Improvements in Joint Domain-Range Modeling for Background Subtraction

Manjunath Narayana
narayana@cs.umass.edu
Allen Hanson
hanson@cs.umass.edu
Erik Learned-Miller
elm@cs.umass.edu

University of Massachusetts Amherst
Massachusetts, USA

Abstract

In many algorithms for background modeling, a distribution over feature values is modeled at each pixel. These models, however, do not account for the dependencies that may exist among nearby pixels. The joint domain-range kernel density estimate (KDE) model by Sheikh and Shah [7], which is not a pixel-wise model, represents the background and foreground processes by combining the three color dimensions and two spatial dimensions into a five-dimensional joint space. The Sheikh and Shah model, as we will show, has a peculiar dependence on the size of the image. In contrast, we build three-dimensional color distributions *at each pixel* and allow neighboring pixels to influence each other's distributions. Our model is easy to interpret, does not exhibit the dependency on image size, and results in higher accuracy. Also, unlike Sheikh and Shah, we build an explicit model of the prior probability of the background and the foreground at each pixel. Finally, we use the adaptive kernel variance method of Narayana *et al.* [5] to adapt the KDE covariance at each pixel. With a simpler and more intuitive model, we can better interpret and visualize the effects of the adaptive kernel variance method, while achieving accuracy comparable to state-of-the-art on a standard backgrounding benchmark.

1 Introduction

Background subtraction, often a first step in segmenting moving objects in videos, is most commonly achieved by modeling the background color likelihoods at each pixel. Stauffer and Grimson [8] use a parametric Gaussian mixture model to estimate the likelihoods at each pixel. A non-parametric model was introduced by Elgammal *et al.* [1], where the likelihoods at each pixel are modeled using a kernel density estimate (KDE) by using the data samples from previous frames in history. These pixel-wise models do not allow for the observations at one location to influence the estimated distribution at a different but nearby location. By including each pixel's position information and modeling the likelihoods using a five-dimensional distribution in a joint domain-range representation, Sheikh and Shah [7] allow pixels in one location to influence the distributions in another location. They show that this sharing of spatial information leads to more accurate background subtraction. Their background model is a *single* distribution in the joint domain-range space. As we will see later, their classification criterion, based on the ratio of likelihoods in this five-dimensional space, has an undesirable dependence on the size of the image. Like Sheikh and Shah, we model the foreground and background likelihoods by a KDE using pixel samples from previous video frames. However, we model the processes using a three-dimensional color distribution at each pixel. Our distributions are conditioned on spatial location, rather than being joint distributions over position and color. Our modeling avoids the dependence on the image size and yields better results.

Recent work on kernel estimate based background modeling by Narayana *et al.* [5] has shown that adapting the kernel variance values for each pixel yields significantly better results than using a uniform kernel variance for all pixels. We use a similar approach for adapting the kernel variance at each pixel and show through both synthetic and real data examples that such a scheme is useful. The improvements we present over their approach are the separation of the foreground process into 'previously seen' and 'previously unseen' foreground processes and the use of explicit spatial priors for the three processes - background (bg), previously seen foreground (fg), and previously unseen foreground (fu). Our probabilistic formulation with likelihoods and a spatially dependent prior for each process leads to a posterior distribution over the processes.

Texture features like local binary and ternary patterns are robust to lighting changes and can be useful in background subtraction [2, 4]. A hybrid approach using both color and texture features combines the strengths of both feature spaces [5, 9]. In our system, we combine color features with scale invariant local ternary patterns (SILTP) [4], a very effective texture representation.

Benchmark comparisons on a standard data set show that our system's performance is comparable to the results of Narayana *et al.*, which are the best reported results on our chosen benchmark. The advantage of our model over that of Narayana *et al.*, is that our probabilistic model is more intuitive. The results from our model can be understood more clearly and the various constants and factors in the model can be interpreted more meaningfully.

2 Joint domain-range KDE

Sheikh and Shah [7] model the spatial dependencies between observed intensities by proposing a joint domain-range representation of the pixels. The background and foreground processes are modeled with non-parametric density estimation using samples from previous frames. Each pixel a is represented as a five-dimensional vector, $a = [a_x, a_y, a_r, a_g, a_b]$. The background sample set B consists of the set of pixels that were classified as background in the previous frames in the video, $B = \{b_i : i \in [1 : n_B]\}$, where n_B is the number of pixels in the background set.

