Universal Multimode Background Subtraction

Hasan Sajid, *Member, IEEE*, and Sen-Ching Samson Cheung, *Senior Member, IEEE*

***Abstract*— In this paper, we present a complete change detection system named multimode background subtraction. The universal nature of system allows it to robustly handle multitude of challenges associated with video change detection, such as illumination changes, dynamic background, camera jitter, and moving camera. The system comprises multiple innovative mechanisms in background modeling, model update, pixel classification, and the use of multiple color spaces. The system first creates multiple background models of the scene followed by an initial foreground/background probability estimation for each pixel. Next, the image pixels are merged together to form mega-pixels, which are used to spatially denoise the initial probability estimates to generate binary masks for both RGB and YCbCr color spaces. The masks generated after processing these input images are then combined to separate foreground pixels from the background. Comprehensive evaluation of the proposed approach on publicly available test sequences from the CDnet and the ESI data sets shows superiority in the performance of our system over other state-of-the-art algorithms.**

***Index Terms*— Computer vision, change detection, background model bank, background subtraction, color spaces, binary classifiers, foreground segmentation, pixel classification.**

I. Introduction

VIDEO change detection or Background Subtraction (BS) is one of the most widely studied topics in computer vision. It is a basic pre-processing step in video processing and therefore has numerous applications including video surveillance, traffic monitoring, human detection, gesture recognition, etc. Typically, a BS process produces a foreground (FG) binary mask given an input image and a background (BG) model.

BS is a difficult problem because of the diversity in background scenes and the changes originated from the camera itself. Scene variations can be in many forms such as, to name just a few, dynamic background, illumination changes, intermittent object motion, shadows, highlights, camouflage as well as a multitude of environmental conditions like rain, snow, and change in sunlight [1]. Likewise, the changes linked to camera can be due to auto-iris, camera jitter, sensor noise and pan-tilt-zoom. Existing state-of-the-art techniques can address only a subset of these challenges and most of them are sensitive to illumination changes, camera/background motion and environmental conditions [2], [3]. No single technique exists that is able to simultaneously handle all key challenges and produce satisfactory results.

In this paper, we propose a BS system that is robust against various challenges associated with real world videos. The proposed approach uses a Background Model Bank (BMB) that comprises of multiple Background (BG) models of the scene. To separate foreground pixels from changing background pixels caused by scene variations or camera itself, we apply Mega-Pixel (MP) based spatial denoising to pixel level probability estimates on different color spaces to obtain multiple Foreground (FG) masks. They are then combined to produce a final output FG mask. The major contribution of this paper is a universal background subtraction system called Multimode Background Subtraction (MBS) with following major innovations: Background Model Bank (BMB), model update mechanism, MP-based spatial denoising of pixel-based probability estimates, fusion of multiple binary masks, and use of multiple color spaces for BS process. Preliminary results of using our system to handle illumination changes and camera movements were presented in [4] and [5] respectively. Improvements upon these prior works include:

- a detailed analysis of the fusion of appropriate color spaces for BS,
- a novel model update mechanism, and
- a novel MP-based spatial denoising and a dynamic model selection scheme that significantly reduces the number of parameters and improve computational speed.

BS is well-researched topics in computer vision, therefore, we demonstrate the performance of MBS by providing a comprehensive comparison with 15 other state-of-the-art BS algorithms on a set of publicly-available challenging sequences across 12 different categories, totalling to 56 video sets. To avoid bias in our evaluations, we have adopted the same sets of metrics as recommended by the CDnet 2014 [2]. The extensive evaluation of our system demonstrates better foreground segmentation and superiority of our system in comparison with existing state-of-the-art approaches.

The rest of paper is organized as follows. Relevant work is discussed in Section II. We present and discuss our contributions in Section III and overall system in section IV, followed by experiments and result comparison in Section V and conclude the paper in Section VI.

II. Related Work

There are a plethora of BS techniques, many of which reviewed in surveys like [6], [7], and [8]. We can broadly divide these into four categories: pixel-based, region-based, frame-based and learning based [9].

Manuscript received June 8, 2015; revised February 19, 2016 and November 8, 2016; accepted April 11, 2017. Date of publication April 19, 2017; date of current version May 9, 2017. This work was supported by the National Science Foundation under Grant 1237134. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hamid Rahim Sheikh. *(Corresponding author: Hasan Sajid.)*

H. Sajid was with the University of Kentucky, Lexington, KY 40506 USA. He is now with the National University of Sciences and Technology, Islamabad 44000, Pakistan (e-mail: hasan.sajid@smme.nust.edu.pk).

S.-C. S. Cheung is with the University of Kentucky, Lexington, KY USA (e-mail: sccheung@ieee.org).

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

Digital Object Identifier 10.1109/TIP.2017.2695882

Pixel-based algorithms form a pixel-wise statistical model of the scene. The algorithms in this category are based on simple statistics from mean, mode, running average to complex multimodal distributions [6], [7]. Although methods relying on simple statistics like unimodal Gaussian methods are very fast and computationally inexpensive, they produce relatively poor segmentation results due to the limited capacity in modelling real world changes such as camera noise, moving background, camera jitter, sudden illumination changes etc. The most popular multimodal techniques in pixel based category are pixel-wise Gaussian Mixture Model (GMM) [10] and Kernel Density Estimates (KDE) [11].

The GMM based techniques model the per-pixel distribution of values observed overtime with a mixture of Gaussians. The multimodal nature of these techniques allow them to cope with dynamic background. GMM has been widely used for different BS systems and various improved versions have been proposed. For example in [12], the authors take advantage of color and texture invariance and combine them with GMM algorithm resulting in a more robust algorithm. However, the improvement has proved to be computationally expensive and unsuitable for real time operation. In [13], instead of fixing the number of components for each pixel authors estimate the appropriate number of components for each pixel dynamically and thus it overcomes the problem of choosing right number of components for each pixel. In [14], the authors further combined motion with pixel-level GMM appearance models. Other improvements in GMM-based techniques are summarized in [8].

Another popular algorithm in this category are based on KDE such as [11] and [15]. For each pixel, these methods accumulate values from pixel's recent history and then estimate the probability distribution of the background values. The distribution is then used to classify whether a pixel belongs to foreground or background. The kernel density estimator helps to overcome two problems inherent in GMM based models; (a) choice of suitable shape for pixel probability distribution function and, (b) constant need for parameter estimation.

Sample consensus is another non-parametric method that relies on recently observed pixels to determine if the new incoming pixel is a FG or BG. SuBSENSE is an example of sample consensus methods that uses pixel-level feedback loop mechanism to continuously update and maintain the pixel's model [16]. A spatiotemporal feature descriptor is also used for increased sensitivity, which however entails high computational costs.

Codebook is another class of techniques that has been reported in [17] and [18]. It comprises of a codebook for each pixel which is a compressed form of background. Each codebook has multiple codewords that are based on a sequence of training images using a color distortion metric. Incoming pixels are matched against all background codewords for classification.

Regardless of the choice of statistical models, pixel-based algorithms in general suffer from a lack of inter-pixel spatial dependencies and the constant need of updating the distribution parameters or model. However, it is difficult to determine an appropriate update rate to differentiate true foreground from drastic background changes such as those caused by sudden variation in illumination or fast moving object.

The second class of techniques are region-based techniques. Unlike their pixel-based counterparts, region-based techniques exploit local spatial relationships among pixels. In [15], the authors enforce spatial context among pixels by incorporating pixel locations into their background and foreground KDEs using a Markov Random Field framework. Another region based method is presented in [19] which uses statistical circular shift moments (SCSM) in image regions for change detection. Although these methods incorporate spatial information, their ability in handling change events at various speeds is questionable - there does not seem to be a rational approach in determining proper time interval for model update.

A different region-based approach, introduced in [20], [21], and [22], models spatial dependencies by considering blocks of different sizes instead of pixels individually. The basic underlying assumption is that the neighbouring pixels undergo similar variation as the pixel itself. The blocks are formed over a sequence of training images, followed by training a Principal Component Analysis (PCA) Model for each spatial block. In [21], classification is done by comparing a block in current frame to its reconstruction from PCA coefficients and declaring it as background if the reconstruction is close. In contrast to [21] and [22] performs classification using threshold based on difference between current image and the back projection of PCA coefficients. PCA-based techniques are more robust against noise and illumination changes in comparison to their pixel based counterparts but lack any update mechanism.

