Background Subtraction on Distributions

Teresa Ko, Stefano Soatto, and Deborah Estrin

Vision Lab
Computer Science Department
University of California, Los Angeles
405 Hilgard Avenue, Los Angeles – CA 90095
{tko,soatto,destrin}@cs.ucla.edu

Abstract. Environmental monitoring applications present a challenge to current background subtraction algorithms that analyze the temporal variability of pixel intensities, due to the complex texture and motion of the scene. They also present a challenge to segmentation algorithms that compare intensity or color distributions between the foreground and the background in each image independently, because objects of interest such as animals have adapted to blend in. Therefore, we have developed a background modeling and subtraction scheme that analyzes the *temporal variation* of intensity or color *distributions*, instead of either looking at temporal variation of point statistics, or the spatial variation of region statistics in isolation. Distributional signatures are less sensitive to movements of the textured background, and at the same time they are more robust than individual pixel statistics in detecting foreground objects. They also enable slow background update, which is crucial in monitoring applications where processing power comes at a premium, and where foreground objects, when present, may move less than the background and therefore disappear into it when a fast update scheme is used. Our approach compares favorably with the state of the art both in generic low-level detection metrics, as well as in application-dependent criteria.

1 Introduction

Background subtraction is a popular pre-processing step in many visual monitoring applications, as it facilitates the detection of objects of interest ("foreground"). Even when the cameras are fixed in the infrastructure, however, naive background modeling and subtraction results in large numbers of false detections because of changes in illumination and fine-scale motion in the scene. Natural environments such as the forest canopy present an extreme challenge because the foreground objects, by necessity, blend with the background, and the background itself changes due to the motion of the foliage and the rapid transition between light and shadow. For instance, images of birds at a feeder station exhibit a larger per-pixel variance due to changes in the background than due to the presence of a bird. Rapid background adaptation fails because birds, when present, are often moving less than the background and often end up being incorporated into it.

Even a summary inspection of a short video will easily convince the reader that neither analysis of the temporal variation of a single pixel, common to many background subtraction methods, nor analysis of the spatial statistics of each image in isolation, common to many image segmentation algorithms, is sufficient to detect the presence

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part III, LNCS 5304, pp. 276–289, 2008.

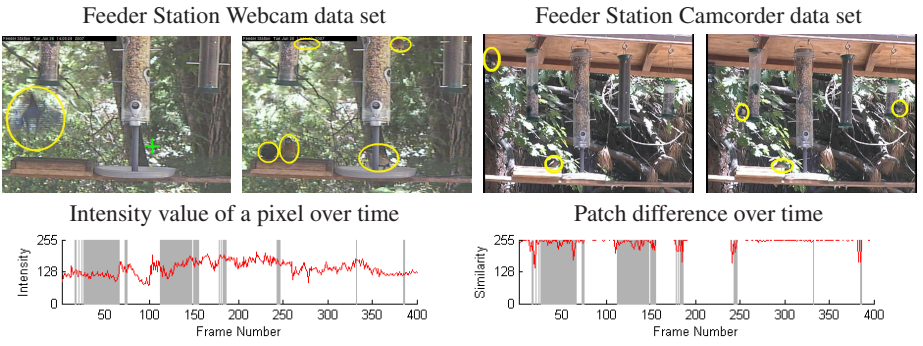

Fig. 1. Birds are difficult to detect due to their similarity with the background and the large temporal variability of the background. Examination of the intensity value of a pixel over time reveals minimal variability in the presence of birds (gray regions). This motivates the use of distributional signatures.

of birds. This is illustrated in Fig. 1. Therefore, we advocate background processing algorithms that analyze *not individual images* but video, by comparing *not single-pixel statistics*, but spatial distributions of pixel intensities or color. An additional peculiarity of environmental monitoring sequences is the *coarse temporal sampling*, dictated by energy considerations since the cameras deployed in natural environments are battery-operated. This makes learning the temporal dynamics of the background motion, as done in Dynamic Textures [1], impossible. On the other hand, only coarse localization of foreground objects is required, as input to subsequent stages where scientists can measure biodiversity by counting the number of birds visiting the feeder station, or placing a bounding box for subsequent processing such as species classification. The benefit of an automated approach to background modeling and foreground detection is readily measured in the amount of person-time saved by scientists in analyzing these long and tedious sequences.

The resulting approach, consisting of analyzing the temporal variation of intensity distributions, rather than pixel values, is a departure from traditional background subtraction schemes. *We represent the signature of each pixel using a distribution of pixel intensities in a neighborhood, and use the Bhattacharyya distance to compare such distributions over time.* This distribution signature is relatively insensitive to small movements of the highly textured background, and at the same time is not tied to individual pixel values for detecting foreground objects. This enables slower background updates, and therefore minimizes the probability that the foreground object be incorporated into the background. Indeed, the background update rate can be chosen depending on the application within a broad range.