Using a KDE, the likelihood of the pixel under the Sheikh and Shah background model [7] is

$$P(a|\mathtt{bg};\Sigma^{\mathrm{B}}) = \frac{1}{\mathrm{n_B}} \sum_{\mathrm{i=1}}^{\mathrm{n_B}} \mathrm{G}(\mathrm{a} - \mathrm{b_i}; \Sigma^{\mathrm{B}}), \tag{1}$$

where $G(x;\Sigma^B)$ is a five-dimensional Gaussian with zero mean and diagonal covariance Σ^B. In our model, since we allow pixel samples to contribute probabilistically to the KDE based on the samples' probability of belonging to the background, we have the following modification of the Sheikh and Shah model:

$$P(a|\mathtt{bg};\Sigma^{\mathrm{B}}) = \frac{1}{\sum_{\mathrm{i}=1}^{\mathrm{n_B}} \mathrm{P}(\mathtt{bg}|\mathrm{b_i})} \sum_{\mathrm{i}=1}^{\mathrm{n_B}} \mathrm{G}(\mathrm{a}-\mathrm{b_i};\Sigma^{\mathrm{B}})\mathrm{P}(\mathtt{bg}|\mathrm{b_i}), \tag{2}$$

where $P(\mathtt{bg}|\mathrm{b_i})$ is the probability that pixel b_i is background.

Our background model is similar to the described model, with the distinction that at each pixel, we model a three-dimensional color distribution *conditioned* on the pixel's spatial location. Denoting the color of the pixel a by $a_{\mathbf{c}}$ and the position by $a_{\mathbf{x}}$, we have

$$P(a_{\mathbf{c}}|\mathtt{bg},\mathrm{a}_{\mathbf{x}};\Sigma^{\mathrm{B}}) = \frac{1}{\mathrm{K}_{\mathtt{bg}}} \sum_{\mathrm{i}=1}^{\mathrm{n_B}} \mathrm{G}(\mathrm{a}_{\mathbf{c}}-\mathrm{b}_{\mathrm{i}\mathbf{c}};\Sigma^{\mathrm{B}}_{\mathtt{range}})\mathrm{G}(\mathrm{a}_{\mathbf{x}}-\mathrm{b}_{\mathrm{i}\mathbf{x}};\Sigma^{\mathrm{B}}_{\mathtt{domain}})\mathrm{P}(\mathtt{bg}|\mathrm{b_i}), \tag{3}$$

where $a_{\mathbf{c}} = [a_r, a_g, a_b]$, and $b_{i\mathbf{c}} = [b_{ir}, b_{ig}, b_{ib}]$ are the color values for pixel a and background pixel b_i respectively. Similarly, $a_{\mathbf{x}} = [a_x, a_y]$ and $b_{i\mathbf{x}} = [b_{ix}, b_{iy}]$ are the position values of the pixels. The matrix Σ^B is separated into its color (range) and position (domain) covariance matrices $\Sigma^B_{\mathtt{range}}$ and $\Sigma^B_{\mathtt{domain}}$. $K_{\mathtt{bg}}$ is the appropriate normalization factor that we explain next.

2.1 Normalization of the kernel estimates

Considering that each background sample b_i contributes probabilistically to the background likelihood depending on its probability of being background and its distance in position from the location $a_{\mathbf{x}}$ at which the likelihood is being computed, we have the following normalization factor:

$$K_{\mathtt{bg}} = \sum_{i=1}^{n_B} G(a_{\mathbf{x}} - b_{i\mathbf{x}};\Sigma^B_{\mathtt{domain}})P(\mathtt{bg}|\mathrm{b_i}). \tag{4}$$

By changing the normalization constant in the KDE, we achieve a likelihood model that is specific to each location in the image. We have a pixel-wise model, but the likelihood at any given pixel location is affected by pixel samples from its spatial neighborhood.

Many other existing methods [1, 8] also model distributions at each pixel location. Recently, in the tracking literature, Sevilla and Learned-Miller [6] have used the term "Distribution Fields" for models that use a distribution at each pixel location, and in which each distribution is estimated from a local neighborhood. Because we estimate the distributions from local neighborhoods in a similar manner, we refer to our method as "distribution field backgrounding" (DFB).

