Theses and Dissertations--Electrical and Computer Engineering

Electrical and Computer Engineering

2014

A Universal Background Subtraction System

Hasan Sajid
University of Kentucky, hasan.sajid@gmail.com

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Hasan Sajid, Student

Dr. Sen-Ching Samson Cheung, Major Professor

Dr. Cai-Cheng Lu, Director of Graduate Studies

A UNIVERSAL BACKGROUND SUBTRACTION SYSTEM

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in Electrical Engineering
in the College of Engineering
at the University of Kentucky

By

Hasan Sajid

Lexington, Kentucky

Director: Dr. Sen-Ching Samson Cheung, Professor of Electrical and Computer Engineering

Lexington, Kentucky

2014

ABSTRACT OF THESIS

A UNIVERSAL BACKGROUND SUBTRACTION SYSTEM

Background Subtraction is one of the fundamental pre-processing steps in video processing. It helps to distinguish between foreground and background for any given image and thus has numerous applications including security, privacy, surveillance and traffic monitoring to name a few. Unfortunately, no single algorithm exists that can handle various challenges associated with background subtraction such as illumination changes, dynamic background, camera jitter etc. In this work, we propose a Multiple Background Model based Background Subtraction (MB^2S) system, which is universal in nature and is robust against real life challenges associated with background subtraction. It creates multiple background models of the scene followed by both pixel and frame based binary classification on both RGB and YCbCr color spaces. The masks generated after processing these input images are then combined in a framework to classify background and foreground pixels. Comprehensive evaluation of proposed approach on publicly available test sequences show superiority of our system over other state-of-the-art algorithms.

KEYWORDS: Background Subtraction, Color Spaces, Binary Classifiers, Foreground Segmentation, Pixel Classification.

Author's signature: Hasan Sajid

Date: July 30, 2014

A UNIVERSAL BACKGROUND SUBTRACTION SYSTEM

By

Hasan Sajid

Director of Thesis: Sen-Ching S. Cheung

Director of Graduate Studies: Cai-Cheng Lu

Date: July 30, 2014

This work is dedicated to my family, especially my father, wife and daughter

ACKNOWLEDGEMENTS

First I would like to thank my advisor, Dr. Sen-Ching Samson Cheung, who has guided me research at the Multimedia Information Analysis lab. I had a wonderful learning experience in my thesis research. Next, I would like to thank to my thesis committee members, Dr. Kevin Donohue and Dr. Nathan Jacobs, for their time and valuable suggestions.

I would also like to thank my friends and all the MIA lab members who have supported me and helped me during my thesis research. They have always been willing to help me when I encountered difficulties.

Finally, I would like to thank my family for their unconditional support and encouragement. Without them, I would never have been able to finish my thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this chapter, we provide an introduction to background subtraction, its applications and our contribution. In first section, we give an overview and concept of background subtraction. Next, we briefly touch the applications and then outline our contributions in third section. The last section presents organization of the thesis.

1.1. Background

Background Subtraction (BS) is one of the most widely studied topics in computer vision. Typically, a BS process produces foreground (FG) binary mask given an input image and a background (BG) model.

BS is a difficult problem primarily because of diversity in background scenes and challenges that are linked to camera itself. Scene variations can be in many forms such as dynamic background, illumination changes, intermittent object motion, shadows, environmental conditions (rain, snow, night etc), highlights and camouflage to name a few [8]. Likewise the challenges linked to camera can be due to camera jitter, sensor noise and/or camera movement (pan, tilt and zoom) etc. The existing state-of-the-art techniques can address only a subset of these challenges but most are sensitive to illumination changes, camera/background motion and environmental conditions [22][23]. No single technique exists that is able to simultaneously handle all key challenges or produce satisfactory results if not accurate.

1.2. Applications

Background subtraction is a basic pre-processing step in video processing and therefore has numerous applications. One of the typical examples include traffic monitoring, where background subtraction algorithms has been widely used to monitor and control traffic flow by counting number of vehicles at different signals at different times of day.

Another example is video surveillance, tracking and privacy, in which the subject is being segmented out from the video using background subtraction algorithms for further processing. The shape of foreground mask produced by these algorithms can also be employed for human detection and gesture recognition.

These are only a few of many applications these algorithms offer but clearly indicates the need for a robust background subtraction algorithm.