We use bird monitoring in natural scenes as a motivating application to bring attention to a far larger class of significant scenes not previously addressed in the literature. A number of pertinent questions about the impact of climate change on our ecosystem are most readily answered by monitoring fine-scale interactions between animals and plants and their environment. Such fine scale measurements of species distribution, feeding habits, and timing of plant blooming events require continuous monitoring in

the natural environment, and are plagued by the same challenges as the feeder station monitoring described in this paper. There is inherent pressure to increase spatial coverage at the cost of reducing the size of the objects of interest in the image, thereby creating a more challenging detection and recognition task. Similarly, increasing temporal coverage (lifetime) pushes for lower sampling rates limiting the applicable methods.

The field of computer vision has made great strides in addressing challenging problems by simplifying the problem with key assumptions. When trying to use techniques developed previously for other use cases, we found that these assumptions did not hold under a large class of our use case scenes. In this paper, we demonstrate an approach that is more generally applicable to a larger set of natural scenes than previous work.

2 Related Work

The most straightforward approach to segment foreground from background, frame differencing [2], thresholds the difference between two frames. Large changes are considered foreground. To resolve ambiguity due to slow moving objects, Kameda and Minoh [3] use a "double difference" that classifies foreground as a logical "add" of the pairwise difference between three consecutive frames.

Another approach is to build a representation of the background that is used to compare against new images. One such approach captures a background image when no foreground objects are present, assuming some user control over the environment. A compromise between differencing neighboring frames and differencing against a known background image is to adapt the background over time by incrementally incorporating the current image into the background. Migliore *et al.* [4] integrate frame differencing and background modeling to improve overall performance.

As needed, added complexity in the model would allow for added complexity in the background scene. W^4 [5] was one of the first to incorporate more powerful statistics by modeling the variance found in a set of background images with the maximum and minimum intensity value and the maximum difference between consecutive frames. Pfinder [6] uses the mean and the variance of pixel value. If all that is known about a distribution is the mean and variance, the most reasonable assumption based on maximal entropy is the Gaussian distribution. The assumption then is that the pixel value follows a Gaussian distribution, and a likelihood model is used to compare the likelihood of background and foreground for a particular pixel. When this assumption does not adequately account for the variance, a Mixture of Gaussians (MoG) can be used [7,8] to further improve the accuracy of the estimate. A MoG model is capable of handling a range of realistic scenarios, and is widely used [9,10].

Rather than extending the MoG model, Elgammal *et al.* [11] show it is possible to achieve greater accuracy under the same computational constraints as the MoG when using a non-parametric model of the background. Another significant contribution of this work was the incorporation of spatial constraints into the formulation of foreground classification. In the second phase of their approach, pixel values that could be explained away by distributions of neighboring pixels were reclassified as background, allowing for greater resilience against dynamic backgrounds. Sheikh and Shah unify the temporal and spatial consistencies into a single model [12]. The result is highly accurate

segmentations of objects even when occluding a dynamic background. Similar models include [13,14,15].

A different approach, taken by Oliver *et al.* [16], looks at global statistics rather than the local constraints used in the previously described work. Similar to eigenfaces, a small number of "eigenbackgrounds" are created to capture the dominant variability of the background. The assumption is that the remaining is due to foreground objects. A threshold on the difference between the original image and the part of the image that can be generated by the eigenbackgrounds differentiates the foreground objects from the background.

Rather than implicitly modeling the background dynamics, many approaches have explicitly modeled the background as composed of dynamic textures [17]. Wallflower [18] uses a Wiener filter to predict the expected pixel value based on the past K samples whose α's are learned. Monnett *et al.* [1] model the background as a dynamic texture, where the first few principal components of the variance of a set of background images (similar to [16]) comprise an autoregressive model in the same vein as [18]. For computational efficiency, Kahl *et. al.* [19] illustrates that using "eigenbackgrounds" on shiftable patches in an image is sufficient to capture the variance in dynamic scenes.

The inspiration for this work came from Rathi's success in single image segmentation using the Bhattacharyya distance [20]. Similarly, the consistency of the Bhattacharyya distance under motion is used for tracking in [21]. While many distances between distributions exist (*e.g.*, Kullback-Leibler divergence [22], χ^2 [23], Earth Mover's Distance [24], *etc.*), the Bhattacharyya distance is considered here due to its low computational cost.

3 Background Model

In our approach, a background model is constructed for each pixel location, including pixel values with temporal and spatial proximity. A distribution is constructed for each pixel location on the current image and compared to the background model for classification.

