

Foreground-Adaptive Background Subtraction

J. Mike McHugh, Janusz Konrad, Venkatesh Saligrama, and Pierre-Marc Jodoin

Abstract—Background subtraction is a powerful mechanism for detecting change in a sequence of images that finds many applications. The most successful background subtraction methods apply probabilistic models to background intensities evolving in time; non-parametric and mixture-of-Gaussians models are but two examples. The main difficulty in designing a robust background subtraction algorithm is the selection of a detection threshold. In this paper, we adapt this threshold to varying video statistics by means of two statistical models. In addition to a non-parametric background model, we introduce a foreground model based on small spatial neighborhood to improve discrimination sensitivity. We also apply a Markov model to change labels to improve spatial coherence of the detections. The proposed methodology is applicable to other background models as well.

I. INTRODUCTION

Networked video cameras are extensively used today for perimeter surveillance, motor traffic analysis and environmental monitoring. One of the most important tasks in all such applications is *change detection* (and, related, *motion detection*), i.e., automatic segmentation of a video sequence into static and changed (e.g., moving) areas. Among the numerous algorithms developed to date, the simplest ones are based on thresholding intensity differences (e.g., between consecutive video frames or between the current frame and a background frame). Since the detection results are sensitive to threshold selection (false positives versus misses), various threshold adaptation methods have been proposed. Among the most successful are those combining frame differencing with Markov random field (MRF) modeling of change labels [1].

Even with adaptation, frame differencing performs poorly if the background is not truly static (e.g., fluttering leaves, water waves), and thus methods based on probabilistic background models have been developed. While some methods model the probability density function (PDF) of background pixel luminance using a single Gaussian, a mixture of Gaussians [2], or a non-parametric kernel model [3], other methods apply Bayesian framework to a vector of features, such as color, gradients, co-occurrences [4]. Once model parameters have been estimated from previous frames, the detection process involves thresholding the resulting PDF. Due to thresholding of probabilities instead of intensities the approach is more robust and constitutes a powerful tool for change detection.

In parallel, change detection methods were developed based on the maximum *a posteriori* probability (MAP) criterion.

J.M. McHugh, J. Konrad and V. Saligrama are with Boston University, Department of Electrical and Computer Engineering, Boston MA 02215, USA ([jmmchugh, jkonrad, srv]@bu.edu). P.-M. Jodoin is with Université de Sherbrooke, Département d'informatique, Sherbrooke, QC, Canada J1K 2R1 (Pierre-Marc.Jodoin@usherbrooke.ca).

This work was supported by the Dean Catalyst Award from the College of Engineering, Boston University, and by the National Science Foundation under award CNS-0721884.

While some methods were formulated in discrete domain and used MRFs as prior models [5], [6], other methods used variational formulations in continuous domain only embodying the spirit of the MAP criterion. Although MAP-inspired change detection performs well, it is computationally complex; approximate, but faster, solution methods were developed [6].

In this paper, we revisit background subtraction from the hypothesis testing point of view, and make two contributions. Assuming spatial ergodicity, we augment the background model with an explicit foreground model and estimate its parameters from a small *spatial* neighborhood. This is unlike in the past approaches where the foreground model's PDF was assumed uniform [1]–[3], [6] or was estimated from several past frames [7], thus requiring slow object motion or object tracking. The new model improves the method's discrimination sensitivity at the cost of a slight increase in computational complexity. We also embed a Markov model into the hypothesis test to spatially characterize the change labels. This is unlike simple intensity thresholding under MRF constraint [1] or computationally-involved MAP-MRF formulations [6], [7]. Our approach is, in a sense, a compromise that, using a spatially-variable detection threshold, offers an improved spatial coherence of the detections at the cost of modestly increasing computational complexity. It should be noted that the proposed models can be combined with various background models (e.g., non-parametric kernel or mixture of Gaussians).

II. BACKGROUND SUBTRACTION AS A HYPOTHESIS TEST

Background subtraction involves two distinct processes that work in a closed loop: *background modeling* and *foreground detection*. In background modeling, a model of the background in the field of view of a camera is created and periodically updated, for example to account for illumination changes. In foreground detection, a decision is made as to whether a new intensity fits the background model; the resulting *change label field* is fed back into background modeling so that no foreground intensities contaminate the background model.