2.2 Foreground and new object models

In the Sheikh and Shah system, along with the background process model, an explicit foreground model is maintained with pixel samples that have been classified as foreground in the previous frames. Similar to Equation 3, our foreground likelihood is

$$P(a_{\mathbf{c}}|\mathtt{fg},\mathrm{a}_{\mathbf{x}};\Sigma^{\mathrm{F}}) = \frac{1}{\mathrm{K}_{\mathtt{fg}}} \sum_{\mathrm{i}=1}^{\mathrm{n_F}} \mathrm{G}(\mathrm{a}_{\mathbf{c}}-\mathrm{f}_{\mathrm{i}\mathbf{c}};\Sigma^{\mathrm{F}}_{\mathtt{range}})\mathrm{G}(\mathrm{a}_{\mathbf{x}}-\mathrm{f}_{\mathrm{i}\mathbf{x}};\Sigma^{\mathrm{F}}_{\mathtt{domain}})\mathrm{P}(\mathtt{fg}|\mathrm{f_i}). \tag{5}$$

The foreground sample set F consists of the foreground pixels from previous frames, $F = \{f_i : i \in [1 : n_F]\}$, where n_F is the number of pixels in the foreground set, Σ^F is the covariance matrix for the foreground model, and $K_{\mathtt{fg}}$ is the normalization factor, analogous to $K_{\mathtt{bg}}$. For efficiency, we compute the likelihoods by considering only samples that lie close enough to $a_{\mathbf{x}}$ to have significant probability density.

To account for the emergence of new objects in the scene, Sheikh and Shah add a uniform distribution to the foreground likelihood. If $P_{KDE}(a|\mathtt{fg})$ is the foreground likelihood obtained using a KDE from previous samples, the Sheikh and Shah foreground likelihood is

$$P(a|\mathtt{fg}) = \alpha \times \gamma + (1-\alpha) \times \mathrm{P_{KDE}(a|\mathtt{fg})}. \tag{6}$$

γ is a uniform distribution with magnitude $\frac{1}{R\times G\times B\times X\times Y}$, where R, G, and B, are the number of possible intensities for red, green, and blue colors respectively, X and Y are the number of columns and rows in the image, and $0 \leq \alpha \leq 1$ is a mixing factor.

The above uniform factor has a peculiar effect. When the size of an image is changed, the uniform likelihood of observing a new foreground object changes. Increasing the size of the image reduces the likelihood of seeing foreground objects. In contrast, we account for appearance of new colors in the scene by placing a uniform distribution over color space at each pixel location in the image, thus avoiding the above effect. Our previously unseen foreground likelihood, $P(\mathbf{c}|\mathtt{fu},\mathbf{x})$, which models “new objects”, has a magnitude $\frac{1}{R\times G\times B}$ for all locations $\mathbf{x} = (x,y)$ in the image.

2.3 Classification

In the Sheikh and Shah model, the classification of pixels is done based on the likelihood ratios of the background and foreground processes. The decision criterion based on the likelihood ratios of the five-dimensional likelihoods can be represented as

$$\begin{aligned} P(a_{\mathbf{c}},a_{\mathbf{x}}|\mathtt{bg}) &\overset{?}{\gtrless} P(a_{\mathbf{c}},a_{\mathbf{x}}|\mathtt{fg}) \\ P(a_{\mathbf{c}}|a_{\mathbf{x}},\mathtt{bg}) \times \mathrm{P}(\mathrm{a}_{\mathbf{x}}|\mathtt{bg}) &\overset{?}{\gtrless} P(a_{\mathbf{c}}|a_{\mathbf{x}},\mathtt{fg}) \times \mathrm{P}(\mathrm{a}_{\mathbf{x}}|\mathtt{fg}). \end{aligned} \tag{7}$$

The classification decision hence depends on the factors $P(a_{\mathbf{x}}|\mathtt{bg})$ and $P(a_{\mathbf{x}}|\mathtt{fg})$. These factors are the prior probability of a particular pixel location given the background or foreground label. For any pixel location $a_{\mathbf{x}}$, these factors can depend upon parts of the image that are arbitrarily far away. This is because the prior likelihood of a given pixel location being foreground will be smaller if more pixels from another part of the image are detected as foreground, and larger if fewer pixels elsewhere are detected as foreground (since $P(a_{\mathbf{x}}|\mathtt{fg})$ must integrate to 1). Furthermore, these factors will change when the image size is changed, hence affecting the classification. Our model avoids this drawback of the Sheikh and Shah model by computing the posterior probability of the background label conditioned on the pixel location. That is, our model does not have this arbitrary dependence on the size of the image.