3.1 Modeling the Background

The background model for the pixel located at the *i*th row and *j*th column is in general a non-parametric density estimate, denoted by $p_{ij}(x)$. The feature vector, $x \in \mathbb{R}^3$, is some color-space representation of the pixel value. For computational reasons, we consider the simplest estimate, given by the histogram

$$p_{ij}(x) = \frac{1}{|S|} \sum_{s \in S} \delta(s - x), \tag{1}$$

where S, the set of pixel values contributing to the estimate is defined as

$$S = \{x_t(a, b) \mid |a - i| < c, |b - j| < c, 0 \leq t < T\}, \tag{2}$$

where $x_t(a, b)$ is the colorspace representation of the pixel at the *a*th row and *b*th column of the image taken at time *t*. The feature vector, x, is quantized to better approximate the true density.

Elgammal *et al.* and Sheikh and Shah similarly model the background as a non-parametric density estimate. A generalized form of Eq. (1),

$$p_{ij}(x) = \frac{1}{|S|} \sum_{s \in S} K(s - x), \tag{3}$$

where K is a kernel function that satisfies $\int K(x)dx = 1, K(x) = K(-x), \int xK(x)dx = 0$, and $\int xx^T K(x)dx = I_{|x|}$, can better approximate the true distribution when the size of S is small. Elgammal *et al.* construct a model using an independent Gaussian kernel for each pixel using only sample points of close temporal proximity,

$$S = \{x_t(i, j) \mid 0 \leq t < T\}. \tag{4}$$

While Sheikh and Shah also use the independent Gaussian kernel, they model the entire background with a single density estimate of the same form as Eq. (1) except the feature vector, x, is appended with the pixel location, (i, j). The set of pixels used to construct the estimate is

$$S = \{x_t(a, b) \mid 0 \leq a < h, 0 \leq b < w, 0 \leq t < T\} \tag{5}$$

for a $w \times h$ image. Also, Sheikh and Shah adopt the simpler δ kernel function when the algorithm is optimized for speed.

3.2 Temporal Consistency

To detect foreground at time τ, a distribution, $q_{ij,\tau}(x)$, is similarly computed for the pixel located in the ith row and jth column using only the image at time τ according to

$$q_{ij,\tau}(x) = \frac{1}{|S_\tau|} \sum_{s \in S_\tau} \delta(s - x), \tag{6}$$

where S_τ, the set of pixel values contributing to the estimate is defined as

$$S_\tau = \{x_\tau(a, b) \mid |a - i| < c, |b - j| < c\}, \tag{7}$$

The Bhattacharyya distance between $q_{ij,\tau}(x)$ and the corresponding background model distribution for that location, $p_{ij,\tau-1}(x)$, calculated from the previous frames, is computed to determine the foreground/background labeling. The Bhattacharyya distance between two distributions is given by

$$d = \int_X \sqrt{p_{ij,\tau-1}(x) q_{ij,\tau}(x)} dx, \tag{8}$$

where X is the range of valid x's. d ranges from 0 to 1. Larger values imply greater similarity in the distribution. A threshold on the computed distance, d, is used to distinguish between foreground and background.

While subtle, enforcing spatial consistency in the current image results in dramatic improvements in the performance of our scheme compared to previous work, as will be demonstrated in the following sections. This approach can be viewed as a hybrid between pixel- and texture-level comparisons. The distribution computed on the image at time τ, $q_{ij,\tau}(x)$ could be viewed as a feature vector describing the texture at that location. But, rather than building a density estimate in this feature space as the background model, we treat each pixel independently in our distribution. While previous approaches have mentioned the use of generic image statistics as a pixel-level representation which could take into account neighboring statistics, a direct application of their approach would have been prohibitively expensive in terms of memory and computation. Our approach is one that compromises model precision so as to feasibly fit on commonplace devices and, as we shall show later on, still works well for our application domain. We expect that this approach will work well for large but consistent background changes and foreground objects that are similar in appearance to the background, exhibit sudden motion and periods of stationarity, and do not necessarily dominate the scene.

3.3 Updating the Background

The background is adapted over time so that the probability at time τ is

$$p_{ij,t}(x) = (1 - \alpha)p_{ij,\tau-1}(x) + \alpha q_{ij,\tau}(x), \tag{9}$$

where α is the adaptation rate of the background model.

4 Experimentation

We compare our approach against a set of background subtraction methods that handle large variances in the background as well as frame differencing for a baseline.

In our experiments with Elgammal's Non-Parametric Background model, we vary the decision threshold on the difference image and α, which determines the relative pixel values considered as shadowed. A difference image was constructed from the probability of a pixel value coming from the background model. The implementation was kindly provided by the author of the paper. We implemented the speed-optimized version of Sheikh's Bayesian model due to the computational constraints of the application domain, and varied the number of bins used to approximate the background model. We also tested our implementation of Oliver's Eigenbackground model with a varying number of eigenbackgrounds used to model the background. The results presented are the best precision-recall pairs found across these parameters.