Let I be a grayscale image sampled on 2-D lattice Λ : $I[\mathbf{n}]$, $\mathbf{n} \in \Lambda \subset \mathbb{R}^2$; extension to color images is straightforward. We denote a sequence of such images $I^{(k)}[\mathbf{n}]$, with k being the frame number. We consider a non-parametric background model [3] for its simplicity and performance. At each background location \mathbf{n} of frame k , this model uses intensity from N recent frames to estimate background PDF:

$$P_{\mathcal{B}}(I^{(k)}[\mathbf{n}]) = \frac{1}{N} \sum_{i=1}^N \mathcal{K}(I^{(k)}[\mathbf{n}] - I^{(k-i)}[\mathbf{n}]), \quad (1)$$

where \mathcal{K} is a zero-mean Gaussian with variance σ^2 that, for simplicity, we consider constant throughout the sequence.

Since the model is based only on recent frames, it adapts to slow background changes such as illumination variations.

If a similar foreground model is available, then change labels can be estimated by evaluating intensity in a new frame against these two models. This entails testing labels \mathcal{B} (background) and \mathcal{F} (foreground) at each pixel of the current image by means of a binary hypothesis test [1] (from now on the superscript k is omitted to simplify notation):

$$\frac{P_{\mathcal{B}}(I[\mathbf{n}])}{P_{\mathcal{F}}(I[\mathbf{n}])} \stackrel{\mathcal{B}}{\underset{\mathcal{F}}{\gtrless}} \eta \frac{\pi_{\mathcal{F}}}{\pi_{\mathcal{B}}}, \quad (2)$$

where on left are the probabilities of observing $I[\mathbf{n}]$ given it is the projection of either the background scene ($P_{\mathcal{B}}$) or a foreground object ($P_{\mathcal{F}}$). On the right-hand side, are the prior probabilities of observing background ($\pi_{\mathcal{B}}$) or foreground ($\pi_{\mathcal{F}}$), and a cost term η . While the ratio $\pi_{\mathcal{F}}/\pi_{\mathcal{B}}$ biases the decision based on the priori probabilities, the cost term η accounts for unequal penalties assigned to the four decision/truth scenarios (\mathcal{B}/\mathcal{B} , \mathcal{B}/\mathcal{F} , \mathcal{F}/\mathcal{F} , \mathcal{F}/\mathcal{B}).

Without an explicit foreground model, $P_{\mathcal{F}}$ is usually considered uniform. Assuming fixed prior probabilities and collecting all constants in (2), this leads to a fixed-threshold background test $P_{\mathcal{B}}(I[\mathbf{n}]) \stackrel{\mathcal{B}}{\underset{\mathcal{F}}{\gtrless}} \theta$. A pixel is labeled as moving if its probability $P_{\mathcal{B}}(I[\mathbf{n}])$ is sufficiently small. However, this simple test is prone to randomly-scattered false positives, even for low θ . Although additional post-processing has been proposed to correct such errors [3], it is *ad hoc* and thus difficult to generalize. We address this by adapting the threshold based on explicit foreground and label field models.

III. FOREGROUND MODELING

In most background subtraction algorithms, $P_{\mathcal{F}}$ is assumed uniform thus preventing any decision bias by moving objects. An exception is the work of ElGammal *et al.* [3], who proposed foreground modeling for human body, and of Sheikh and Shah [7], who proposed a general foreground model using past frames. While the first model is object-specific, the second one necessitates slow object motion as otherwise background samples contaminate $P_{\mathcal{F}}$. Although this could be mitigated by object tracking, such an approach would be illogical (track an object in order to detect it?). Instead, we propose a foreground model based on small *spatial* neighborhood, i.e., in the *same* frame. Recently, we have demonstrated that ergodicity in time also holds spatially; local-in-time and local-in-space models produce equivalent background characteristics [8].