Another region based method named Multiscale Spatio-Temporal uses a three-level spatio-temporal color/luminance Gaussian pyramid BG model for each pixel [23]. While it is robust against dynamic background and shadows, selecting an appropriate update rate is challenging for this method.

Frame-based methods create statistical BG models for the entire frame. Many of the frame-based techniques are based on a shading model, which calculates the ratio of intensities between an input image and the reference frame or BG model [9], [24]. Frame-based techniques have not gained as much as popularity as pixel based approaches but are known to offer more robust solution against gradual as well as sudden illumination changes [8].

Based on the shading model, Pilet *et al.* [25] propose a Statistical Illumination (SI) model that uses GMM to model the distribution of the ratio of intensities. In this method, spatial dependence is incorporated in the framework by learning a spatial-likelihood model. Although this technique is robust against global illumination changes, it is not able to handle local illumination changes [9].

Eigen Background (EB) is a frame-based method that builds an Eigen space over expected illumination changes and reconstructs the BG image by projecting an input image on the learned Eigen space [26]. The performance of EB strongly depends on an ad-hoc threshold and whether the global and local illumination changes can be well represented by a linear combination of background scenes in training set.

Vosters *et al.* present an improved frame-based technique by combining both EB and SI models in [9] at the expense of higher computational cost. EB reconstructs the BG image and then SI model segments the image into FG and BG regions. The authors also improve SI by introducing an online instead of an offline spatial-likelihood model.

Another frame-based technique is Tonal Alignment (TA) [27]. For an input image, it first uses the change detection algorithm in [28] to extract out BG pixels, subset of which are then used for histogram specification transform computation. This transformation tonally aligns the input and background image. FG segmentation is done by pixel-wise comparison between the input and the tonally-aligned background image. TA is able to handle global illumination changes but also fails to deal with local lighting changes. Apart from these, there exist methods such as those in [32] and [33] that take advantage of illumination invariant features such as texture with edge or color. However, they suffer from the possible absence of texture in certain areas of image or poor color discrimination in low lighting conditions.

The fourth class of methods apply traditional machine learning on different features to build the BG model. For example, in [29], the authors combine Haar, color, and gradient features for each pixel in a kernel density framework, and apply SVM for segmentation. Neural network based approaches have also gained popularity in recent years. SC_SOBS [30] models the BG with weights of a neural network, whereas a weightless neural network named CwisarDH is proposed in [31]. It buffers previous FG values to robustly handle intermittent objects. The dependence on training data with positive and negative labels makes these methods impractical for real world deployment.

III. System Innovations

Background Subtraction can be summarized as a five-step process: pre-processing, background modelling, foreground detection, data validation and model update. Pre-processing involves simple image processing on input video such as format conversion and image resizing for subsequent steps. Background modelling is responsible for constructing a statistical model of the scene, followed by pixel classification in the foreground detection step. In the data validation step, falsely-detected foreground pixels are removed to form the final foreground mask [6]. The final step is to update the model if necessary.

Our innovations primarily fall in the use of multiple color spaces, background model bank for background modelling process, MP formation and label correction for foreground detection, and a novel model update procedure. In the following sub-sections, we detail each of these innovations.

A. Multiple Color Spaces for BS

The choice of color space is critical to the accuracy of foreground segmentation. Many different color spaces including RGB, YCbCr, HSV, HSI, lab2000, normalized-RGB (rgb) have been used for background subtraction. Among these color spaces, we focus on the four most widely-used color spaces: RGB, YCbCr, HSV and HSI [34], [35].

RGB is a popular choice for a number of reasons: (a) the brightness and color information are equally distributed in all three color channels; (b) it is robust against both environmental and camera noise [34]; (c) it is the output format of most cameras and its direct usage in BS avoids the computation cost of color conversion [35].

The use of the three other color spaces: YCbCr, HSV and HSI are motivated by human visual system (HVS). The defining color perception in HVS is that it tends to assign a constant color to an object even under changing illumination over time or space [34], [36]. These color spaces segregate the brightness and color information, with YCbCr on Cartesian coordinates whereas HSV and HSI on polar coordinates. While the color constancy makes the BS process more robust against shadow, highlights and illumination changes, the foreground detection is less discriminatory if brightness information is not used [34], [36]–[38].

In comparative studies on color spaces [34], [35], [37], [39], YCbCr has been shown to outperform RGB, HSI and HSV color spaces and is considered to be the most suitable color space for foreground segmentation [34], [35], [37]. Due to its independent color channels, YCbCr is the least sensitive to noise, shadow and illumination changes. RGB is ranked second with HSI and HSV at the bottom as their polar coordinate descriptions are quite prone to noise [34]. The conversion from RGB to YCbCr is also computationally less expensive than to HSI or HSV.

Based on the above comparison, YCbCr is a natural choice for segmentation. However, [36] and [37] also identify potential problems with the YCbCr color space: when current image contains very dark pixels, the chance of misclassification increases since dark pixels are close to the origin in RGB space. The fact that all chromaticity lines in RGB space meet at the origin makes dark pixels close or similar to any chromaticity line. Such scenario does not occur only when illumination levels are low globally, but also happens when portion of the image becomes darker. This is common especially in indoor scenes with complex illumination sources and scene geometry. Shadows casted by objects is one such example. The exclusive use of YCbCr color space in such situations will result in a decrease in foreground segmentation accuracy.

Inspired by the HVS, we propose to use two color spaces: RGB and YCbCr to handle different illumination conditions. We then choose the appropriate channels for the scene in question. This is different from all existing techniques that employ all channels and only one color space. RGB and Y channels are used under poor lighting conditions since chromatic information is uniformly distributed across RGB channels and Y represents intensity only. During good lighting conditions, we also employ the color channels (Cb and Cr) of YCbCr color space to increase foreground segmentation accuracy. During intermediate lighting conditions, both RGB and YCbCr color spaces complement each other in providing a robust FG/BG classification.

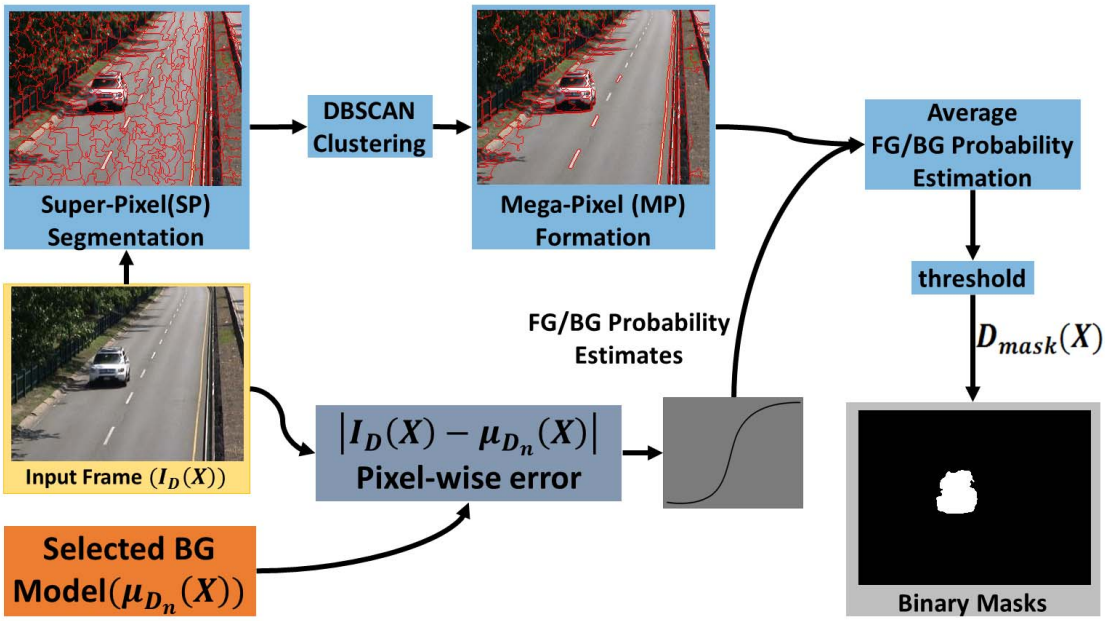

Fig. 1. Binary classification and mask generation.