1.3. Contribution of Thesis

In this thesis, 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 true foreground pixels from changing background pixels caused by scene variations or camera itself, we apply both pixel and frame level binary classification 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 real time universal background subtraction system with following major innovations: the background model, analysis and blending of RGB and YCbCr color spaces for BS and fusion of pixel and frame level Binary

Classifiers (BC). Another important contribution of this thesis is a comprehensive evaluation of ours and other state-of-the-art algorithms on a set of publicly available challenging sequences across 11 different categories totaling to 52 video sets. This is unlike the other algorithms in which authors tend to select certain metrics, choose or make test sequences of their own and compare with algorithms of their own choice. This makes overall comparison somewhat unfair and biased. The extensive evaluation of our system illustrate better foreground segmentation and superiority of our system in comparison with existing state-of-the-art approaches.

1.4. Organization

The rest of thesis is organized as follows: relevant work is discussed in chapter 2. The proposed system is detailed in chapter 3, followed by experiments and result comparison in chapter 4. The thesis is concluded in chapter 5.

Chapter 2

Related Work

There are a plethora of BS techniques, many of which are reviewed in surveys like [7], [11] and [21]. We can broadly divide these into 3 categories: pixel-based, region-based, and frame-based [1].

2.1. Pixel based techniques

Pixel-based algorithms are based on forming a statistical BG model for each pixel separately. Such algorithms are based on simple statistics such as mean value to complex multimodal distributions. The most simple techniques in this category include use of previous frame as background model, median value of pixels from a fixed number of recent images, running average and modeling of each pixel as a Gaussian to name a few [7][11].

Most of the techniques based on these simple statistics including unimodal Gaussian methods are very fast and computationally inexpensive but produce poor segmentation results due to complex real world scenarios such as camera noise, moving background, camera jitter, sudden illumination changes etc. The most popular techniques in pixel based category are pixel-wise Gaussian Mixture Model (GMM) [9] and kernel densities [10].

The GMM based techniques model per-pixel distribution of values observed overtime with mixture of Gaussians. The multimodal nature of these techniques allow them to cope with various real life challenges such as dynamic background. It has gained a lot of popularity and various improved versions have been presented in [21]. For example, in [37] authors take advantage of color and texture invariance and combine them with

GMM algorithm resulting in a more robust algorithm but it has proved to be computationally expensive and unsuitable for real time operation.

Another improvement to GMM algorithm is introduced in [38]. This improvement overcomes one of the problems of fixed number of components for GMM. In this scheme, 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.

Apart from GMM, many algorithms based on non-parametric kernel density estimates exist. Most popular techniques in this category are [10] and [12]. For each pixel, these methods accumulate values from pixel's recent history and then builds histogram of background values. The histogram is then used to classify that whether a pixel belongs to foreground or background. The kernel density estimates 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.

The pixel based algorithms in general suffer from loss of inter-pixel spatial dependencies and try to address this issue by constantly 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 caused by sudden variation in illumination or fast moving object.

Codebook [39] and [40] is another class of techniques that have been reported in literature. It comprises of codebook for each pixel and is basically a compressed form of background. Each codebook has codewords, which are formed based on a sequence of training images using a color distortion metric. Incoming pixels are matched against corresponding codebooks for classification. These techniques generally require long training sequences and do not have model update mechanism i.e. creation of new codewords if there is permanent change in scene[2].

2.2. Region based techniques

The second class of techniques are region-based techniques, which unlike their pixel based counterpart exploit local spatial relationships among pixels. In [12], authors present non-parametric kernel density estimate to model probability of foreground and background pixels but they include pixel location in the model. This is done using Maximum A Posteriori - Markov Random Field framework, which enforces spatial context among pixels. Although this method incorporates spatial information but the ability of these methods to handle events at various speeds raises the question of determining proper time interval for model update[2]. Another region based method is presented in [4], which uses statistical circular shift moments (SCSM) in image regions for change detection.

Apart from these, there are a number of region based techniques [2], [41] and [42] that take into account 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 using a sequence of images, which is followed by training a Principal Component Analysis (PCA) Model for each block. In [41], classification is done by comparing a block in current frame to its reconstruction from PCA coefficients and declared as background if observation is close. In contrast to [41], [42] performs classification using threshold based on difference between current image and back projection of PCA coefficients.

These techniques are more robust against noise and illumination changes in comparison to their pixel based counterparts but lack any update mechanism.

2.3. Frame based techniques

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 [1][13]. 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 [21].

Based on the shading model, Pilet *et al.*[3] proposed a method that makes use of the ratio of intensities between an input image and background image. The ratio of intensities are then modeled as a Mixture of Gaussians(MoG) resulting in a Statistical Illumination(SI) model. In this method, spatial dependence is also incorporated in the framework by learning a spatial-likelihood model. Although this technique is robust to global illumination changes, it is not able to handle local illumination changes [1].