Let $e[\mathbf{n}]$ be a change label at \mathbf{n} ($e[\mathbf{n}] = 0$ for \mathcal{B} and 1 for \mathcal{F}). Using the simplified likelihood ratio test $P_{\mathcal{B}}(I[\mathbf{n}]) \stackrel{\mathcal{B}}{\underset{\mathcal{F}}{\gtrless}} \theta$, we find an initial detection mask $e_0[\mathbf{n}]$ first. Then, for each \mathbf{n} we define a set of neighbors belonging to the foreground:

$$\mathcal{N}_{\mathcal{F}_i}(\mathbf{n}) = \{\mathbf{m} \in \mathcal{N}(\mathbf{n}) : e_i[\mathbf{m}] = \mathcal{F}\},$$

where $i = 0$ and $\mathcal{N}(\mathbf{n})$ is a small neighborhood around \mathbf{n} , and calculate the foreground probability using the kernel-based method (1), but with same-time samples replacing recent ones:

$$P_{\mathcal{F}_i}(I[\mathbf{n}]) = \frac{1}{|\mathcal{N}_{\mathcal{F}_i}(\mathbf{n})|} \sum_{\mathbf{m} \in \mathcal{N}_{\mathcal{F}_i}(\mathbf{n})} \mathcal{K}(I[\mathbf{n}] - I[\mathbf{m}]). \quad (3)$$

If there are no detected pixels in the neighborhood of \mathbf{n} ($\mathcal{N}_{\mathcal{F}_i}(\mathbf{n}) = \emptyset$), there is presumably no foreground object at \mathbf{n} and we revert to the naive assumption that $P_{\mathcal{F}}$ is uniform.

At iteration i , this results in a refined likelihood ratio test:

$$\frac{P_{\mathcal{B}}(I[\mathbf{n}])}{P_{\mathcal{F}_i}(I[\mathbf{n}])} \stackrel{\mathcal{B}}{\underset{\mathcal{F}}{\gtrless}} \theta \quad (4)$$

that produces a new label $e_{i+1}[\mathbf{n}]$. This test is iterated with the new label field defining a new neighborhood $\mathcal{N}_{\mathcal{F}_{i+1}}$ and new PDF $P_{\mathcal{F}_{i+1}}$ which, in turn, produce a new estimate e_{i+2} . In our experience, only a few iterations are needed; as shown in Figs. 3.b and 3.c, the evolution of both $P_{\mathcal{F}_i}$ and e takes place mostly in the first 2-3 iterations. However, since we introduce a positive feedback, the threshold θ must be carefully selected to discourage false detections; otherwise errors may compound [9]. False negatives, on the other hand, will be corrected by Markov model if several neighbors are correctly detected.

Note that $P_{\mathcal{F}_i}(I[\mathbf{n}])$ in (4) can be considered a scale for the threshold θ . If for intensity $I[\mathbf{n}]$ the foreground probability $P_{\mathcal{F}_i}(I[\mathbf{n}])$ is lower, then the effective threshold $\theta \cdot P_{\mathcal{F}_i}(I[\mathbf{n}])$ is reduced thus encouraging assignment of the background label. Our introduction of the foreground model can be interpreted as threshold adaptation.

IV. MARKOV MODELING OF CHANGE LABELS

So far, we have assumed fixed prior probabilities $\pi_{\mathcal{F}}$, $\pi_{\mathcal{B}}$ for the detection labels. Intuitively, however, a pixel surrounded by foreground labels should be more likely to receive a foreground label than a pixel with background neighbors. This can be accomplished by modeling labels as a Markov random field E of which e is a particular realization. MRF models have been successfully used in motion detection [1] reducing scattered false detections and smoothing region boundaries.

We propose a Markov model within the binary hypothesis test (2) while maintaining non-parametric $P_{\mathcal{B}}$, $P_{\mathcal{F}}$. Our approach extends early methods using single-Gaussian $P_{\mathcal{B}}$ and uniform $P_{\mathcal{F}}$ [1], and shares Markovianity with more recent formulations [6], [7]. While in [6] $P_{\mathcal{B}}$ is a two-Gaussian mixture and $P_{\mathcal{F}}$ is uniform, we use more accurate non-parametric models. Also, despite the use of accelerated simulated annealing in [6], the computational complexity is high. Although one can seek local minima by means of one-at-a-time search, such as the iterated conditional modes algorithm, in this case a binary MAP-MRF solution is identical to our binary hypothesis test. Also, our approach uses spatial ergodicity whereas the one in [7] is based on temporal ergodicity which necessitates slow motion or tracking of foreground objects.

