# UNIVERSAL MULTIMODE BACKGROUND SUBTRACTION BY USING DIFFERENT COLOR SPACE CONVERSIONS

KATTHI SHASHIKALA (M.E)[1]
GOLLA NIRMALA(M.E(ASSISTANT PROFESSOR))[2]
Stanley College Of Engineering And Technology For Women Abids, Hyderabad, Telangana, INDIA[1,2]
shashikatthi447@gmail.com[1] gnirmala@stanley.edu.in[2]

## ABSTRACT

In this paper, we display an entire change discovery framework named multimode foundation subtraction. The general idea of framework enables it to powerfully deal with huge number of difficulties related with video change location, for example, enlightenment changes, dynamic foundation, camera jitter, and moving camera. The framework contains various creative components in foundation displaying, show refresh, pixel order, and the utilization of different shading spaces. The framework initially makes different foundation models of the scene took after by underlying frontal area/foundation likelihood estimation for every pixel. Next, the picture pixels are combined to shape megapixels, which are utilized to spatially denoise the underlying likelihood appraisals to create paired veils for both RGB and YCbCr shading spaces. The covers created in the wake of handling these info pictures are then consolidated to isolate forefront pixels from the foundation. Thorough assessment of the proposed approach on freely accessible test successions from the CDnet and the ESI informational collections demonstrates predominance in the execution of our framework over other best in class calculations.

***Key Words***— 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 motionand 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, totaling 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].

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 modeling 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).

The GMM based techniques model the per-pixel distribution of values observed overtime with a mixture of Gaussians. The multimodal nature of these techniques allows them to cope with dynamic background. GMM has been widely used for different BS systems and various improved versions have been proposed. Another popular algorithm in this category is based on KDE. 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. Suspense is an example of sample consensus methods that uses pixel-level feedback loop mechanism to continuously update and maintain the pixel's model. 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 previous models. It comprises of a codebook for each pixel which is a compressed form of background. Each codebook has multiple code words that are based on a sequence of training images using a color distortion metric. Incoming pixels are matched against all background code words 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 classes of techniques are region-based techniques. Unlike their pixel-based counterparts, region-based techniques exploit local spatial relationships among pixels. In previous models, 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 previous models 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 the previous models spatial dependencies by considering blocks of different sizes instead of pixels individually. The basic underlying assumption is that the neighboring 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 previous models, 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 previous models, 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 Spatiotemporal uses a three-level Spatio-temporal color/luminance Gaussian pyramid BG model for each pixel. 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].

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.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. 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). For an input image, it first uses the change detection algorithm in 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 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 classes of methods apply traditional machine learning on different features to build the BG model. For example, 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 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 modeling, 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 modeling 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.

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; (c) it is the output format of most cameras and its direct usage in BS avoids the computation cost of color conversion.

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. 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.

In comparative studies on color spaces 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. 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. 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, previous methods 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 are 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.

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.

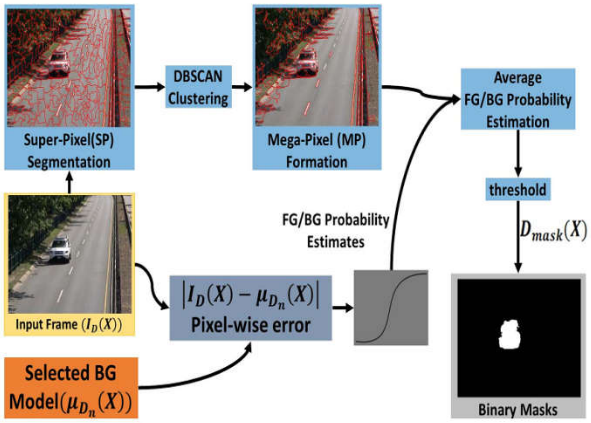


Fig.1. Binary classification and mask generation.

### *B. Background Modeling:*

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 is 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 $\mu_p$ and $\mu_q$ are defined as:

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

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 andCr. 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.

$$err_D(X) = |I_D(X) - \mu D_n(X)| \quad (3)$$

Where D denotes the color channel in question, ID (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 i p for each pixel by passing them through a sigmoid function.

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

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 i p and color information of the neighbourhood pixels under the framework of Super-Pixels (SP) [41].

SPs offer 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 ⊆ E to approximate a graph 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.