2.4 Location-specific priors for background and foreground processes

We define suitable priors for the three processes involved - background, previously seen foreground, and unseen foreground. The classified pixel labels from the previous frame can be used as a starting point for building the priors for the current frame. We assume that a pixel that is classified as background in the previous frame has a 95% probability of being background in the current frame as well. The pixel has a 2.5% probability of being one of the seen foreground objects, and a 2.5% probability of coming from a new foreground object. For a foreground pixel in the previous frame, we assume that due to object motion, there is a 50% probability of this pixel becoming background, a 25% probability of this pixel belonging to the same foreground object as in the previous frame, and a 25% probability that it becomes a new object. Note that the use of the term “new object” does not necessarily mean that a new unseen object has appeared at that pixel location, but simply that the color being observed at the pixel is not explained well by either the existing foreground or background colors in the vicinity of the pixel and that a uniform color distribution best explains the color. The reason a pixel may be classified as a “new object” could be that the object position in the image has changed significantly compared to the previous frame or that the object color appearance has changed due to motion.

The above scheme for informative priors is applied in a soft manner. For instance, a pixel that has probability p of being background in the previous frame will have a background prior equal to $p \times .95 + (1-p) \times .5$. Also, since objects typically move by a few pixels from the previous frame to the current frame, we apply a smoothing (7×7 Gaussian filter with a standard deviation value of 1.75) to the classification results from the previous frame before computing the priors for the current frame. Let $\tilde{P}_{t-1}(\mathtt{bg})$ be the smoothed background posterior image from the previous frame. The priors for the current frame are

$$\begin{aligned} P(\mathtt{bg}|\mathbf{x}) =& \tilde{P}_{t-1}(\mathtt{bg}|\mathbf{x}) \times .950 + (1 - \tilde{\mathrm{P}}_{\mathrm{t-1}}(\mathtt{bg}|\mathbf{x})) \times .500 \\ P(\mathtt{fg}|\mathbf{x}) =& \tilde{P}_{t-1}(\mathtt{bg}|\mathbf{x}) \times .025 + (1 - \tilde{\mathrm{P}}_{\mathrm{t-1}}(\mathtt{bg}|\mathbf{x})) \times .250 \\ P(\mathtt{fu}|\mathbf{x}) =& \tilde{P}_{t-1}(\mathtt{bg}|\mathbf{x}) \times .025 + (1 - \tilde{\mathrm{P}}_{\mathrm{t-1}}(\mathtt{bg}|\mathbf{x})) \times .250. \end{aligned} \tag{8}$$

The posterior probability for the background label given a pixel a is

$$\begin{aligned} P(\mathtt{bg}|\mathrm{a}) =& \frac{\mathrm{P}(\mathrm{a}_{\mathbf{c}}|\mathtt{bg}, \mathrm{a}_{\mathbf{x}}; \Sigma^{\mathrm{B}}) \times \mathrm{P}(\mathtt{bg}|\mathrm{a}_{\mathbf{x}})}{\sum_{\mathrm{l}=\mathtt{bg},\mathtt{fg}} \mathrm{P}(\mathrm{a}_{\mathbf{c}}|\mathrm{l}, \mathrm{a}_{\mathbf{x}}; \Sigma^{\mathrm{l}}) \times \mathrm{P}(\mathrm{l}|\mathrm{a}_{\mathbf{x}}) + \mathrm{P}(\mathrm{a}_{\mathbf{c}}|\mathtt{fu}, \mathrm{a}_{\mathbf{x}}) \times \mathrm{P}(\mathtt{fu}|\mathrm{a}_{\mathbf{x}})} \\ P(\mathtt{fg}|\mathrm{a}) =& 1 - \mathrm{P}(\mathtt{bg}|\mathrm{a}). \end{aligned} \tag{9}$$

Note that mixing a uniform distribution to the foreground likelihood with a factor α to account for new objects, as done by Sheikh and Shah, is equivalent to our method of treating the existing foreground objects and new foreground objects as separate processes with their own likelihoods and priors. However, the use of explicit priors allows our system to be extended more easily. For instance, at the image boundary regions, we can use a higher prior for unseen foreground to model the higher probability of new objects entering the scene from outside the camera’s field of view.

3 Adaptive kernels for background modeling

Narayana *et al.* [5] showed recently that in kernel estimate based background modeling, using an adaptive pixel-wise kernel improves results significantly. For each pixel location, for the background model, a set of variance values for both spatial and color dimensions is tried and the configuration that results in the highest likelihood is chosen for that particular pixel. Mathematically, the likelihood and normalization factor Equations 3 and 4 now include a location-specific covariance matrix:

$$
\begin{aligned}
P(a_{\mathbf{c}}|\mathtt{bg},\mathrm{a}_{\mathbf{x}};\Sigma^{\mathrm{B}_{\mathrm{a}_{\mathbf{x}}}}_{\mathtt{range}},\Sigma^{\mathrm{B}_{\mathrm{a}_{\mathbf{x}}}}_{\mathtt{domain}})&=\frac{1}{\mathrm{K}_{\mathtt{bg}}}\sum_{\mathrm{i}=1}^{\mathrm{n}_{\mathrm{B}}}\mathrm{G}(\mathrm{a}_{\mathbf{c}}-\mathrm{b}_{\mathrm{i}\mathbf{c}};\Sigma^{\mathrm{B}_{\mathrm{a}_{\mathbf{x}}}}_{\mathtt{range}})\mathrm{G}(\mathrm{a}_{\mathbf{x}}-\mathrm{b}_{\mathrm{i}\mathbf{x}};\Sigma^{\mathrm{B}_{\mathrm{a}_{\mathbf{x}}}}_{\mathtt{domain}})\mathrm{P}(\mathtt{bg}|\mathrm{b}_{\mathrm{i}}),\\
K_{\mathtt{bg}}&=\sum_{i=1}^{n_B}G(a_{\mathbf{x}}-b_{i\mathbf{x}};\Sigma^{B_{a_{\mathbf{x}}}}_{\mathtt{domain}})P(\mathtt{bg}|\mathrm{b}_{\mathrm{i}}),
\end{aligned}
\tag{10}
$$

where $\Sigma^{B_{a_{\mathbf{x}}}}_{\mathtt{domain}}$ and $\Sigma^{B_{a_{\mathbf{x}}}}_{\mathtt{range}}$ represent the location-specific spatial and color dimension variances at location $a_{\mathbf{x}}$. For each pixel location $a_{\mathbf{x}}$, the optimal variance for the background process is selected by maximizing the likelihood of the background at pixel a under different variance values:

$$
\{\sigma^{B^*_{a_{\mathbf{x}}}}_{\mathtt{domain}},\sigma^{B^*_{a_{\mathbf{x}}}}_{\mathtt{range}}\}=\underset{\sigma^{B_{a_{\mathbf{x}}}}_{\mathtt{domain}},\sigma^{B_{a_{\mathbf{x}}}}_{\mathtt{range}}}{\arg\max}\,P(a_{\mathbf{c}}|\mathtt{bg},\mathrm{a}_{\mathbf{x}};\Sigma^{\mathrm{B}_{\mathrm{a}_{\mathbf{x}}}}_{\mathtt{range}},\Sigma^{\mathrm{B}_{\mathrm{a}_{\mathbf{x}}}}_{\mathtt{domain}}).
\tag{11}
$$

Here, $\sigma^{B_{a_{\mathbf{x}}}}_{\mathtt{domain}}\in R^{B}_{\mathtt{domain}}$ and $\sigma^{B_{a_{\mathbf{x}}}}_{\mathtt{range}}\in R^{B}_{\mathtt{range}}$. $R^{B}_{\mathtt{domain}}$ and $R^{B}_{\mathtt{range}}$ represent the set of spatial and color dimension variances from which to choose the optimal variance. These constitute the diagonal elements in the covariance matrices $\Sigma^{B_{a_{\mathbf{x}}}}_{\mathtt{domain}}$ and $\Sigma^{B_{a_{\mathbf{x}}}}_{\mathtt{range}}$.

4 Results

We present results on I2R videos [3], a standard data set with nine videos taken in different settings. The videos have several challenging features like moving leaves and waves, strong object shadows, and moving objects becoming stationary for a long duration. Each video has 20 frames for which the ground truth has been marked. We use the F-measure to judge accuracy [4]. As done by Narayana *et al.*, we use a Markov random field (MRF) to post-process the labels and also discard any foreground detections smaller than 15 pixels in size.

In order to study the effects of the different normalizations of Equations 2 and 4, which represent the difference between the Sheikh and Shah model and ours, we present results of our background classification system for both normalization procedures. Table 1 shows the results comparing the two normalization schemes for different settings of spatial and color covariances in our system. The columns are labeled ‘a’ and ‘b’ referring to the Sheikh and Shah model and our model respectively. The table shows that using a pixel-wise model results in a higher accuracy than the Sheikh and Shah model for all parameter settings except one.