The subsequent parts of the algorithm suggested by Elgammal *et al.* and Sheikh and Shah used to provide clean segmentations (*i.e.*, morphological operations and minimum cuts on Markov Random Field, respectively) could arguably be used by any of the algorithms tested, so the informativeness of the results could be obscured by the varying abilities of these operations. Therefore, when comparing these algorithms, we look only at the construction of the background model and the subsequent comparison against the current frame.

4.1 Data Sets

We collected video sequences from two independent cameras pointed at a feeder station. This feeder station had been previously set up to aid biologists in observing avian behavior. The characteristics of data sets evaluated in this paper are:

- Feeder Station Webcam: Images are captured by webcam at 1 frame per second. The size of these images are 480×704.
- Feeder Station Camcorder: Images are captured at full NTSC speed. The size of the resulting images is 480×640.
- Sheikh's: This dataset consists of $\backsim 70$ frames of only background, followed by a person and a car traversing the scene in the opposite direction.

Examples of these data sets are shown in Figs. 2, 4, and 5. Each data set was hand labeled by defining the outline of the foreground objects.

4.2 Metrics

We use several metrics in our evaluation to try to best characterize the performance. The most direct measure we use is the precision and recall of each pixel, following the methodology of Sheikh's work [12]. They are defined as follows:

$$Precision = \frac{\#TruePositives}{\#TruePositives + \#FalsePositives} \tag{10}$$

$$Recall = \frac{\#TruePositives}{\#TruePositives + \#FalseNegatives}. \tag{11}$$

This measures the accuracy of the approach at the pixel level, but does not capture precisely its ability to give reasonable detections of birds for higher level classification. In some cases, performance may easily be improved by morphological operations or the use of Markov Random Fields, evidenced in [11,12]. On the other hand, if the predicted foreground pixels that do correspond to birds are not connected (or cannot reasonably be connected through some morphological operation), each disjoint set of foreground pixels could be interpreted as a bird on its own, resulting in partial birds being fed to a classifier or, worse, discarded because the region is too small. To better quantify the performance of our approach, we also look at the precision and recall of birds. If 10% of the bird is detected as a single blob, a then it is counted as a true positive. Otherwise, the bird is considered a miss, and counted as a false negative. The smaller blobs are discarded, and those remaining blobs are counted as false positives. These values are used to compute the final precision and recall of birds.

More important is the end goal of understanding the nature of a bird's visit. At the application level, we would like an algorithm that successfully detects a wide range of objects, rather than the one that may happen to dominate our test sequence. To that end, we treat a visit to the feeder station as one event. Visit accuracy is the percent of visits where the bird is detected at least once.

4.3 Results

A challenge of the data sets considered is the absence of long periods of only background. For the Feeder Station data sets, we tested with 20 background images. For Sheikh's data set which had more background images, we used 100 background images. As shown in Figure 3, our approach significantly outperforms the others on the Feeder Station data sets, and performs comparably on Sheikh's data set. Representative frames of the Feeder Station Webcam are shown in Figure 2 and of the Feeder Station Camcorder in Figure 4.

While most birds are detected correctly, our approach has trouble when the bird is similar to parts of the background, such as in image A in Figure 2, where the bird is