To support our claim of using multiple color spaces, a detailed quantitative analysis is presented in section V by comparing segmentation accuracy across 12 different categories using each color space separately, two color spaces combined, and by dynamically choosing color channels.

B. Background Modelling

BG modelling is one of most important steps in a BS process and the accuracy of the model used directly impacts the segmentation results. Most BG models use a variant of multi-modal pixel-wise statistical background model. Such an approach has two problems: first, it is difficult to determine the correct number of modes for modelling the pixel probability distribution function. Second, and more importantly, inter-pixel dependencies are overlooked, which leads to poor segmentation results.

In order to model the BG, we propose Background Model Bank (BMB), which comprises of multiple BG models instead of a single BG model. To form BMB, each background training image is treated as a BG model with selected color channels stacked together as a vector. This initial set of BG models are then merged together into a number of average BG models using an iterative sequential clustering procedure. Two BG mean models (p and q in vector form) with correlation measure greater than the pre-defined parameter $corr\_th$ are merged and replaced by their average. The correlation measure is defined as

$$Corr(p,q) = \left(\frac{(p-\mu_p)(q-\mu_q)'}{\sqrt{(p-\mu_p)(p-\mu_p)'}\sqrt{(q-\mu_q)(q-\mu_q)'}} \right) \quad (1)$$

where μ_p and μ_q are defined as:

$$\mu_p = \frac{1}{|X|}\sum_j p_j \text{ and } \mu_q = \frac{1}{|X|}\sum_j q_j \quad (2)$$

This process continues in an iterative fashion unless there are no more average BG models with $Corr > corr\_th$.

The use of frame-level clustering is motivated by physical laws that govern scene geometry. Typically real-life scenes comprise of different types of objects. The variety in configurations and interactions between different types of matter and objects generate very intricate and infinite scene geometry. Examples include variations caused by illumination changes, dynamic changes, camera shaking, camera movement etc. This diversity makes it difficult to accurately capture and model the scene. The use of multiple BG models allows us to capture scene more accurately while keeping spatial dependencies intact.

Another advantage of BMB is that it is computationally simpler than other multi-mode approaches – as we will demonstrate, we choose a model at frame level and ignore the rest of the BG models in the BMB. While there is an additional cost on choosing the model at frame level, it incurs minimal cost because of simple comparison with average BG models than those that rely on pixel-based multi-mode distributions.

As our experimental results in Section V will demonstrate, our multiple BG models can capture scene diversity and camera variations accurately. Comparing to more complex multi-modal or non-parametric techniques, our model obtain equal or better results using only simple binary classifier for pixel classification, resulting in efficient implementation.

C. Binary Classification

In this sub-section, we discuss the binary mask generation for each of the selected color channels. It is a four step process: color channel activation/deactivation, pixel-level probability estimation, MP formation and average probability estimation. Fig. 1 depicts the binary mask generation process.

1) Color-Channels Activation/Deactivation: This step is responsible to activate/deactivate the color channels Cb and

Cr. Both color channels are used if the mean intensity of input image is greater than empirically determined parameter *channel_th*, which otherwise are not employed.

2) Pixel-Level Probability Estimation: Pixel-wise error,$err_D(X)$ is calculated between each color channel from both RGB and YCbCr spaces and the chosen BG model as follows.

$$err_D(X) = \left| I_D(X) - \mu_{D_n}(X) \right| \tag{3}$$

where D denotes the color channel in question,$I_D(X)$ is the input image, and $\mu_{D_n}(X)$ is the chosen average BG model.

Once we have calculated the error for each individual pixel, we estimate an initial probability ip for each pixel by passing them through a sigmoid function.

$$ip(err_D(X)) = \frac{1}{\left(1 + e^{-err_D(X)}\right)} \tag{4}$$

The rationale behind this conversion is that the higher the error the more likely that the pixel belongs to the FG.

3) Mega-Pixel Formation: The primarily goal of this step is to introduce spatial denoising by considering the initial probability estimates ip and color information of the neighbourhood pixels under the framework of Super-Pixels (SP) [41].

SPs offers advantage in terms of capturing local context and significant reduction in computational complexity. These algorithms combine neighboring pixels into one pixel based on similarity measure such as color, texture, size etc. We use the ERS algorithm in [41] to segment the input frame into M SPs. In [41], the SP segmentation is formulated as a graph partitioning problem. For a graph $G = (V, E)$ and M number of SPs, the goal is to find a subset of edges $A \subseteq E$ to approximate a graph $\overline{G} = (V, A)$ with at least M connected sub-graphs. The clustering objective function comprises of two terms: the entropy rate H of a random walk and a balancing term B.

$$\begin{aligned} &\max_A \; H(A) + \lambda B(A), \\ &s.t.\; A \subseteq E \; and \; N_A \geq M \end{aligned} \tag{5}$$

where N_A is the number of connected components in $\overline{G}$. A large entropy term favors compact and homogeneous clusters, whereas the balancing term encourages clusters with similar size. For more details, we refer readers to [41].

To mitigate over-segmentation, SPs are combined to form much bigger Mega-Pixels (MPs) using DBSCAN clustering [42]. DBSCAN is a density based clustering algorithms in which clusters are defined as high density areas, whereas the sparse regions are treated as outliers or borders to separate clusters. Two SPs are merged together into a MP under the following criteria:

$$MP = \begin{cases} 1 & dist \leq colorthreshold \cap SPs\, are\, adjacent \\ 0 & dist > colorthreshold \\ & \cup SPs\, are\, non-adjacent \end{cases}$$

For any two adjacent SPs yand$z$, distance function is based on mean Lab color difference and is defined as:

$$dist = \left|\mu_y^L - \mu_z^L\right| + \left|\mu_y^a - \mu_z^a\right| + \left|\mu_y^b - \mu_z^b\right| \tag{6}$$

$$\mu_y^{ch} = \frac{1}{Y}\sum_{np=1}^{Y} ch(np) \tag{7}$$

where μ_y^{ch} represents the mean value of color channel $ch = \{L, a, b\}$ of SP y. np is the pixel index and Y is the total number of pixels in SP y. Our implementation of DBSCAN is based on [43]. Fig. 1 depicts the overall MP formation process. Notice the road SPs correctly merged as a single MP.

4) Average Probability Estimation and Labelling: The next step is to compute the average probability of a MP y, denoted as AP_y, with a total of Y pixels:

$$AP_y = \frac{1}{Y}\sum_{np=1}^{Y} ip(np) \tag{8}$$

where np is the pixel index and ip is the initial FG/BG probability estimate of each pixel. The AP is then assigned to each pixel belonging to that MP. Finally, to obtain Binary Mask $D_{mask}(X)$ for each color channel D, the average probability measure is thresholded using an empirically determined parameter *prob_th*.

The use of MP and its respective AP allow us to assign the same probability to each pixel belonging to the same object and therefore increases the segmentation accuracy. For example, all the pixels belonging to the road in Fig. 2 should be BG. Clearly, in Fig. 2, as we move from left to right, road pixels with erroneous probability estimates would be averaged out using neighbouring pixels via SPs or MP, thereby improving the segmentation accuracy. As MPs respect edge integrity, the average probability of a MP represents the same object or part rather than using FG/BG probability estimates for each individual pixel or SPs.

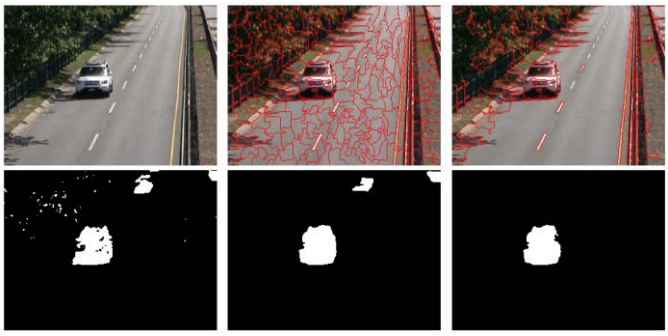

Fig. 2. Comparison of segmentation with probability measure of each pixel individually (left), SP based average motion probability estimation (middle), and MP based average motion probability estimation (right).