Table 1 also shows that using the adaptive kernel variance for the background improves accuracy in our model. Interestingly, the adaptive kernel variance reduces the accuracy when using the Sheikh and Shah normalization. Figure 1 shows images that characterize the performance of the two normalization schemes and the effect of using adaptive kernel variance in both schemes (corresponding to the columns 3a, 3b, 4a, and 4b in Table 1). We see that the Sheikh and Shah normalization (Figure 1b) causes many foreground pixels to be misclassified as background. This is because Equation 2 is biased towards whichever process has a smaller spatial neighborhood - the background process in this case. If the neighborhood is large, pixel samples that are spatially far away contribute little to the numerator, but heavily to the denominator. Figure 2 illustrates this phenomenon with a synthetic example. Consider that a red foreground object was present in front of a pink background in the previous frame and that the foreground pixel samples from this image are used to compute the foreground

Column num	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
4*$\sigma^{B}_{\texttt{domain}}$ $\longrightarrow$	3	3	3	3	3	3	[1 3]	[1 3]
4*$\sigma^{B}_{\texttt{range}}$ $\longrightarrow$	15	15	45	45	45	45	[5 15 45]	[5 15 45]
4*$\sigma^{F}_{\texttt{domain}}$ $\longrightarrow$	3	3	3	3	12	12	[12]	[12]
4*$\sigma^{F}_{\texttt{range}}$ $\longrightarrow$	15	15	45	45	45	45	[15]	[15]
AirportHall	48.91	**53.64**	65.70	**66.37**	**70.13**	67.95	65.52	**68.28**
Bootstrap	56.16	**58.90**	65.32	**66.96**	**71.77**	69.17	71.38	**71.86**
Curtain	49.79	**49.96**	69.55	**71.22**	**87.34**	85.66	79.76	**93.57**
Escalator	25.10	**35.32**	42.54	**53.01**	53.70	**54.01**	54.02	**66.37**
Fountain	52.07	**56.02**	58.84	**59.00**	57.35	**77.11**	49.89	**77.43**
ShoppingMall	58.97	**62.67**	67.23	**70.28**	**74.12**	70.95	74.43	**76.46**
Lobby	**23.90**	23.27	**23.56**	22.55	**27.88**	21.64	**33.34**	13.24
Trees	48.22	**62.35**	75.22	**78.35**	**85.80**	82.61	**85.57**	83.88
WaterSurface	46.61	**46.78**	**57.76**	55.63	**78.16**	**75.80**	64.03	**93.81**
Average	45.53	**49.88**	58.41	**60.37**	**67.62**	67.21	64.22	**71.66**

Table 1: *F-measure* for Sheikh and Shah model (left) and our model (right) for different kernel variances. Columns 4a and 4b correspond to the adaptive variance procedure with variance values given in brackets. Bold entries correspond to the higher of the two values for a given parameter setting. For all but one parameter setting, the pixel-wise model (ours) is more accurate than the Sheikh and Shah model for most videos. Blue entries correspond to the best F-measure for each video. Using our model with adaptive variance results in the highest accuracy for most videos.

likelihoods at each pixel in the current frame. For simplicity, we consider the case where the object has not moved from the previous frame to the current. Applying the Sheikh and Shah normalization scheme, we see that as the size of the neighborhood for foreground samples is increased from 1 to 3, the likelihood values for the foreground pixels decrease dramatically (compare Figures 2c and d). Using our normalization method, the dependence between the spatial neighborhood and likelihood values is eliminated (Figures 2e and f).

In our normalization scheme, since the denominator term weights each sample's probability by its spatial distance, the likelihood equations 3 and 5 are not inherently biased based on the spatial neighborhood of the processes. However, the undesirable effect of our normalization is that there are more false positive foreground classifications (Figure 1c, row 1) and once a region is classified as foreground, it tends to remain classified as foreground (1 c, row 3). In light of this observation, it is obvious why using the adaptive kernel method for the background process helps our model (Figure 1e). The adaptive kernel for the background process, by selecting the best of the available kernel variances, in effect "tries hard" to classify each pixel as background. For instance, background pixels that have been occluded by a foreground object for many frames can correctly be recovered when the object moves away by matching the revealed pixel colors to nearby background locations using a larger spatial kernel. In effect, the kernel adaptation allows the background to "spread" back into a region that has recently been foreground. When a pixel is not well explained by the background model despite the selection procedure, it gets labeled as foreground. In contrast, the adaptive procedure hurts the performance of the Sheikh and Shah model because it tends to further bias the decision towards the background label (Figure 1d).

With our probabilistic model, we can interpret the effect of the adaptive kernel variance method of Narayana *et al*. more easily in Figures 3 and 4. Consider a synthetic scene with no foreground objects, but in which the colors in the central greenish part of the background have been displaced at random by one or two pixel locations to simulate spatial uncertainty.