Suppose that the label field realization e is known for all \mathbf{m} except \mathbf{n} . This assumption is reasonable since the estimation process is often iterative. Thus, the background detection task is reduced to deciding a label for $e[\mathbf{n}]$ only. Let $e^{\mathcal{B}}$ denote the label field produced when $e[\mathbf{n}] = 0$ and let $e^{\mathcal{F}}$ denote the case when $e[\mathbf{n}] = 1$. Then, the decision rule at \mathbf{n} is:

$$\frac{P(\mathcal{I} = I|e^{\mathcal{B}})}{P(\mathcal{I} = I|e^{\mathcal{F}})} \stackrel{\mathcal{B}}{\underset{\mathcal{F}}{\gtrless}} \eta \frac{P(E = e^{\mathcal{F}})}{P(E = e^{\mathcal{B}})},$$

where $P(\mathcal{I} = I|e^{\mathcal{B}})$ is the joint probability that the whole random field \mathcal{I} assumes realization I (i.e., $\mathcal{I}[\mathbf{m}] = I[\mathbf{m}], \forall \mathbf{m}$),

given label field realization $e^{\mathcal{B}}$, and similarly for $P(\mathcal{I} = I|e^{\mathcal{F}})$. The ratio of these joint probabilities can be simplified if we allow the intensities, while dependent on the label field, to be mutually independent spatially, that is $P(\mathcal{I} = I|e) = \prod_{\mathbf{m}} P(\mathcal{I}[\mathbf{m}] = I[\mathbf{m}]|e)$. Since $P(\mathcal{I} = I|e^{\mathcal{B}})$ and $P(\mathcal{I} = I|e^{\mathcal{F}})$ differ only at \mathbf{n} , common terms corresponding to $\mathbf{m} \neq \mathbf{n}$ cancel out and the above reduces to:

$$\frac{P(\mathcal{I}[\mathbf{n}] = I[\mathbf{n}]|e^{\mathcal{B}})}{P(\mathcal{I}[\mathbf{n}] = I[\mathbf{n}]|e^{\mathcal{F}})} \stackrel{\mathcal{B}}{\geq} \eta \frac{P(E = e^{\mathcal{F}})}{P(E = e^{\mathcal{B}})}, \quad (5)$$

or using the hypothesis test notation from (2):

$$\frac{P_{\mathcal{B}}(I[\mathbf{n}])}{P_{\mathcal{F}}(I[\mathbf{n}])} \stackrel{\mathcal{B}}{\geq} \eta \frac{P(E = e^{\mathcal{F}})}{P(E = e^{\mathcal{B}})}. \quad (6)$$

Since E is a MRF, the *a priori* probabilities on the right-hand side are Gibbs distributions characterized by the natural temperature γ , cliques c , and potential function V defined on c . Choosing the cliques and potential function is crucial to the Gibbs model's effectiveness. We use 2-element cliques $c = \{\mathbf{n}, \mathbf{m}\}$ from the second-order, spatial Markov neighborhood (eight immediate neighbors), commonly used in image processing. An extension to higher neighborhood orders and more complex cliques is straightforward and increases spatial coherence of labels e . As for the potential, it should produce no penalty within a patch of identical labels e , resulting in a high probability, but it should penalize label fields e with severe fragmentation. The Ising potential is commonly used in such scenarios for its simplicity:

$$V(\mathbf{n}, \mathbf{m}) = \begin{cases} 0 & \text{if } e[\mathbf{n}] = e[\mathbf{m}] \\ 1 & \text{if } e[\mathbf{n}] \neq e[\mathbf{m}] \end{cases}. \quad (7)$$

A penalty of 1 is incurred for neighbors different from $e[\mathbf{n}]$.