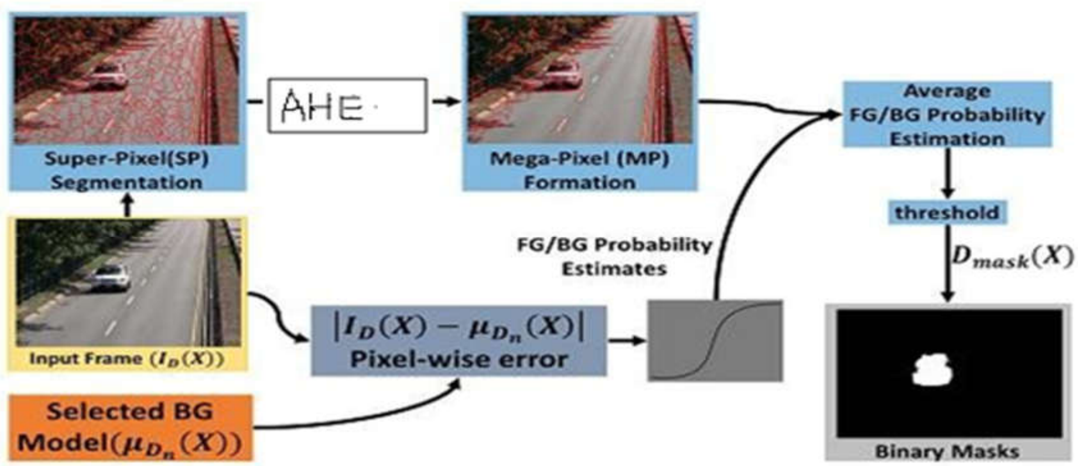


Fig B:Binary Mask for Single Frame Using Adaptive Histogram Equalization

$$\max_A \quad H(A) + \lambda B(A),$$
$$s.t. A \subseteq E \text{ and } N_A \geq M \quad (5)$$

Where NA is the number of connected components in 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 yandz, distance function is based on mean Lab color difference and is defined as:

$$dist = \left|\mu_y^L - \mu_z^L\right| + \left|\mu_y^a - \mu_z^a\right| + \left|\mu_y^b - \mu_z^b\right| (6)$$

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

Where µch y 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:

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

Where np is the pixel index and i p 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 Dmask (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 neighboring 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.

### *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:

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

Whereth 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:

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

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:

$$U = f(rate\ of\ Change) (10)$$

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:

$$U = m * rate\ Of\ Change \qquad (11)$$

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:

$$\mu_n(X) = (1-U).\mu_n(X) + U.I_t(X) \qquad (12)$$

Where It (X) represents current input frame at time t and (µn (X)) is the chosen BG model for current frame and is being updated.

The dynamic model update mechanism allows catering 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.2. 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):

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

where, I and µn are vector forms of I (X) and µn(X) respectively. µI and µ are defined as:

$$\mu_I = \frac{1}{|X|}\sum\nolimits_j I_j \quad and \quad \mu = \frac{1}{|X|}\sum\nolimits_j \mu_{nj} \qquad (14)$$

**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 Dmask(X). The BM generation is discussed in detail in section III.C.

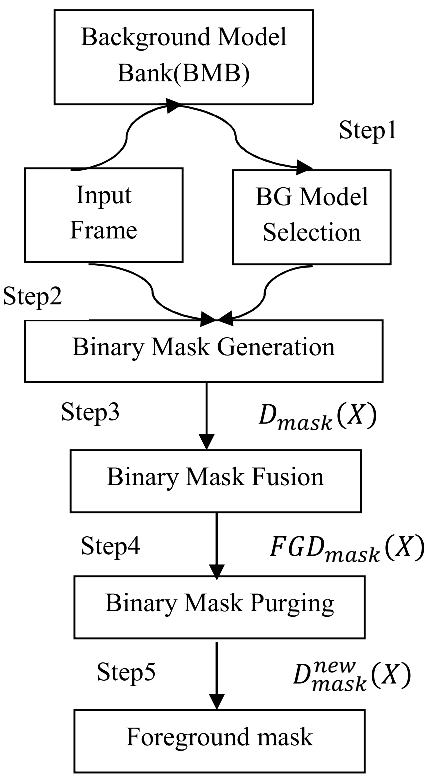


Fig.3. Universal Multimode Subtraction System.

**Step 3**: Binary Masks Aggregation/Fusion

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

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

For YCbCr color space, if Cb and Cr channels are deactivated then FG DY 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:

$$FGD_{mask}(X) = Dilate(FGD_{mask}^{RGB}(X))\&Dilate\left(FGD_{mask}^{YCbCr}(X)\right)(16)$$

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:

$$D_{mask}^{new}(X) = D_{mask}(X).Dilate(FGD_{mask}(X))(17)$$

**Step 5**: Foreground Mask

In the final step of the process, FG mask is obtained by the logical OR of all the Dmask new (X) masks.