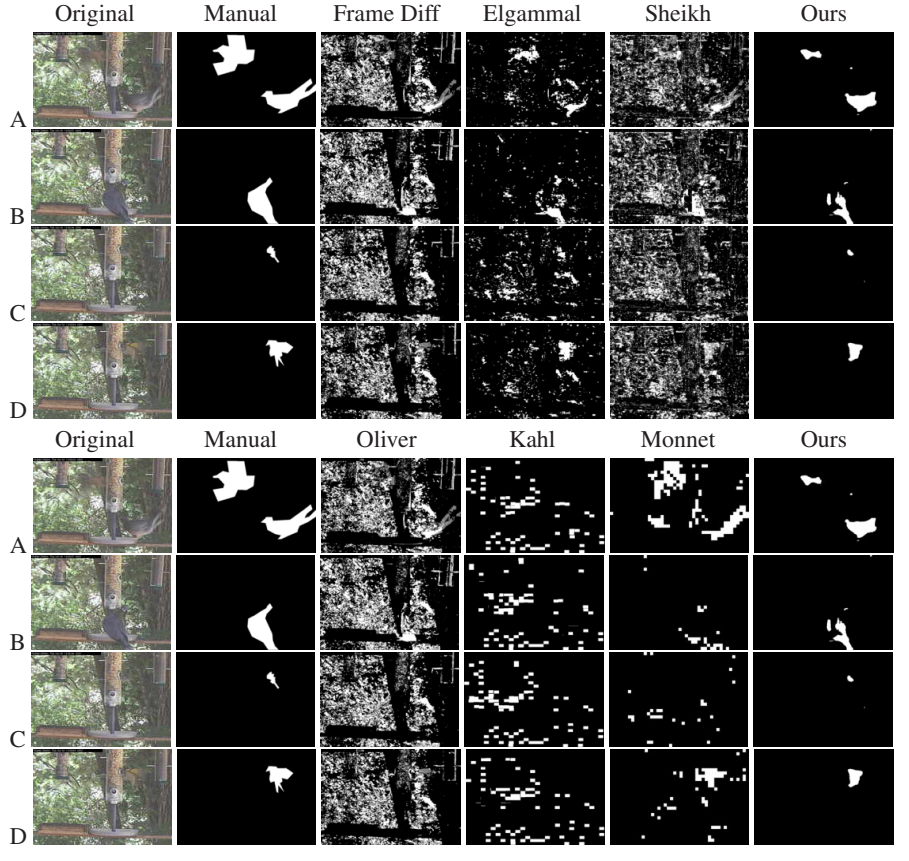

Fig. 2. Results on the Feeder Station Webcam Dataset. Parameters were varied to find the maximum precision subject to pixel recall $> 50\%$. 20 background images were used to train the background models. Our approach works well under most situations, but is unable to deal well with narrow regions. This is shown in row A, where the tail is lost on the bottom bird. Also, another weakness is when the background is of similar color. The bird is broken up into many segments, as shown in row B.

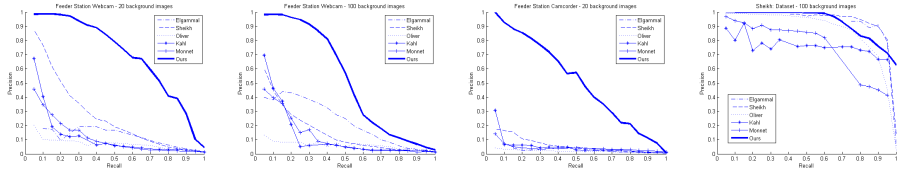

Fig. 3. Comparison of Elgammal *et al.*, Sheikh and Shah, Oliver *et al.*, Kahl *et al.*, Monnet *et al.* and our approach using Precision-Recall curves on selected data sets. Our approach outperforms other approaches on the Feeder Station data sets and is comparable to others (as shown with Sheikh's data set).

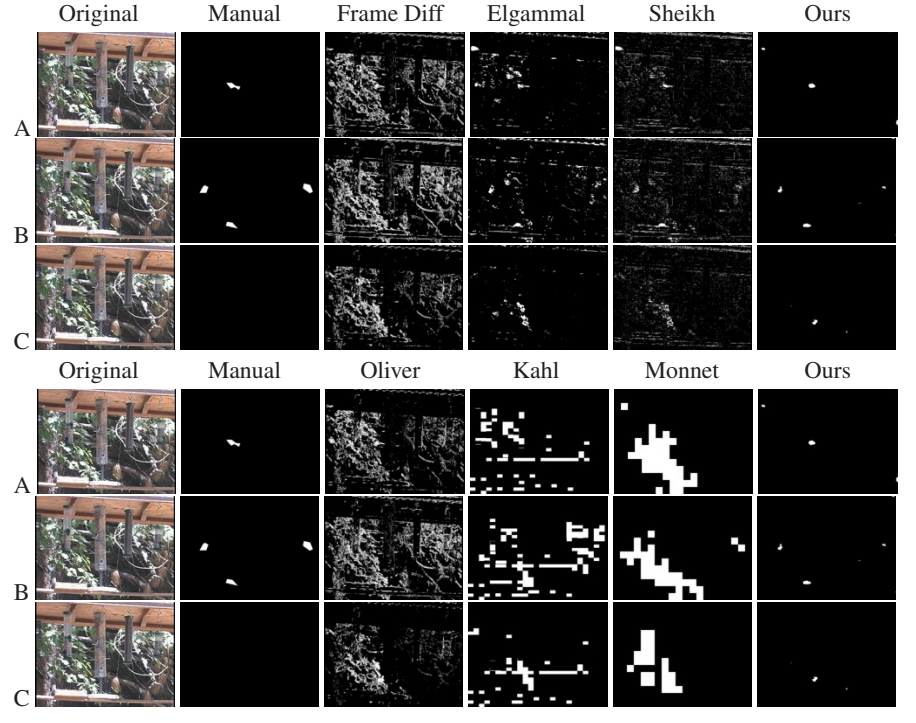

Fig. 4. Results on the Feeder Station Camcorder Dataset. Parameters were varied to find the maximum precision subject to pixel recall $> 50\%$. 20 background images were used to train the background models. Most of the previous approaches a challenged by the slight movement of the camera during the sequence run and the strongest response is to the movement of the leaves in the background. Our approach better filters out this movement. In some cases, the automatic approaches actually outperformed the manual labeling. In row A, the detected region in our approach corresponds to birds, verified after the fact.

separated into multiple segments. Surprisingly, though, in the Feeder Station Camcorder data set, our approach detected birds that were missed by the manual labeling, as in the upper left of image A in Figure 4.