D. Model Update

This section explains model update mechanism of the proposed system. Model update is an essential component of an algorithm to deal with scene changes that take place with the passage of time. The classic approach for model update is to replace old values in the model with new ones after a number of frames or time period. Such updating mechanisms can be problematic since the update rate is difficult to determine. For

example, a person sitting idly in a scene may become a part of background if update rate is too fast. Another scenario could be of a forgotten luggage, in which question arises as when should it become a part of background or should it ever become a part of background?

An update mechanism should be able to address two questions. First, is there a need for model update at all? Second, what is the appropriate update rate? We argue that *rate of change in number of FG pixels* can serve as a good measure to trigger model update and to determine an appropriate update rate. In a typical surveillance scene, the number of FG pixels fluctuates in a relatively narrow range and a significant change can serves as a trigger for departure from the old BG model:

$$modelupdate = \begin{cases} 1 & if\ rateOfChange \geq th \\ 0 & otherwise \end{cases}$$

where *th* is an empirically-determined parameter that signifies a significant enough change for model update. The *rateOfChange* is calculated based on the deviation of the number of FG pixels in current frame from the running mean. Formally, we define it as:

$$rateOfChange = \frac{\sum_{x \in X} O_t(X) - \frac{1}{h}\left(\sum_{i=t-h-1}^{t-1} \sum_{x \in X} O_i(X)\right)}{\frac{1}{h}\left(\sum_{i=t-h-1}^{t-1} \sum_{x \in X} O_i(X)\right)} \tag{9}$$

where $O_t(X)$ is the output binary mask of current input image at time t.

Once model update mechanism is triggered and *rateOfChange* is calculated, an update rate function f is used to map rate of change to determine an appropriate update rate U and defined as:

$$U = f(rateOfChange) \tag{10}$$

In order to understand the need for an update rate function f, we must first understand how and what type of changes can occur in a scene. Changes in BG can occur at different rates from slow to abrupt. The gradual illumination change in daylight from sunrise to sunset is a good example of a slowly changing BG and requires a slow update rate. Whereas on the contrary, there can be abrupt changes such as caused by sudden illumination changes in indoor environments or due to a moving camera. Situations such as these require a fast update rate. Failure to determine an appropriate update rate can result in too many false positives. Hence it is necessary for the algorithm to be able to dynamically determine appropriate update rate for changing BG.

There are different options for choosing an update rate function f ranging from simple linear to complex functions. Two candidates are a linear function or an exponential function based on the simplicity of parameters and their effectiveness. A linear function provides a straightforward direct relationship between the model update rate and the rate of change. Exponential function can be used when a more aggressive response i.e. higher update rate is desired for any small change in BG. Such function may be more suitable for coping sudden illumination changes and PTZ camera movements. In our experiments, we have used a simple linear function:

$$U = m * rateOfChange \tag{11}$$

where m is the slope and can be set by the user to any value between zero to one. For example with m set to 0.75 and a rate of change of 1, the calculated update rate would be 0.75, i.e. less weightage is given to old BG model and current frame is given more weightage in updating the BG model.

After determining the update rate, the models are then updated as follows:

$$\mu_n(X) = (1 - U).\mu_n(X) + U.I_t(X) \tag{12}$$

where $I_t(X)$ represents current input frame at time t and $(\mu_n(X))$ is the chosen BG model for current frame and is being updated.

The dynamic model update mechanism allows to cater for various scenarios in which conventional approaches fail. For example, no model update will be applied when there is no FG in the scene or FG is not changing as the rate of change is close to zero. Lastly, whenever there is a change in BG, it is able to dynamically determine update rate and then update BG model.

IV. System Integration

In this section, we describe how individual components are combined in our system. The proposed system consists of five steps as shown in Fig. 3. Each step is described below.

Step 1: BG Model Selection

The first step is to select an appropriate BG Model for the incoming frame. The selection criterion is based on identifying the BG model in BMB that maximizes the correlation with input image $I(X)$:

$$Corr = \arg\max_{n=1,\ldots,N} \times \left(\frac{(I - \mu_I)(\mu_n - \mu)'}{\sqrt{(I - \mu_I)(I - \mu_I)'}\sqrt{(\mu_n - \mu)(\mu_n - \mu)'}}\right) \tag{13}$$

where, I and μ_n are vector forms of $I(X)$ and $\mu_n(X)$ respectively. μ_I and μ are defined as:

$$\mu_I = \frac{1}{|X|}\sum_j I_j \text{ and } \mu = \frac{1}{|X|}\sum_j \mu_{nj} \tag{14}$$

Step 2: Binary Mask (BM) Generation

In this step, the input image and the selected BG model are first used to estimate an initial probability estimate for each pixel. The input image is simultaneously passed to the MP module, which segments the image in arbitrary number of MPs. Average probability estimates are calculated for each MP using pixel-level probability estimates and then thresholded to generate Binary Mask(BM) for each color channel. We denote the BM for color channel D as $D_{mask}(X)$. The BM generation is discussed in detail in section III.C.

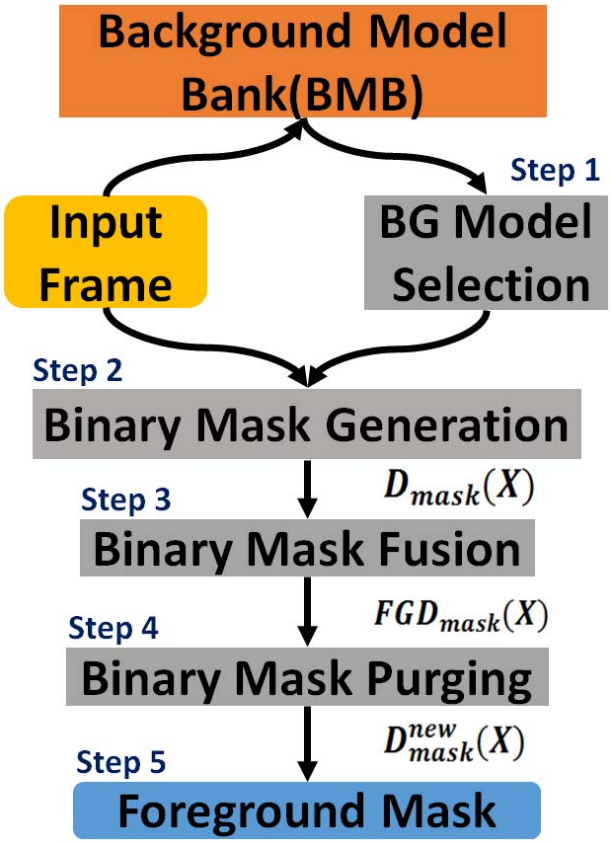

Fig. 3. Universal Multimode Subtraction System.

Step 3: Binary Masks Aggregation/Fusion

The BMs are then used to form Foreground Detection (FGD) masks for RGB and YCbCr color spaces:

$$FGD^{colorspace}_{mask}(X) = \left[\sum\nolimits_{D}\left(D_{mask}(X)\right)\right] > 1 \quad (15)$$

For YCbCr color space, if Cb and Cr channels are deactivated then FGD^{YCbCr}_{mask} will be reduced to the Y channel BM alone. Finally the two FGD masks are combined by taking logical AND between dilated versions of the two to obtain the actual FGD mask:

$$FGD_{mask}(X) = \text{Dilate}(FGD^{RGB}_{mask}(X))\&\text{Dilate}(FGD^{YCbCr}_{mask}(X)) \quad (16)$$

The dilated versions are to ensure that all true foreground pixels are captured in the FGD mask.

Step 4: Binary Masks Purging

The FGD mask is then applied to each of the BMs obtained in step 3. This removes all of the falsely detected foreground regions and increases our confidence in classifying FG and BG pixels in the final step. The resulting component masks are defined as follows:

$$D^{new}_{mask}(X) = D_{mask}(X) \cdot \text{Dilate}(FGD_{mask}(X)) \quad (17)$$

Step 5: Foreground Mask

In the final step of the process, FG mask is obtained by the logical OR of all the $D^{new}_{mask}(X)$ masks.

V. Experiments and Results

In this section, we compare the proposed system with state of the art algorithms on publicly available test sequences. Two datasets are included; CDnet 2014 [2] and ESI [44].