Since, by definition, $e^{\mathcal{F}}$ and $e^{\mathcal{B}}$ differ only at \mathbf{n} , $P(E = e^{\mathcal{F}})$ and $P(E = e^{\mathcal{B}})$ are identical except for the cliques containing \mathbf{n} . With the Ising potential (7), the hypothesis test (5) becomes:

$$\frac{P_{\mathcal{B}}(I[\mathbf{n}])}{P_{\mathcal{F}}(I[\mathbf{n}])} \stackrel{\mathcal{B}}{\geq} \theta \exp\left(\frac{1}{\gamma}(Q_{\mathcal{F}}[\mathbf{n}] - Q_{\mathcal{B}}[\mathbf{n}])\right) \triangleq \vartheta[\mathbf{n}], \quad (8)$$

where $Q_{\mathcal{F}}[\mathbf{n}]$ and $Q_{\mathcal{B}}[\mathbf{n}]$ denote the number of foreground and background neighbors of \mathbf{n} , respectively, and $\vartheta[\mathbf{n}]$ is the effective spatially-variable threshold. Detailed derivation of test (8) can be found in [9]. Clearly, the detected foreground neighbors increase the effective threshold $\vartheta[\mathbf{n}]$ toward declaring \mathcal{F} whereas background neighbors decrease this threshold toward declaring \mathcal{B} . The parameter γ is selected by the user to control the nonlinear behavior of $\vartheta[\mathbf{n}]$ as a function of $Q_{\mathcal{F}}[\mathbf{n}] - Q_{\mathcal{B}}[\mathbf{n}]$; smaller values of γ strengthen the influence of MRF model on the estimate, while larger values weaken it.

V. EXPERIMENTAL RESULTS

The proposed models were tested on ground-truth data and real video. We used the Gaussian kernel (1) with $\sigma^2=6$ and $N=80$, 7×7 -pixel neighborhood \mathcal{N} (3) and natural Gibbs temperature $\gamma=4/5$ (8). In all experiments θ was fixed at 0.00811, i.e., probability of Gaussian random variable with variance σ^2 deviating from the mean by 6.

Fig. 1 shows the error rate performance (false alarms versus misses) as θ is swept from 0 to 0.02, while Fig. 2 shows

the resulting detections. Both figures compare various model combinations applied to a data set with synthetic objects moving against real quasi-static background (fluttering leaves).

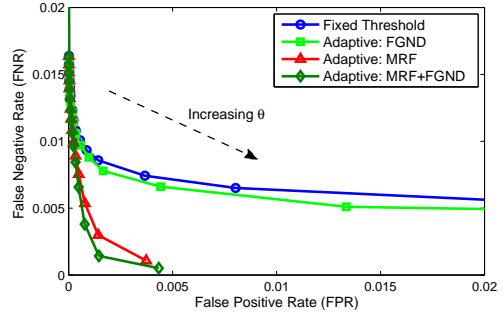


Fig. 1. Performance comparison for background subtraction with fixed and adaptive thresholds on data from Fig. 2. The FPR-FNR pairs depend on θ that increases from 0 to 0.02 in the direction of the arrow.

Clearly, the inclusion of the foreground model improves the performance over the fixed-threshold approach and over the MRF approach, although only slightly in each case (Fig. 1). For example, for a given miss rate (FNR), the foreground-model approach results in a lower false alarm rate (FPR) than the fixed-threshold procedure, while for a given false alarm rate it produces a slightly lower miss rate. Thus, an adjustment of θ in the fixed-threshold approach cannot accomplish the same error rates as the foreground-model approach.

For a given θ , the inclusion of foreground model results in a clear decrease of misses and a slight increase of false alarms (Fig. 2) compared to the fixed-threshold method. However, most of the false alarms and misses are corrected by the addition of Markov model which significantly improves the detection performance (Fig. 1). Again, although gains due to the foreground model are modest, it is clear that the miss rate within the “truck” object is reduced when compared with Markov-only model (Fig. 2 bottom row).