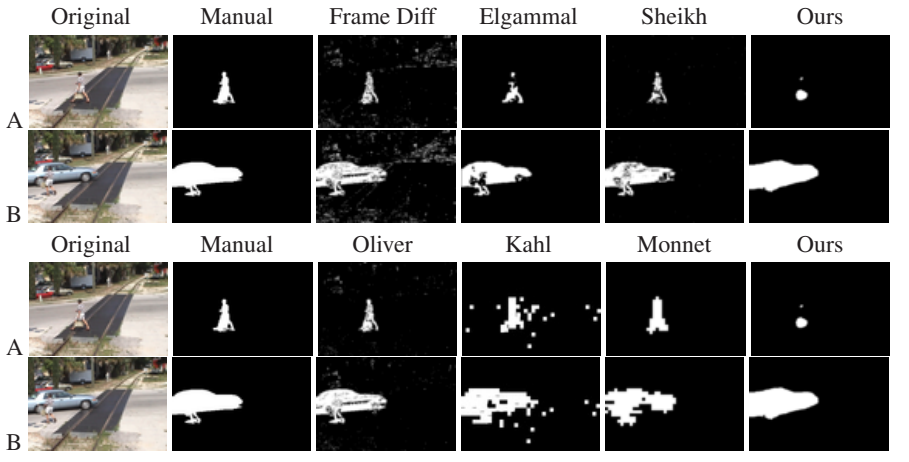

Fig. 5. Results on the Sheikh's dataset. Parameters were varied to find the maximum pixel precision subject to pixel recall $> 80\%$. 100 background images were used to train the background model. Incorporating neighboring pixels in the current image for classification, our approach results in some inherent blurring of decision boundaries, resulting in blob-like classifications rather than the more detailed boundaries of other schemes. In our application domain, this is not a significant drawback.

The suggested methodology for Oliver *et al.*, Kahl *et al.*, and Monnet *et al.* is to include all images in the eigenbackground computation with the assumption that foreground objects do not persist in the same location for long periods of time. If they are infrequent, they will not constitute the principal dimensions of variance. Adopting this assumption, we tested all the algorithms against a background model with 100 frames (where birds are at times present) and see in most cases a degradation in performance as compared to when only 20 frames are used for the background, shown in Figure 3.

One drawback of our approach (though, not particularly relevant for our application) is that accurate boundaries are lost. A comparison is made against the data set used in [12] to illustrate this drawback. We see in this case, as shown in Figure 5, the algorithm suffers from its inherent blurring, missing the full contour of the person in image A. This is quantified in Figure 3. Interestingly enough, frame differencing works relatively well on this data set, especially if coupled with some morphological filtering.

Due to the nature of the monitoring application, background adaptation is particularly difficult, yet obviously necessary. Scientists are interested in monitoring the outdoors for prolonged periods of time and do not necessarily have control over the environment to gather background images whenever desired.

Blind adaptation is very sensitive to the rate of adaptation, as shown in Figure 6. If too slow, the background model does not adequately account for the changes. If too fast, the stationary birds go undetected due to integration into the foreground. Updating only the background model of pixel locations that are clearly background (far from the decision boundary of foreground/background) results in less sensitivity in the choice of α. Since we account for most of the pixel value variance as movement, the remaining changes are slow and due mostly to the lighting changes during the day. When sampling one

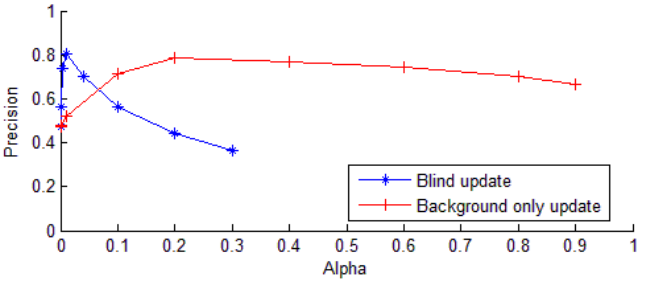

Fig. 6. When blindly updating the background, only a limited range of α's (the adaptation rate of the background) maintain high precision at a fixed recall rate. By selectively updating a pixel's background model when no bird is present at that location, the accuracy is less affected by the selected α.