A. CDnet 2014 Dataset

The CDnet 2014 dataset [2] is one of the most comprehensive datasets available for evaluating BS algorithms. It has 11 different categories: *Baseline (BL), Dynamic Background (DB), Camera Jitter (CJ), Intermittent Object Motion (IOM), Shadow (SHD), Thermal (TH), Bad Weather (BW), Low Framerate (LFR), Night Videos (NV), Pan Tilt Zoom (PTZ)* and*Turbulence (TB)*. Each category has 4 to 6 videos totalling to 53 video test sequences. The authors have clearly identified training and testing data to ensure consistency for comparing state of the art algorithms.

1) Evaluation Metrics: The authors of [2] use the seven evaluation metrics:

1. $Recall\,(Re) : \frac{TP}{TP+FN}$
2. $Specificity\,(Sp) : \frac{TN}{TN+FP}$
3. $FalsePositiveRate\,(FPR) : \frac{FP}{FP+TN}$
4. $FalseNegativeRate\,(FNR) : \frac{FN}{FP+TN}$
5. $PercentageofWrongClassifications\,(PWC) :$

$$100 * \frac{(FN+FP)}{(FN+FP+TN+TP)}$$

6. $Precision\,(Pr) : \frac{TP}{TP+FP}$
7. $F-Measure(FM) : 2 * \frac{Pr.Re}{Pr+Re}$

An additional metric has been introduced by authors of [2] for *Shadow (SHD)* category. This metric determines False Positive Rate in hard-shadow areas (FPR-S). Finally, in order to compare the state of the art algorithms, the authors combine these metrics into an overall average rank (R) and average rank across categories (RC) metrics. For details of these metrics, the readers are referred to [2].

In our evaluation, we primarily use F-Measure (FM) for overall and category-wise comparison purposes for a number of reasons. First, the authors of [2] indicate strong correlation of FM with ranks on CDnet website and in general is considered as a good indicator for comparison purposes. Second, in [16], the authors identifies potential biasness towards "precise" method - change detection is an unbalanced classification problem as there are more BG pixels in comparison to FG pixels. As a result, PWC metric would therefore favour "precise" methods. Furthermore, the ranking relies on two reciprocal metrics, i.e. FPR and Sp, and hence it will favour "precise" method. Third, nonlinearity of overall ranks substantially affect how top methods are ranked and therefore is not a reliable indicator for comparing methods.

2) Parameter Setting: One set of parameters are used for the entire dataset: $corr\_th = 0.99$, $prob\_th = 0.75$, $M = 300$, $colorthreshold = 3$, $channel\_th = 100$, $th = 0.15$ and $m = 0.5$. The parameter setting is based on the set that yields overall best results across all categories. For details of parameter used by other techniques, we refer readers to the website at [3].

TABLE I

MBS EVALUATION WITH RGB, YCb Cr AND BOTH (RGB & YCb Cr) COLOR SPACES ON THE CDNET 2014 DATASET

Method	$FM_{overall}$	FM_{BW}	FM_{LFR}	FM_{NV}	FM_{PTZ}	FM_{TB}	FM_{BL}	FM_{DB}	FM_{CJ}	FM_{IOM}	FM_{SHD}	FM_{TH}	FPR-S
MBS-RGB	0.6708	0.771	0.604	0.473	0.468	0.445	0.858	0.707	0.799	0.735	0.735	0.778	0.591
MBS-YCbCr	0.6040	0.595	0.470	0.295	0.383	0.453	0.751	0.659	0.802	0.718	0.735	0.777	0.397
MBS-Both	0.7030	0.787	0.618	0.393	0.609	0.466	0.888	0.783	0.874	0.750	0.773	0.787	0.436
MBS	0.7179	0.787	0.618	0.534	0.609	0.466	0.888	0.783	0.874	0.763	0.776	0.789	0.465

TABLE II

CATEGORY-WISE COMPARISONS ON THE CDnet 2014 DATASET*

Method	FM_{BL}	FM_{DB}	FM_{CJ}	FM_{IOM}	FM_{SHD}	FM_{LFR}	FM_{NV}	FM_{PTZ}	FM_{BW}	FM_{TB}	FM_{TH}	FPR-S
MBS	0.888	0.783	0.874	0.763	0.776	0.618	0.534	0.609	0.787	0.466	0.789	0.465
FTSG [14]	0.933	0.879	0.751	0.789	0.883	0.625	0.513	0.324	0.822	0.712	0.776	0.500
SuBSENSE [16]	0.950	0.817	0.815	0.656	0.898	0.644	0.559	0.347	0.861	0.779	0.817	0.599
CwisarDH [31]	0.914	0.827	0.788	0.575	0.858	0.640	0.373	0.321	0.683	0.722	0.786	0.554
Spectral-360 [45]	0.933	0.776	0.714	0.560	0.851	0.643	0.483	0.365	0.756	0.542	0.776	0.581
Bin Wang Apr 2014 [46]	0.881	0.843	0.710	0.721	0.812	0.468	0.380	0.134	0.767	0.754	0.759	0.465
SC_SOBS [30]	0.933	0.668	0.705	0.591	0.778	0.546	0.450	0.040	0.662	0.488	0.692	0.603
KNN [13]	0.841	0.686	0.689	0.502	0.746	0.549	0.420	0.212	0.758	0.519	0.604	0.397
RMoG [47]	0.784	0.735	0.701	0.543	0.721	0.531	0.426	0.247	0.682	0.457	0.478	0.309
KDE [11]	0.909	0.596	0.572	0.408	0.803	0.547	0.436	0.036	0.757	0.447	0.742	0.621
SOBS_CF [48]	0.929	0.651	0.715	0.581	0.772	0.514	0.448	0.036	0.637	0.470	0.714	0.589
Mahalanobis distance [49]	0.464	0.179	0.335	0.229	0.335	0.079	0.137	0.037	0.221	0.335	0.138	0.064
GMM [10]	0.824	0.633	0.596	0.520	0.737	0.537	0.409	0.152	0.738	0.466	0.662	0.535
GMM-Zivkovic [50]	0.838	0.632	0.567	0.532	0.732	0.506	0.396	0.104	0.740	0.416	0.654	0.542
Multiscale Spatio-Temporal [23]	0.845	0.595	0.507	0.449	0.791	0.336	0.416	0.036	0.637	0.529	0.510	0.528
Euclidean distance [49]	0.872	0.508	0.487	0.489	0.678	0.501	0.385	0.039	0.670	0.413	0.631	0.576

*In each column, Red font is for best, Green Font for second best, and Blue font for third best results. Out of the 12 categories, MBS ranks first in 3 categories, second in 2 and third in 2. MBS is within the top five for all categories.

TABLE III

COMPLETE RESULTS FOR MBS ON THE CDnet 2014 DATASET

Category	Average Recall	Average Specificity	Average FPR	Average FNR	Average PWC	Average F-Measure	Average Precision
Bad Weather	0.7493	0.9969	0.0031	0.2507	0.7416	0.7876	0.8312
Low Framerate	0.6345	0.9947	0.0053	0.3655	0.9268	0.6189	0.6044
Night Videos	0.6117	0.9722	0.0278	0.3883	3.9754	0.5393	0.4950
PTZ	0.6684	0.9968	0.0032	0.3316	0.5105	0.6099	0.5647
Turbulence	0.4584	0.9984	0.0016	0.5416	0.3149	0.4661	0.5135
Baseline	0.9010	0.9957	0.0043	0.0990	0.6439	0.8882	0.8762
Dynamic Background	0.7957	0.9984	0.0016	0.2043	0.3076	0.7833	0.7734
Camera Jitter	0.8940	0.9926	0.0074	0.1060	1.1785	0.8743	0.8564
Intermittent Object Motion	0.7525	0.9813	0.0187	0.2475	2.6327	0.7636	0.7913
Shadow	0.7901	0.9898	0.0102	0.2099	1.7816	0.7765	0.7662
Thermal	0.7919	0.9916	0.0084	0.2081	1.5217	0.7891	0.7872
Overall	0.7316	0.9917	0.0083	0.2684	1.3214	0.7179	0.7145

3) Quantitative Evaluation: In this section, we compare our proposed MBS system with 15 state of the art algorithms: Flux Tensor with Split Gaussian models(FTSG) [14], suBSENSE [16], CwisarDH [31], Spectral-360 [45], Bin Wang Apr 2014 [46], KNN [13], SC_SOBS [30], Region-based Mixture of Gaussians (RMoG) [47], KDE - ElGammal [11], SOBS_CF [48], Mahalanobis distance [49], GMM-Stauffer & Grimson [10], GMM-Zivkovic [50], Multiscale Spatio-Temporal BG Model [23] and Euclidean distance [49]. Table II presents F-Measure based category-wise comparisons. Table IV provides the overall comparison in terms of $FM_{overall}$. Table III provides the complete statistics of MBS on the CDnet 2014 dataset.