Fig. 3 shows results for video captured by network camera. The combined foreground-MRF model significantly outperforms fixed $P_{\mathcal{B}}$ thresholding as well as the joint $P_{\mathcal{B}}/P_{\mathcal{F}}$ model, and produces accurate results. Figs. 3(b-d) show iterative evolution of one object. Fig. 3(b) shows the background probability $P_{\mathcal{B}}$ as brightness level (top) followed by the evolution of $P_{\mathcal{F}_i}$ (3). Note the complementarity of background and foreground probabilities (where $P_{\mathcal{B}}$ is low, $P_{\mathcal{F}}$ is high) leading to reinforced threshold adaptation. Fig. 3(c) shows the corresponding evolution of the label field; since most gains occur in the first 2-3 iterations, the process may be quickly terminated. A similar convergence can be observed for the Markov model (Fig. 3(d)). Note the characteristic increase of false positives and reduction of misses in foreground evolution, and a reduction of false positives and elimination of the remaining misses in MRF evolution. The final object mask has smooth boundary and is void of misses.

VI. CONCLUSIONS

We have tested the proposed models on other surveillance videos and confirmed the gains reported here [10]. Background subtraction based on these models currently serves as

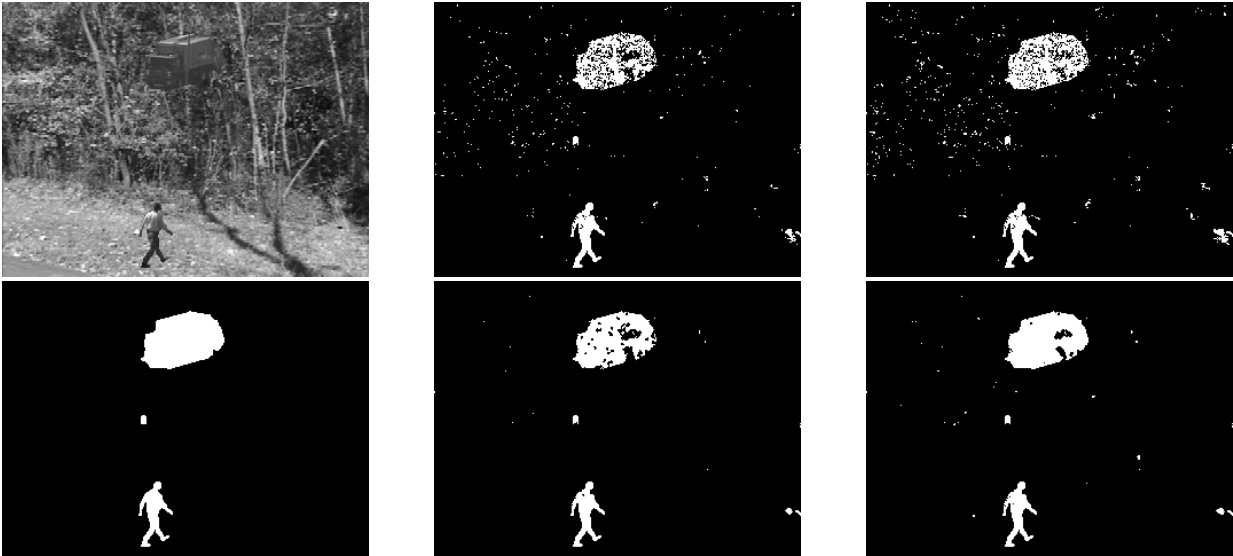


Fig. 2. Background subtraction results for synthetic objects moving against quasi-static background (fluttering leaves): left column – original frame and ground-truth labels, middle column – P_B (top) and $P_B + \text{MRF}$ (bottom) model and right column – P_B/P_F (top) and $P_B/P_F + \text{MRF}$ (bottom) model.