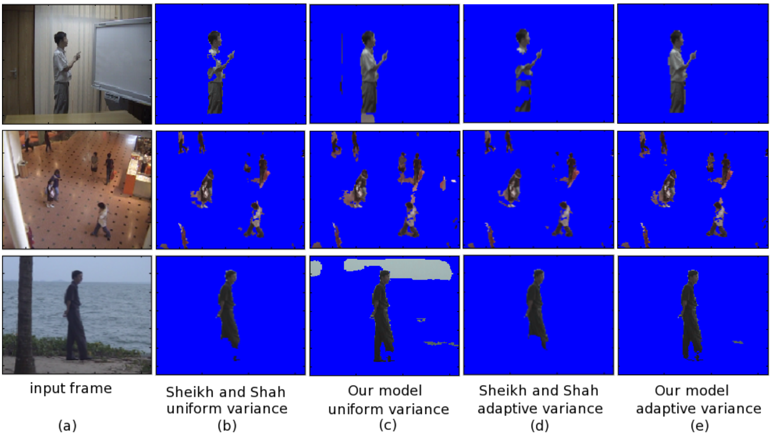
input frame
(a)
Sheikh and Shah uniform variance
(b)
Our model uniform variance
(c)
Sheikh and Shah adaptive variance
(d)
Our model adaptive variance
(e)

Figure 1: Comparing the effect of adaptive kernel variance for the Sheikh and Shah normalization versus our normalization method. The Sheikh and Shah method has a bias towards background label (b), further exacerbated by the adaptive kernel selection (d). Our method tends to classify foreground objects well, but has more false positive foreground pixels (c). Adaptive kernel variance with our normalization yields best results (e).

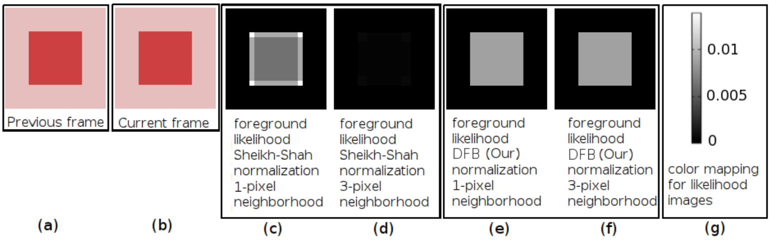
Previous frame
Current frame
foreground likelihood Sheikh-Shah normalization 1-pixel neighborhood
foreground likelihood Sheikh-Shah normalization 3-pixel neighborhood
foreground likelihood DFB (Our) normalization 1-pixel neighborhood
foreground likelihood DFB (Our) normalization 3-pixel neighborhood
0.01
0.005
0
color mapping for likelihood images
(a)
(b)
(c)
(d)
(e)
(f)
(g)

Figure 2: Sheikh and Shah normalization equation leads to a dependency between neighborhood size and likelihood values. Our normalization does not.

As shown in Figure 3, the adaptive kernel variance method models the scene better by applying a high spatial variance for pixels that have moved and a low spatial variance for pixels that have not moved. Similarly, for color variance, Figure 4 shows the resulting likelihoods when uniformly sampled noise is added to the color values in the central part of the image. A small color variance value results in low likelihoods for pixels whose colors have changed. A large color variance results in low likelihoods for pixels that have not changed. The adaptive kernel variance method performs well in both kinds of pixels.

Table 2 shows F-measure values for different methods on the I2R data set. The mixture of Gaussians (MoG) method [8] is a commonly used baseline method. Scale invariant local ternary patterns (SILTP) [4] are effective texture features that are robust to lighting changes in the scene. The variable kernel score (VKS) method [5] uses an adaptive kernel variance method for each pixel location in the image and achieves the best results reported on this data set. Following the same scheme as Narayana *et al.*, we present results from two versions of our system - one called DFB-rgb using RGB color features only and another called

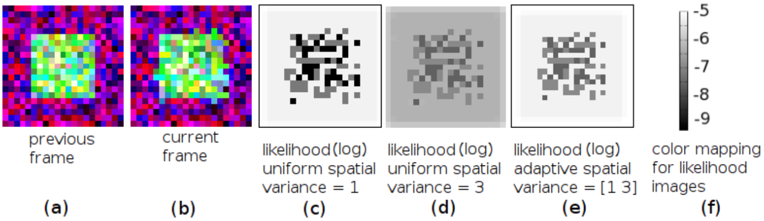

Figure 3: (a) and (b) Spatial uncertainty in the central part of the background. (c) Small uniform variance results in low likelihoods for pixels that have moved. (d) Large uniform variance results in higher likelihoods of the moved pixels at the expense of lowering the likelihoods of stationary pixels. (e) Adaptive variance results in high likelihoods for both the moved and stationary pixels.