Furthermore, we provide three additional configurations of MBS using RGB color space alone, YCbCr color space alone and RGB and YCbCr combined. These are denoted by MBS-RGB, MBS-YCbCr and MBS-Both respectively. The F-Measure based overall and category-wise comparisons for these configurations are presented in Table I. These additional comparisons serve two purposes: (a) to quantitatively analyse the robustness that is offered by selecting appropriate color spaces and channels in comparison with using a single or combination of color spaces for every scene/test sequence, and (b) to analyse strength and weaknesses of color spaces in different categories.

4) Discussions: We first analyse the performance of different MBS configurations namely MBS-RGB, MBS-YCbCr, MBS-Both and MBS (i.e. when appropriate color space and channels are selected). As depicted in Table I, MBS not only has the highest overall F-Measure but it outperforms

TABLE IV
OVERALL COMPARISON ON THE CDNET 2014 DATASET*

Method	$FM_{overall}$
MBS	0.7179
FTSG [14]	0.7283
SuBSENSE [16]	0.7408
CwisarDH [31]	0.6812
Spectral-360 [45]	0.6732
Bin Wang Apr 2014 [46]	0.6577
SC_SOBS [30]	0.5961
KNN [13]	0.5937
RMoG [47]	0.5735
KDE [11]	0.5688
SOBS_CF [48]	0.5883
Mahalanobis [49]	0.2267
GMM [10]	0.5707
GMM-Zivkovic [50]	0.5566
Multiscale Spatio-Temporal [23]	0.5141
Euclidean distance [49]	0.5161

*Red font is for best, Green for second best and Blue for third best result.

other configurations in individual categories as well. This supports the claim that *use of appropriate color space and channels are critical for segmentation accuracy*. The second best configuration is MBS-Both, which combines the strength of both color spaces. Among the two remaining, YCbCr performs marginally better than RGB. MBS-YCbCr is the most robust in handling shadows with the least FPR-S rate of 0.397, while MBS-RGB has the worst performance with the highest FPR-S of 0.591.

As for their performances on each category, NV and BW categories are affected by low lighting conditions and potentially has poor color discrimination problem. There are two important observations; (a) as shown in Table I, F-Measure of MBS-RGB is significantly higher than MBS-YCbCr for both NV and BW categories and, (b) as per Table I, in NV category, using all channels deteriorates the segmentation accuracy because of poor color discrimination in Cb and Cr channels. On the other hand, significant improvement is achieved when the chroma channels are automatically deactivated. This substantiates our earlier claim that *RGB and Y channels are more robust under low lighting conditions or when color discrimination is poor.* The BW category has considerably better illumination conditions, the Cb and Cr channels are retained and the use of all channels produce higher segmentation accuracy than using RGB and YCbCr color spaces alone. In all of the categories except NV, TB and TH (where there is zero color information or poor discrimination), the segmentation results are improved with added advantage of color information.

Next, we compare MBS against other state of the art algorithms. The following seven key points are observed from our comprehensive evaluation and results on different categories.

1. In six out of eleven categories, the proposed system is among top 3 with 1st position in two of them. In DB, BL and SHD categories, MBS is not among top 3 but achieves acceptable results with approximately 80% F-Measure (FM). According to [16], FM $\geq$ 80% is considered an acceptable result.
2. In six out of eleven categories, the proposed system is among top 3 with 1st position in two of them. In DB, BL and SHD categories, MBS is not among top 3 but achieves acceptable results with approximately 80% F-Measure (FM). According to [16], FM $\geq$ 80% is considered an acceptable result.
3. In the NV category, we are ranked at 2nd position with FM of 0.534. Like other top methods, the performance is affected by halos and reflections caused by strong headlights and low visibility.
4. In the LFR category, MBS has FM of 0.618, which is slightly less than FM of 0.644 of top performing method. The result is poor for all methods in this category. It is important to mention that MBS performs poorly in only one of four test sequences in LFR category. This particular test sequence 'port_0_17fps' is recorded at 0.17fps with wavering lighting conditions and intense dynamic behavior of water and boats causing the overall FM to drop down.
5. It is important to note the marked difference in performance of our algorithm against others in the two moving camera categories; PTZ and CJ. Most of the existing state of the art fail due to static camera assumption, whereas the frame level BG modeling and MP spatial denoising approach of proposed system allows it to handle both static and moving camera video sequences and thus make our system universal. We have FM of 87.4% for CJ and 60.9% for PTZ. Although our FM is significantly higher than others, better results (>80% F-Measure) could be achieved for the PTZ category if sufficient training data, especially for the "continuousPan" and "zoomInZoomout" video sequences are available.
6. IOM category involves objects being placed and removed intermittently, our innovative model update mechanism and MP based spatial denoising allows MBS to achieve FM of 76.36% and is placed at 2nd position. The MP approach allows to handle intermittently placed objects. Consider the situation where a box is on a sofa and was learnt as a part of BG. As soon as the box is removed the pixel-level estimates would classify those pixels as FG, however all the neighboring pixels belonging to sofa will average those out and therefore will have no effect on segmentation accuracy while model can be gradually updated. None of the methods in this category is able to produce the defined acceptable FM level $\geq$ 80%. In BW and TH categories, MBS achieves acceptable results and is placed at 3rd and 2nd position respectively. Our worst performance is in TB category. In general for all categories, MBS is always placed among top 5 out of 15 methods.
7. MBS achieves third lowest False Positive Rate Shadow (FPR-S) of 0.465 out of 15 state of the art algorithms. This measure is recommended by the authors of dataset to test algorithms ability to suppress pixels specifically in Shadow regions. It is important to note that none of the top performing methods is able to achieve

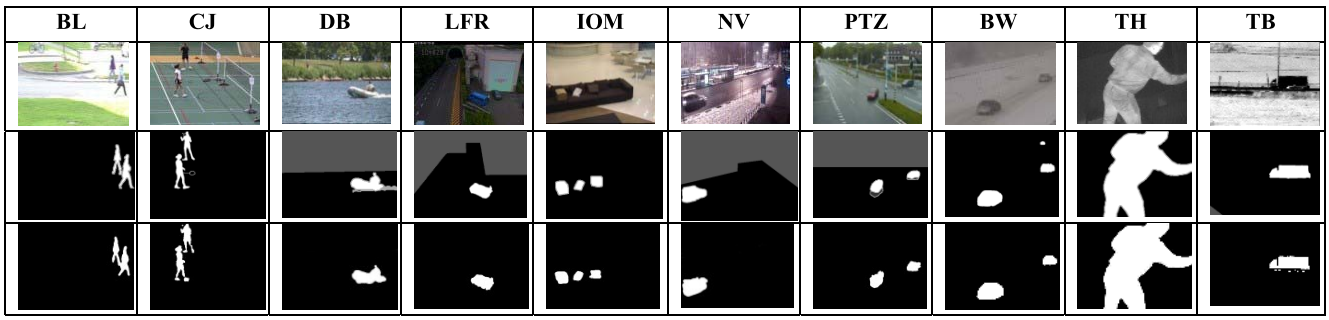

Fig. 4. Foreground Segmentation results of MBS on example frames from CDnet 2014 dataset. Input image (Row 1), Ground truth (Row 2) and MBS output(Row 3).

lower FPR-S than ours, which is significantly lower than top methods.

8. Table IV provides an overall comparison of MBS against 15 state of the art algorithms on CDnet 2014 Dataset. MBS has the third highest overall F-Measure of 0.7179 and is a top performing method. None of the methods except top 3 including ours is able to achieve an overall FM $\geq$ 70%. For metrics of all 15 methods, we refer readers to [2] and [3].