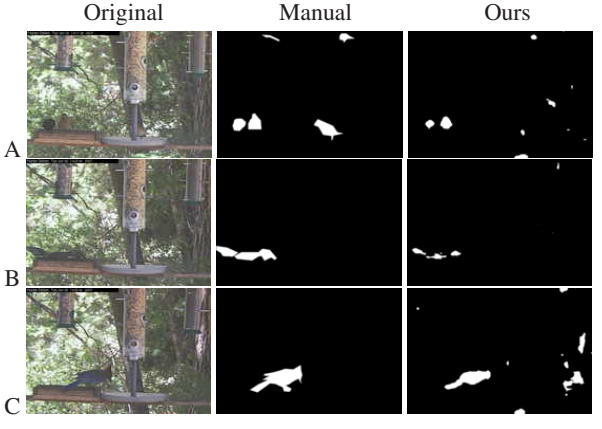

Fig. 7. Results of our approach on sample images over an hour using a background update every minute. Even with very infrequent updates, our approach is able to continually segment birds. Remaining unaddressed situations include specular highlights resulting in false positives, as shown in the row C.

frame per minute, we still achieve 85.21% precision at 50% recall.This allows greater flexibility in system design including situations where full video frame rate capture and analysis is either infeasible or impractical.

Sample frames are shown in Figure 7 illustrating the background model's ability to continuously resolve images over the course of an hour.

Figure 8 shows the corresponding bird precision-recall curves for the Feeder Station Webcam and Camcorder using 20 background images. Most missed birds occured at the boundary of the image. At bird recall $> 59.5\%$ on the Feeder Station Webcam data set, our approach detected 37 of out 39 birds during their visit. The two undetected birds were captured only in one frame each and either suffered from interlacing effects or were only partially shown. At bird recall $> 48.2\%$ on the Feeder Station, 4 out of 4 birds were detected. In Figure 9, we provide a subset of cropped images detected

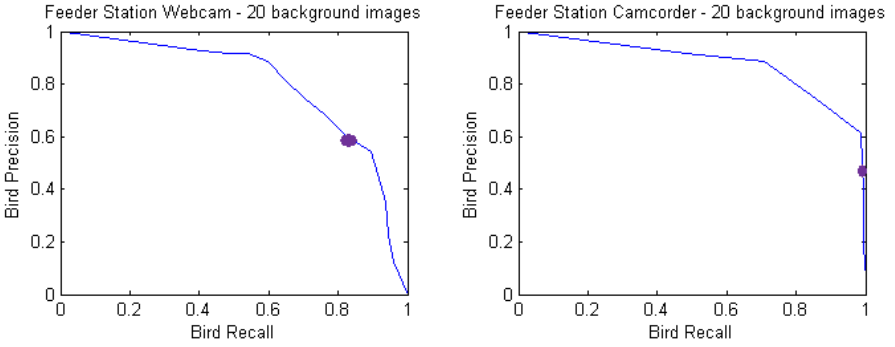

Fig. 8. Precision-Recall Curves for Feeder Station Webcam and Camcorder data sets. Performance suffers in the Feeder Station Webcam when birds are cut off by the capture window. At the point indicated by the purple dot, we detect 37 out of 39 and 4 out of 4 bird visits on the respective data sets.

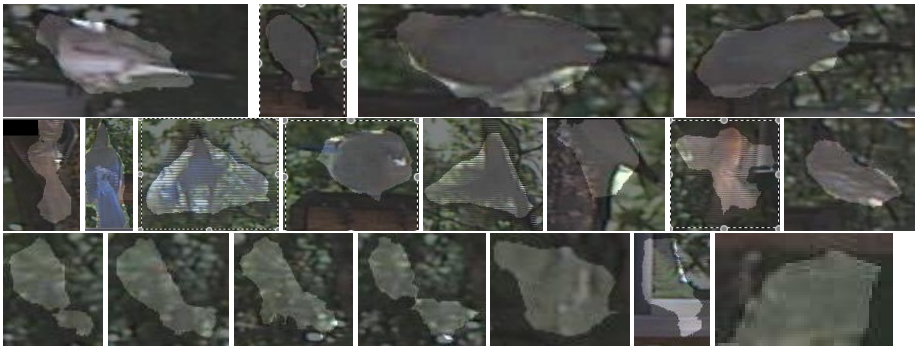

Fig. 9. Sample segmentations from our approach, including both true detections and false alarms

by our approach. We improved the selected area of the detected bird by lowering the pixel difference threshold around our detection. A simple clustering algorithm could separate these images into a reasonably sized set for a biologist to view and provide domain-specific knowledge, such as the species of the bird.

5 Conclusion

In this paper, we call to attention several inherent characteristics of natural outdoor environmental monitoring that pose a challenge to automated background modeling and subtraction. Namely, foreground objects tend to, by necessity, blend into the background, and the background exhibits large variations due to non-stationary objects (moving leaves) and rapid transitions from light to shadow. These conditions present a challenge to the state of the art, which we have addressed with an algorithm that exhibits comparable performance also on standard surveillance data sets.

