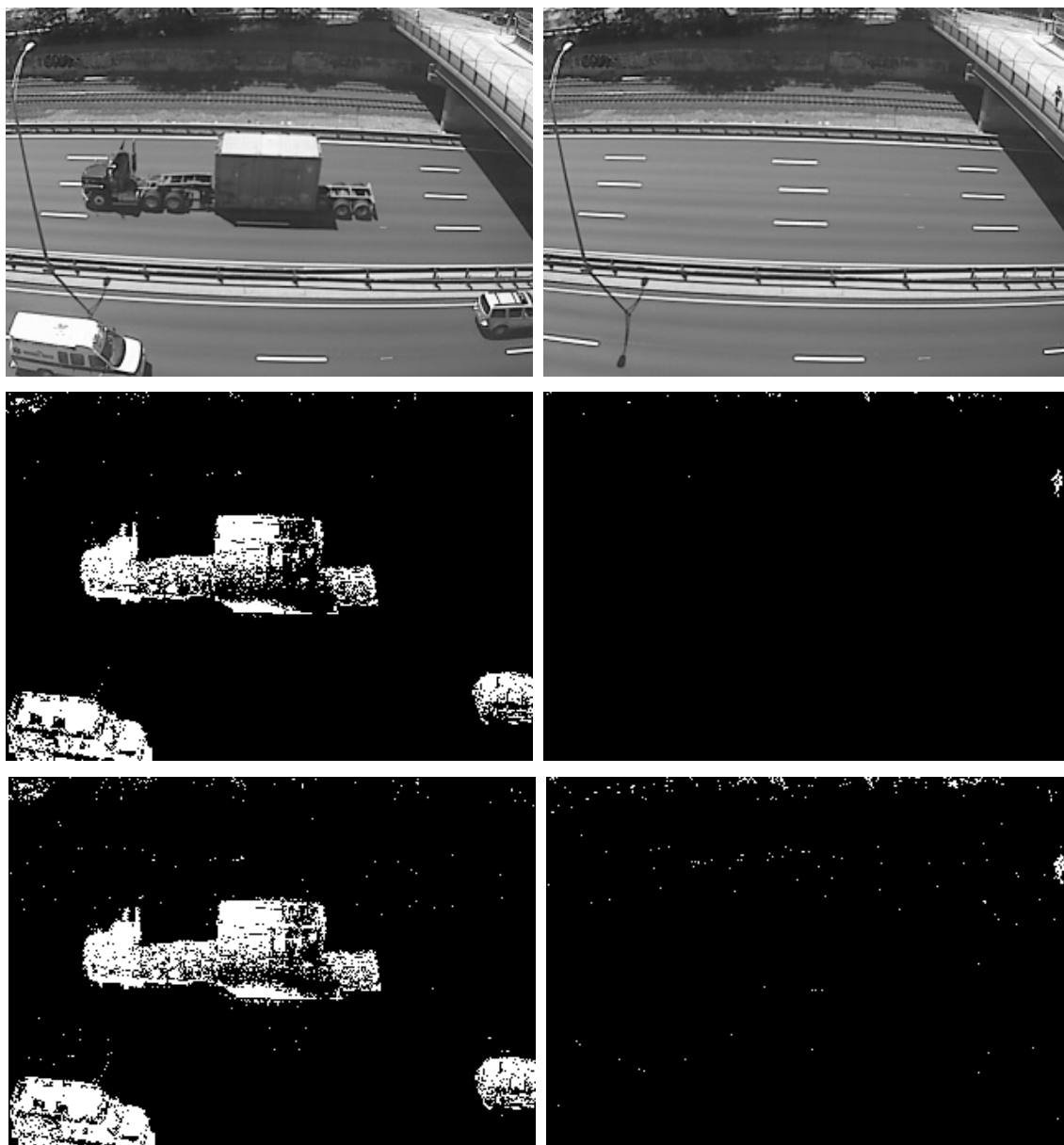


Fig. 2. Detection result for real sequences with medium-density (left) and low-density (right) of moving objects (\mathcal{F}): original frame (top), FDR detection result (middle), and FPR detection result (bottom).

computation at each pixel whereby \mathcal{R}_1 is grown until $p^* \leq f(\theta)$. These non-trivial transformations require more memory to store the PDFs and a substantial amount of time to compute. Reducing this computational complexity is where our future work will concentrate.

6. REFERENCES

- [1] D. Zhang and G. Lu, "Segmentation of moving objects in image sequence: A review," *Circuits, Systems and Signal Process.*, vol. 20, no. 2, pp. 143–183, 2001.
- [2] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 780–785, 1997.
- [3] C. Stauffer and E.L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 747–757, 2000.
- [4] A. Elgammal, R. Duraiswami, D. Harwood, and L.S. Davis, "Background and foreground modeling using nonparametric kernel density for visual surveillance," *Proc. IEEE*, vol. 90, pp. 1151–1163, 2002.
- [5] A. Neri, S. Colonnese, G. Russo, and P. Talone, "Automatic moving object and background separation," *Signal Processing*, vol. 66, no. 2, pp. 219–232, Apr. 1998.
- [6] J. Konrad, "Motion detection and estimation," in *Handbook of Image and Video Processing, 2nd Edition*, A. Bovik, Ed., chapter 3.10, pp. 253–274. Academic Press, 2005.
- [7] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Royal Stat. Soc., B*, vol. 57, no. 1, pp. 289–300, 1995.
- [8] E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley and Sons, 3rd edition, 1986.
- [9] J.M. McHugh, "Probabilistic methods for adaptive background subtraction," M.S. thesis, Boston University, Jan. 2008.