5) Qualitative Results: Fig. 4 presents some sample results of proposed system for different categories of CDnet 2014 dataset. Complete set of results for all categories are available at the official CDnet 2014 website [3] and additionally videos at our website [51].

B. ESI Dataset

Robustness of BS algorithm against sudden illumination changes is very critical to its success in real life scenarios. This is especially true for indoor environments, where sudden lighting change often occurs during door opening and closing, switching light on and off etc. CDnet 2014 dataset lacks such a category. As a result, we include the ESI dataset and, instead of comparing with general BS algorithms, we compare MBS with algorithms that specialize in dealing with this challenge. In our opinion, ESI dataset [44] is the most challenging publicly available test dataset in terms of sudden illumination changes.

ESI dataset comprises of 5 test sequences; sofa, walking, chair, scene1 and scene2 [9]. They have 382, 734, 573, 750 and 154 frames respectively. Since the test sequences sofa, chair and walking have the same background scene/model, we combine these three into a single test sequence "House" comprising of 1689 frames. We now discuss the evaluation metrics, parameter setting for all test sequences and also present quantitative and qualitative results.

1) Evaluation Metrics: For quantitative evaluation of ESI dataset, we use three metrics as defined earlier; precision (Pr), Recall (Re) and F- measure. Precision and Recall are calculated for whole of a test sequence as arithmetic mean over all frames. Using this precision and recall, F-measure is calculated for each test sequence. Overall F-Measure is used for comparison purposes, which is simply mean FM of all test sequences in this category.

TABLE V
PRECISION, RECALL AND F-MEASURE OF MBS ON ESI DATASET

Sequence	IMG	Precision	Recall	F-Measure
House	400	0.7157	0.7541	0.7344
Scene1	500	0.8154	0.7236	0.7667
Scene2	500	0.7358	0.7988	0.7660
Overall	-	0.7556	0.7588	0.7557

2) Parameter Setting: One set of parameters are used for the entire dataset: *corr_th = 0.99, prob_th = 0.75, M = 300, colorthreshold = 3, channel_th = 100, th = 0.15 and m = 0.5*. The parameter setting is based on the set that yields the overall best results across all test sequences. Table V reports the number of training images (IMG) used for making BMB.

3) Quantitative Evaluation: Overall as well as individual results for each test sequence are tabulated in Table V. A comparison of existing state-of-the-art techniques with MBS is depicted in Fig. 6. The techniques include; Eigen background based Statistical Illumination (ESI) [9], Statistical Illumination (SI) [25], Eigen Background (EB) both dynamic and fixed [26], [52], Tonal Alignment (TA) [27] and Adaptive Background Mixture Model (ABMM) [53]. The results for these techniques are obtained from [9].

Clearly, as shown in Fig. 6 MBS outperforms state of the art in handling illumination changes.

4) Qualitative Results: For qualitative results, we choose the ESI technique as benchmark for comparison purposes. Fig. 5 not only presents comparative results of our approach on some of example frames from scene1, scene2 and house test sequences, but also depicts the challenging nature and variation of illumination in these test sequences. Complete comparative video of all test sequences with ground truth and input images can be found at our website [51].

C. Processing Speed

The proposed system is currently implemented in Matlab and run on an Intel core i5 PC with 8GB RAM. For a typical image resolution of 320 $\times$ 240, the current system in its coarse form is able to achieve $\sim$ 10 fps if both color spaces i.e. all color channels are used. Approximately, 70% of processing time is consumed by SP segmentation algorithm, which is an external component. With code optimization and

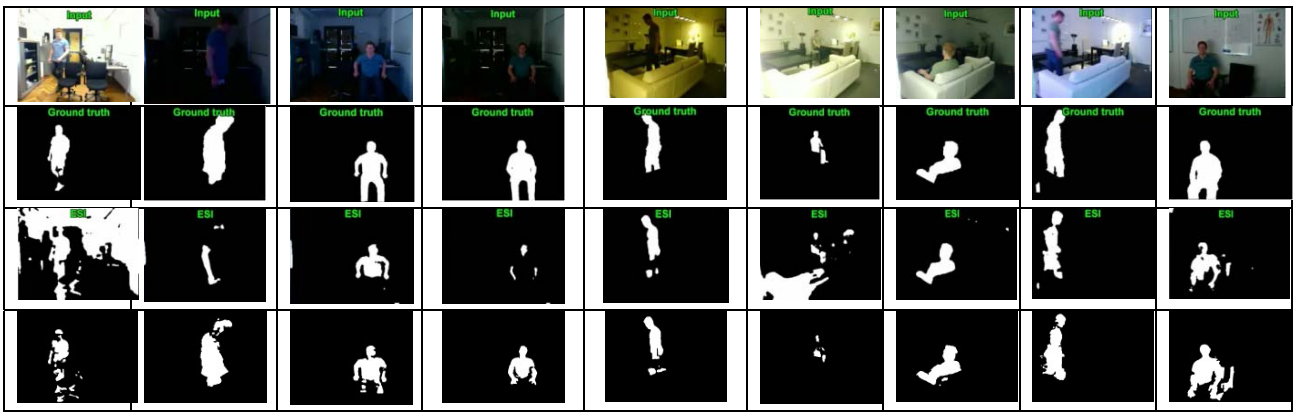

Fig. 5. Foreground segmentation results of example frames from test sequence*scene1 (columns 2-5), house (columns 6-9) and scene2 (column 10)*. Input image (Row 1), Ground truth (Row 2), ESI output (Row 3) and MBS output (Row 4).

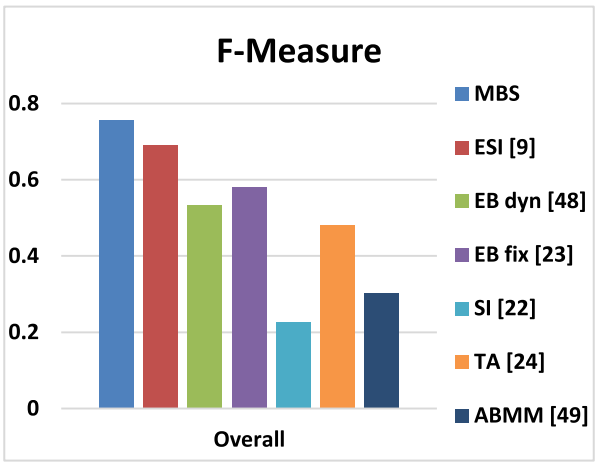

Fig. 6. ESI dataset F-measure.

implementation in C++, the system is expected to meet real time requirements.

VI. Conclusion and Future Work

In this paper, we have presented a universal BG subtraction system that exploits multiple BG models and computationally inexpensive pixel-level comparison to generate initial probability estimates, which undergo spatial denoising by forming MPs. To separate vision tasks based on illumination conditions, we use RGB and Y color channels to for low light vision and CbCr for bright light to provide more accurate foreground segmentation. The introduction of FG dependent model update mechanism eliminates the need to tune parameters for every test sequence.

Comprehensive evaluations of the proposed system over 12 different challenging categories comprising of 56 video test sequences demonstrate the capability and flexibility of proposed system over wide variety of environmental conditions. In 10 out of 12 categories, MBS ranks among top 3 or achieve acceptable results. MBS is clearly a top performing method that outperforms state of the art especially in the moving camera categories and achieves best results for shadow suppression among top methods.

The current implementation of our algorithm is in MATLAB. Code optimization and implementation of algorithm in C/C++ are part of our future work. All results have been made available at official CDnet 2014 website [3].