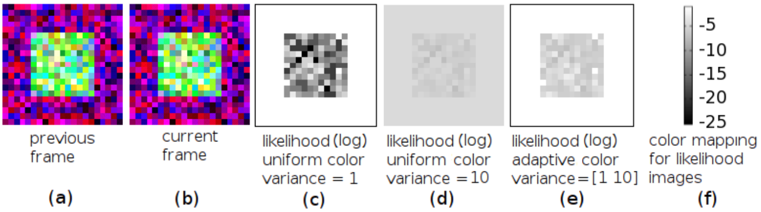

Figure 4: Color uncertainty in the central part of the background is best modeled by using adaptive kernel variances. (c) Small uniform variance results in low likelihoods for pixels that have changed color. (d) Large uniform variance results in higher likelihoods of the altered pixels at the expense of lowering the likelihoods of other pixels. (e) Adaptive variance results in high likelihoods for both kinds of pixels.

DFB-lab+siltp using a hybrid of LAB color space and SILTP texture (computed at 3 scales) features. The parameters used for the variable kernel method are the same as those used in Narayana *et al*. We see that our DFB method performs better than SILTP on most videos and is comparable to the accuracy of VKS. However, DFB has the advantage of being a simpler model that can be better understood in probabilistic terms.

5 Conclusions

We have highlighted some of the issues with modeling the background using a single joint domain-range distribution as done by Sheikh and Shah. By instead extending the domain-range representation to model distributions conditioned on each pixel location, we address these issues. In addition, we incorporate spatial priors for the background and foreground processes by using classification labels from the previous frame. Our background and foreground likelihood models are conceptually easier to interpret than the foreground and background scores of Narayana *et al*. Our model's accuracy is comparable to theirs and also explains better the effect of using the adaptive kernel variance for each pixel location. For future work, the foreground priors can be modeled more accurately by including the objects' tracking information. Integrating the background modeling with object tracking may result

Video	MoG	SILTP [4]	VKS	VKS	DFB	DFB
			rgb	lab+siltp	rgb	lab+siltp
AirportHall	57.86	68.02	70.44	**71.28**	68.28	70.75
Bootstrap	54.07	72.90	71.25	76.89	71.86	**77.64**
Curtain	50.53	92.40	**94.11**	94.07	93.57	94.07
Escalator	36.64	**68.66**	48.61	49.43	66.37	49.99
Fountain	77.85	85.04	75.84	**85.97**	77.43	85.88
ShoppingMall	66.95	79.65	76.48	**83.03**	76.46	82.64
Lobby	68.42	**79.21**	18.00	60.82	13.24	62.60
Trees	55.37	67.83	82.09	**87.85**	83.88	87.64
WaterSurface	63.52	83.15	94.83	92.61	**93.81**	93.79

Table 2: *F-measure* on I2R data. DFB is better than SILTP and comparable to VKS.

in significant improvement in the accuracy of the classification.

6 Acknowledgements

This work was supported in part by the National Science Foundation under CAREER award IIS-0546666 and grant CNS-0619337. Any opinions, findings, conclusions, or recommendations expressed here are the authors' and do not necessarily reflect those of the sponsors.

References

[1] Ahmed M. Elgammal, David Harwood, and Larry S. Davis. Non-parametric model for background subtraction. In *European Conference on Computer Vision*, pages 751–767, 2000.

[2] Marko Heikkilä and Matti Pietikäinen. A texture-based method for modeling the background and detecting moving objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(4):657 –662, 2006.

[3] Liyuan Li, Weimin Huang, Irene Y. H. Gu, and Qi Tian. Foreground object detection from videos containing complex background. In *ACM International Conference on Multimedia*, pages 2–10, 2003.

[4] Shengcai Liao, Guoying Zhao, Vili Kellokumpu, Matti Pietikäinen, and Stan Z. Li. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 1301 –1306, 2010.

[5] Manjunath Narayana, Allen Hanson, and Erik Learned-Miller. Background modeling using adaptive pixelwise kernel variances in a hybrid feature space. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2012.

[6] Laura Sevilla-Lara and Erik Learned-Miller. Distribution fields for tracking. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2012.

[7] Yaser Sheikh and Mubarak Shah. Bayesian modeling of dynamic scenes for object detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27:1778–1792, 2005.

[8] Chris Stauffer and W. Eric L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, volume 2, pages 246–252, 1999.

[9] Jian Yao and Jean-Marc Odobez. Multi-layer background subtraction based on color and texture. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 1–8, 2007.