A side benefit of this approach is that it has relatively low memory requirements, does not require floating point operations, and for the most part, can run in parallel. This makes it a good candidate for embedded processing, where Single Instruction, Multiple Data (SIMD) processors are available. Because the scheme does not depend on a high sampling rate, needed for optical flow or dynamic texture approaches, it lends itself to an adjustable sampling rate. This ability to provide an embedded processor that can easily capture objects facilitates scientific observation of phenomena that are consistently difficult to reach (e.g, environmental, space, underwater, and more general surveillance monitoring).

Acknowledgments

This material is based upon work supported by the Center for Embedded Networked Sensing (CENS) under the National Science Foundation (NSF) Cooperative Agreement CCR-012-0778 and #CNS-0614853, by the ONR under award #N00014-08-1-0414, and by the AFOSR under #FA9550-06-1-0138. Any opinions, ndings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reect the views of the NSF, the ONR or the AFOSR.

References

1. Monnet, A., Mittal, A., Paragios, N., Ramesh, V.: Background modeling and subtraction of dynamic scenes. In: International Conference on Computer Vision, pp. 1305–1312 (2003)
2. Jain, R., Nagel, H.: On the analysis of accumulative difference pictures from image sequences of real world scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence 1(2), 206–214 (1979)
3. Kameda, Y., Minoh, M.: A human motion estimation method using 3-successive video frames. In: International Conference on Virtual Systems and Multimedia, pp. 135–140 (1996)
4. Migliore, D., Matteucci, M., Naccari, M., Bonarini, A.: A revaluation of frame difference in fast and robust motion detection. In: VSSN 2006. Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks, pp. 215–218 (2006)
5. Haritaoglu, I., Harwood, D., Davis, L.S.: W4: real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 809–830 (2000)
6. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pfinder: Real-time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 780–785 (1997)
7. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: Computer Vision and Pattern Recognition, vol. 2, pp. 246–252 (1999)
8. Friedman, N., Russell, S.: Image segmentation in video sequences: A probabilistic approach. In: 13th Conference on Uncertainty in Aritficial Intelligence, pp. 175–181 (1997)
9. Harville, M.: A framework for high-level feedback to adaptive, per-pixel, mixture-of-gaussian background models. In: European Conference on Computer Vision, pp. 543–560 (2002)
10. Tian, Y.L., Lu, M., Hampapur, A.: Robust and efficient foreground analysis for real-time video surveillance. In: Computer Vision and Pattern Recognition, pp. 1182–1187 (2005)
11. Elgammal, A., Harwood, D., Davis, L.: Non-parametric model for background subtraction. In: European Conference of Computer Vision, pp. 751–767 (2000)

12. Sheikh, Y., Shah, M.: Bayesian modeling of dynamic scenes for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(11), 1778–1792 (2005)
13. Mittal, A., Paragios, M.: Motion-based background subtraction using adaptive kernel density estimation. In: IEEE International Conference on Computer Vision and Pattern Recogntion, pp. 302–309 (2004)
14. Pless, R., Larson, J., Siebers, S., Westover, B.: Evaluation of local models of dynamic backgrounds. In: Computer Vision and Pattern Recognition, vol. II, pp. 73–78 (2003)
15. Ren, Y., Chua, C.-S., Ho, Y.-K.: Motion detection with nonstationary background. Machine Vision and Applications 13, 332–343 (2003)
16. Oliver, N.M., Rosario, B., Pentland, A.P.: A bayesian computer vision system for modeling human interactions. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 831–843 (2000)
17. Doretto, G., Cremers, D., Favaro, P., Soatto, S.: Dynamic texture segmentation. In: International Conference of Computer Vision, pp. 1236–1242 (2003)
18. Toyama, K., Krumm, J., Brumitt, B., Meyers, B.: Wallflower: Principles and practice of background maintenance. In: Seventh International Conference on Computer Vision, pp. 255–261 (1999)
19. Kahl, F., Hartley, R., Hilsenstein, V.: Novelty detection in image sequences with dynamic background. In: 2nd Workshop on Statistical Methods in Video Processing (SMVP), European Conference on Computer Vision (2005)
20. Rathi, Y., Michailovich, O., Malcolm, J., Tannenbaum, A.: Seeing the unseen: Segmenting with distributions. In: Signal and Image Processing (2006)
21. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(5), 564–577 (2003)
22. Kullback, S., Leibler, R.: On information and sufficiency. Annals of Mathematical Statistics 22, 79–86 (1951)
23. Lancaster, H.O.: The chi-squared distribution. Biometrics 27, 238–241 (1971)
24. Rubner, Y., Tomasi, C., Guibas, L.J.: A metric for distributions with applications to image databases. In: IEEE International Conference on Computer Vision (1998)