Acknowledgment

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proc. ICCV*, Sep. 1999, pp. 255–261.
[2] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, "CDnet 2014: An expanded change detection benchmark dataset," in *Proc. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2014, pp. 387–394.
[3] *Changedetection Dataset*, accessed on Dec. 15, 2016. [Online]. Available: https://www.changedetection.net
[4] H. Sajid and S.-C. S. Cheung, "Background subtraction under sudden illumination change," in *Proc. IEEE Multimedia Signal Process. (MMSP)*, Sep. 2014, pp. 1–6.
[5] H. Sajid and S.-C. S. Cheung, "Background subtraction for static moving camera," in *Proc. Int. Conf. Image Process.*, Sep. 2015, pp. 4530–4534.
[6] S. C. S.-Ching and C. Kamath, "Robust techniques for background subtraction in urban traffic video," in *Proc. Electron. Imag.*, 2004, pp. 881–892.
[7] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *Proc. CVPR*, Jun. 2011, pp. 1937–1944.
[8] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection—A systematic survey," *Recent Patents Comput. Sci.*, vol. 4, no. 3, pp. 147–176, 2011.
[9] L. P. Vosters, C. Shan, and T. Gritti, "Background subtraction under sudden illumination changes," in *Proc. AVSS*, Sep. 2010, pp. 384–391.
[10] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking,"in CVPR, 1999.
[11] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proc. ECCV*, 2000, pp. 751–767.
[12] P. D. Z. Varcheie, M. Sills-Lavoie, and G.-A. Bilodeau, "A multiscale region-based motion detection and background subtraction algorithm," *Sensors*, vol. 10, no. 2, pp. 1041–1061, 2010.
[13] Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, 2006.

[14] R. Wang, F. Bunyak, G. Seetharaman, and K. Palaniappan, "Static and moving object detection using flux tensor with split Gaussian models," in *Proc. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2014, pp. 414–418.

[15] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1778–1792, Nov. 2005.

[16] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "SuBSENSE: A universal change detection method with local adaptive sensitivity," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 359–373, Jan. 2015.

[17] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in *Proc. ICIP*, Oct. 2004, pp. 3061–3064.

[18] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground–background segmentation using codebook model," *Real-Time Imag.*, vol. 11, no. 3, pp. 172–185, 2005.

[19] S.-C. Liu, C.-W. Fu, and S. Chang, "Statistical change detection with moments under time-varying illumination," *IEEE Trans. Image Process.*, vol. 7, no. 9, pp. 1258–1268, Sep. 1998.

[20] O. Barnich and M. van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, 2011.

[21] M. Seki, T. Wada, H. Fujiwara, and K. Sumi, "Background subtraction based on cooccurrence of image variations," in *Proc. CVPR*, Jun. 2003, pp. II-65–II-72.

[22] P. W. Power and J. A. Schoonees, "Understanding background mixture models for foreground segmentation," *J. Image Vis. Comput.*, vol. 2002, pp. 10–11, Nov. 2002.

[23] X. Lu, "A multiscale spatio-temporal background model for motion detection," in *Proc. ICIP*, Oct. 2014, pp. 3268–3271.

[24] K. Skifstad and R. Jain, "Illumination independent change detection for real world image sequences," *Comput. Vis., Graph. Image Process.*, vol. 46, no. 3, pp. 387–399, Jun. 1989.

[25] J. Pilet, C. Strecha, and P. Fua, "Making background subtraction robust to sudden illumination changes," in *Proc. ECCV*, 2008, pp. 567–580.

[26] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 831–843, Aug. 2000.

[27] L. di Stefano, F. Tombari, S. Mattoccia, and E. de Lisi, "Robust and accurate change detection under sudden illumination variations," in *Proc. ACCV*, Nov. 2007.

[28] F. Tombari, L. di Stefano, and S. Mattoccia, "A robust measure for visual correspondence," in *Proc. Int. Conf. Image Anal. Process.*, Sep. 2007, pp. 376–381.

[29] B. Han and L. S. Davis, "Density-based multifeature background subtraction with support vector machine," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 1017–1023, May 2012.

[30] L. Maddalena and A. Petrosino, "The SOBS algorithm: What are the limits?" in *Proc. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 21–26.

[31] M. D. Gregorio and M. Giordano, "Change detection with weightless neural networks," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 403–407.

[32] L. Li and M. K. H. Leung, "Integrating intensity and texture differences for robust change detection," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 105–112, Feb. 2002.

[33] X. Zhao, W. He, S. Luo, and L. Zhang, "MRF-based adaptive approach for foreground segmentation under sudden illumination change," in *Proc. Int. Conf. Inform., Commun. Signal Process.*, Dec. 2007, pp. 1–4.

[34] F. Kristensen, P. Nilsson, and V. Öwall, "Background segmentation beyond RGB," in *Proc. ACCV*, 2006, pp. 602–612.

[35] M. Balcilar, F. Karabiber, and A. C. Sonmez, "Performance analysis of Lab2000HL color space for background subtraction," in *Proc. IEEE Int. Symp. Innov. Intell. Syst. Appl.*, Jun. 2013, pp. 1–6.

[36] T. Horprasert, D. Harwood, and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in ICCV, 1999.

[37] Z. Chen, N. Pears, M. Freeman, and J. Austin, "Background subtraction in video using recursive mixture models, spatio-temporal filtering and shadow removal," in *Advances in Visual Computing*. Berlin, Germany: Springer-Verlag, 2009.

[38] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003.

[39] J. Chengjun, C. Guiran, C. Wei, and J. Huiyan, "Background extraction and update method based on histogram in ycbcr color space," in *Proc. Int. Conf. E-Business E-Government*, May 2011, pp. 1–4.

[40] T. Gevers, A. Gijsenij, J. Van de Weijer, and J.-M. Geusebroek, *Color in Computer Vision: Fundamentals and Applications*. Hoboken, NJ, USA: Wiley, 2012.

[41] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. CVPR*, Jun. 2011, pp. 2097–2104.

[42] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *Kdd*, vol. 96, no. 34, pp. 226–231, 1996.

[43] P. D. Kovesi. *MATLAB and Octave Functions for Computer Vision and Image Processing*, accessed on Apr. 15, 2016. [Online]. Available: http://www.peterkovesi.com/matlabfns/

[44] *Esi Dataset*, accessed on Dec. 20, 2015. [Online]. Available: https://sites.google.com/site/-tommasogritti/publications/background-subtraction-data

[45] M. Sedky, M. Moniri, and C. C. Chibelushi, "Spectral-360: A physics-based technique for change detection," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, 2014, pp. 399–402.

[46] B. Wang and P. Dudek, "A fast self-tuning background subtraction algorithm," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, 2014, pp. 395–398.

[47] S. Varadarajan, P. Miller, and H. Zhou, "Spatial mixture of gaussians for dynamic background modelling," in *Proc. IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Aug. 2013, pp. 63–68.

[48] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," *Neural Comput. Appl.*, vol. 19, no. 2, pp. 179–186, 2010.

[49] Y. Benezeth, P.-M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Comparative study of background subtraction algorithms," *J. Electron. Imag.*, vol. 19, no. 3, 2010.

[50] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proc. Int. Conf. Pattern Recognit.*, 2004, pp. 28–31.

[51] *Hasan Sajid*, accessed on Apr. 22, 2017. [Online]. Available: https://sites.google.com/site/hasansajid/research

[52] B. Han and R. Jain, "Real-time subspace-based background modeling using multi-channel data," in *Advances in Visual Computing*. Berlin, Germany: Springer-Verlag, 2007.

[53] P. KaewTraKulPong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Video-Based Surveillance Systems*. Norwell, MA, USA: Kluwer, 2002.

Hasan Sajid received the B.S. degree in mechatronics engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2007, and the M.S. and Ph.D. degrees in electrical engineering from the University of Kentucky, USA, in 2014 and 2016, respectively. He is currently an Assistant Professor with the Department of Robotics and AI, School of Mechanical and Manufacturing Engineering, NUST. His research interests include traffic and crowd analytics, video surveillance, indoor localization, computer vision, machine learning, and robotics. He was a recipient of the U.S. State Department Fulbright Scholarship for the M.S. degree Program at the University of Kentucky in 2012.

Sen-Ching Samson Cheung (M'91–SM'07) received the Ph.D. degree from the University of California at Berkeley, Berkeley, CA, USA, in 2002. He was a Computer Scientist with the Scientific Data Mining Group, Lawrence Livermore National Laboratory. In 2004, he joined the University of Kentucky (UKY), Lexington, KY, USA, where he is currently an Associate Professor with the Department of Electrical and Computer Engineering. He has a joint appointment with the UKY Center of Visualization and Virtual Environments. His work spans a number of different areas in multimedia, including video copy detection, data mining, video surveillance, privacy protection, encrypted domain signal processing, and computational multimedia for therapy. He is a Senior Member of the ACM. He is/was an Associate Editor of the IEEE TRANSACTIONS OF MULTIMEDIA, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, *Signal Processing: Image Communications*, and the *EURASIP Journal on Information Security*.