## V. EXPERIMENTS&RESULTS

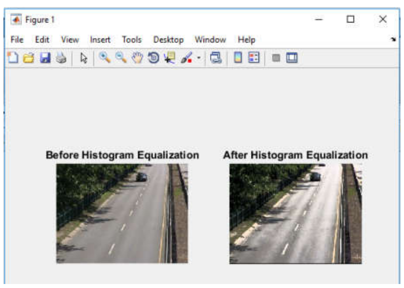


Fig 1:Before and After Adaptive Histogram Equalization

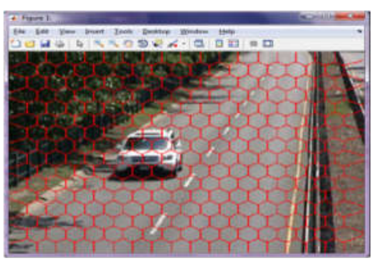

Fig 2: Super Pixel Image

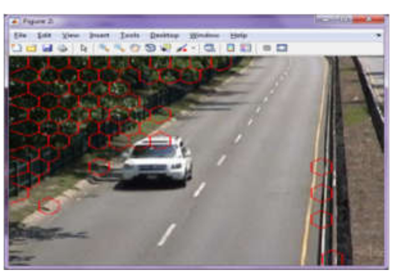

Fig 3:Mega Pixel Image

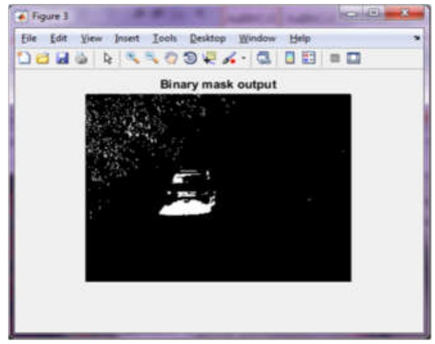


Fig 4: Binary Mask Output

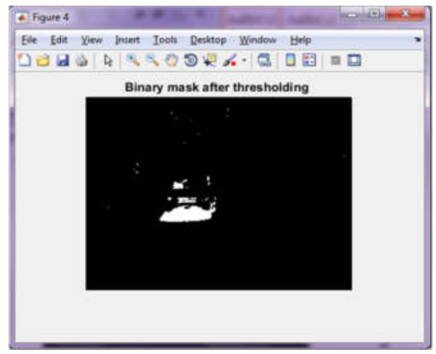


Fig 5:Binary Mask After Thresholding

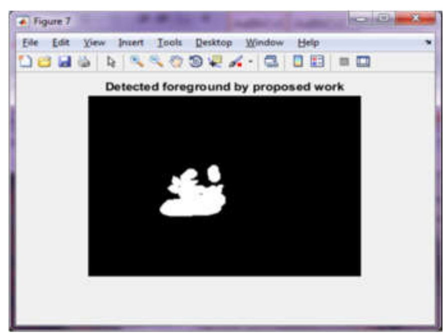


Fig 6:Detected Foreground by Proposed Work

## VI. CONCLUSION&FUTURE WORK

In this paper, we have displayed an all inclusive BG subtraction framework that adventures various BG models and computationally modest pixel-level correlation with produce starting likelihood gauges, which experience spatial denoising by shaping MPs. To isolate vision errands in view of brightening conditions, we utilize RGB and Y shading channels to for low light vision and CbCr for splendid light to give more precise forefront division. The presentation of FG subordinate model refresh system wipes out the need to tune parameters for each test grouping.

Far reaching assessments of the proposed framework over 12 distinctive testing classes containing 56 video test arrangements exhibit the capacity and adaptability of proposed framework over wide assortment of ecological conditions. In 10 out of 12 classifications, MBS positions among top 3 or accomplish worthy outcomes. MBS is plainly a best performing strategy that beats cutting edge particularly in the moving camera classes and accomplishes best outcomes for shadow concealment among top strategies.

## Future Work:

- In our proposed method we are using RGB to HSV colour space conversion.
- Our future work is to apply the different color space conversions,the use of these conversions to provide more accurate results.
- The introduction of FG dependent model update mechanism eliminates the need to tune parameters for every test sequence.

## 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].Availablehttps://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.

Shashikala Katthi received the B.Tech degree in Electronics and Communication Engineering from the Jawaharlal Nehru Technological University in 2016.she is interested in areas of Embedded systems and Signal Processing.

Nirmala Golla received the M.E degree in Embedded systems from the Osmania University and B.Tech degree in Electronics and Communication Engineering from the Jawaharlal Nehru Technological University. she is working as an Assistant Professor with the department of Electronics and Communication Engineering. she is interested in areas of Embedded systems and Signal Processing.