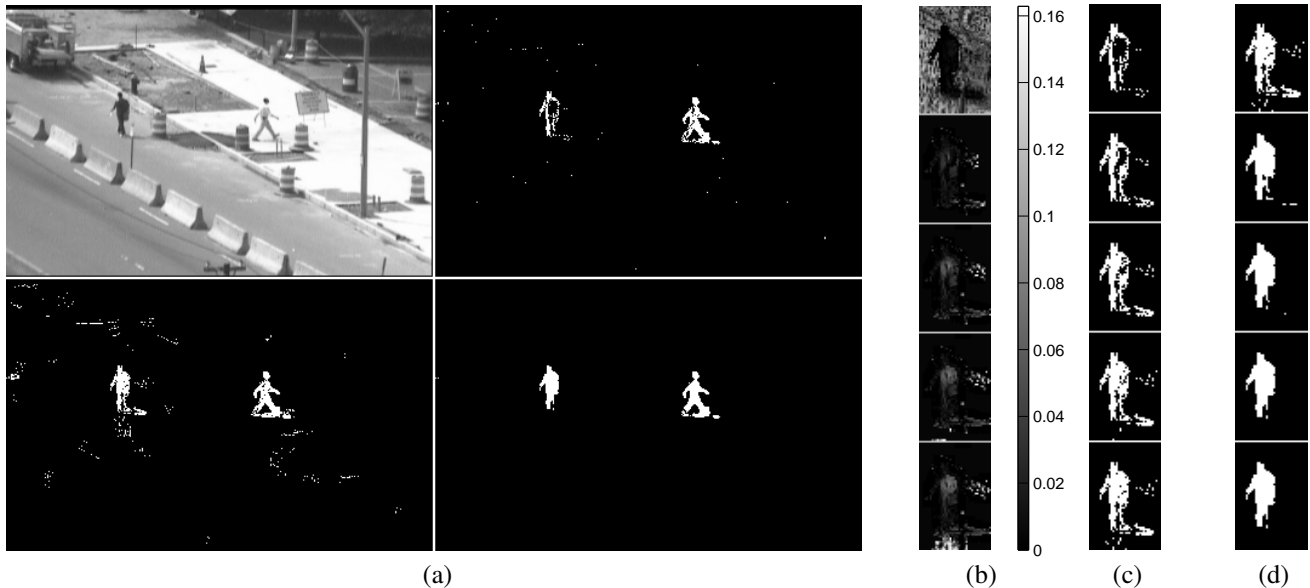


Fig. 3. Background subtraction results for real video: (a) original frame (top-left), detection obtained using P_B model (top-right), P_B/P_F model (bottom-left), and $P_B/P_F + \text{MRF}$ model (bottom-right); and iterative evolution of one object: (b) probabilities $P_B, P_{F_0}, P_{F_1}, P_{F_2}, P_{F_3}$; (c) labels computed using P_F model (P_B followed by $P_B/P_{F_0}, P_B/P_{F_1}, P_B/P_{F_2}, P_B/P_{F_3}$), and (d) labels computed using additional MRF model.

our main change detection algorithm in research on visual behavior analysis and classification. However, the inclusion of a foreground model tends to grow the detected regions rather than shrink them. It is thus critical that the initial label field have as few false positives as possible. This can be accomplished *via* false discovery rate control that uses thresholding of significance scores instead of probabilities [11].

REFERENCES

- [1] T. Aach and A. Kaup, "Bayesian algorithms for adaptive change detection in image sequences using Markov random fields," *Signal Process., Image Commun.*, vol. 7, pp. 147–160, 1995.
- [2] C. Stauffer and E. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 747–757, 2000.
- [3] A. Elgammal, R. Duraiswami, D. Harwood, and L. Davis, "Background and foreground modeling using nonparametric kernel density for visual surveillance," *Proc. IEEE*, vol. 90, pp. 1151–1163, 2002.
- [4] L. Li, W. Huang, I. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1459–1472, Nov. 2004.
- [5] N. Paragios and V. Ramesh, "A MRF-based approach to real-time subway monitoring," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 2001, pp. 58–65.
- [6] J. Migdal and E. L. Grimson, "Background subtraction using Markov thresholds," in *Proc. IEEE Workshop Motion and Video Computing, WACV/MOTIONS'05*, vol. 2, 2005, pp. 58–65.
- [7] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 11, pp. 1778–1792, 2005.
- [8] P.-M. Jodoin, M. Mignotte, and J. Konrad, "Statistical background subtraction using spatial cues," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, pp. 1758–1763, Dec. 2007.
- [9] J. McHugh, "Probabilistic methods for adaptive background subtraction," Master's thesis, Boston University, Jan. 2008.
- [10] http://iiss.bu.edu/jkonrad/Research/VSNs/Background_Subtraction/background_subtraction.html.
- [11] J. McHugh, J. Konrad, V. Saligrama, P.-M. Jodoin, and D. Castanon, "Motion detection with false discovery rate control," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 2008.