Eigen background (EB) is a frame-based method that builds an eigenspace over expected illumination changes and reconstructs the BG image by projecting input image on the learned eigenspace [6]. 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 version by combining both EB and SI models in [1] at the expense of higher computation cost. EB reconstructs the BG image and then SI model segments the image into FG and BG regions. They also improved SI by introducing an online instead of an offline spatial-likelihood model.

Another, frame based technique is Tonal Alignment (TA) [14]. For an input image, change detection algorithm [15] is used to extract out BG pixels, subset of which are then used for histogram specification transform computation. The transformation then tonally aligns the input and background image. FG segmentation is done by pixel wise comparison of input and tonally aligned background image. TA is able to handle global illumination changes but fails to deal with local lighting changes.

Apart from these, there exists methods [17][18] in literature that take advantage of illumination invariant features such as texture with edge or color and combine them for reconstructing BG images but they suffer from the possibility of texture absence in certain areas of image or poor color discrimination in low lighting conditions.

Chapter 3

Background Subtraction System

Background Subtraction can be generalized as a four step process: preprocessing, background modeling, foreground detection, and data validation. Preprocessing involves simple image processing on input video such as format conversion, image resizing etc for subsequent steps. Background modeling is responsible for constructing a statistical model of the scene, which is followed by pixel classification in foreground detection step. The final step, data validation removes the falsely detected foreground pixels and outputs the final foreground mask [7].

Two of our innovations are related to background modeling and foreground detection steps and therefore we focus more on these two steps. For better understanding of the system, first we discuss aforementioned innovations and their motivation. Then, we provide a general overview of system as how these contributions when combined together result in a robust universal BS system. Next we detail the system parameter settings and their sensitivity. Lastly, we briefly discuss system's real time performance.

3.1. Fusion of RGB and YCbCr Color Spaces

The choice of color space is very critical to accuracy of foreground segmentation. Different color space including RGB, YCbCr, HSV, HSI, lab2000, normalized-RGB (rgb) have been employed by existing state of the art techniques. Among these color spaces, we focus on the most widely used color spaces for background subtraction: RGB, YCbCr, HSV and HSI [28][31].

RGB has been an automatic choice because, (a) first, the brightness and color information is equally/uniformly distributed in all of the 3 channels, (b) second, it handles noise well [28] (b) third, it is the output format of most devices and camera and (c) fourth, it is computationally inexpensive in comparison to other color spaces[31].

The remaining three color spaces YCbCr, HSV and HSI differ from RGB and are motivated by human visual system, which tends to assign a constant color to an object even under changing illumination over time or space [28][29]. These color spaces segregate the brightness and color information, which makes these color spaces more robust against noise, shadow, highlights and illumination changes but at the cost of loss of brightness information [27][28][29][30][32]. YCbCr uses Cartesian coordinates whereas HSV and HSI color spaces use polar coordinates.

In comparative studies on color spaces [27][28][30][31], YCbCr has been proven to overall outperform RGB, HSI and HSV color spaces and considered most suitable color space for foreground segmentation [28][30][31]. YCbCr is least sensitive to noise due to independent color channels followed by RGB, while HSI and HSV are affected by noise due to their polar coordinate description [28]. YCbCr is second to RGB in terms of computational cost. For shadow and highlights, [27], [29] and [31] clearly indicate the superiority of YCbCr in handling shadow and illumination changes in comparison to other color spaces.

In YCbCr, all 3 channels are independent of each other. Y channel represents the luminance whereas Cb & Cr channels represent chrominance. RGB and YCbCr color spaces are related by eq 3.1 as in [28]:

$$Y = 0.257R + 0.504G + 0.098B + 16,$$

$$Cb = -0.148R - 0.291G + 0.439B + 128,$$

$$Cr = 0.439R - 0.368G - 0.071B + 128 \tag{3.1}$$

Based on the above comparison, YCbCr becomes a natural choice for segmentation purposes but , [29] and [30] identifies problematic behavior of YCbCr color space, when current image contains very low RGB values. The misclassification chances of dark pixel increases since dark pixels are close to the origin in RGB space and the fact that all chromaticity lines in RGB space meet at the origin, thus the color point is considered to be close or similar to any chromaticity line. It is not necessary that such scenario occur only when illumination levels are low globally but rather it is also true when an image is affected partially or certain portions of image are darker. In real life videos, this is common depending on position of illumination sources and scene geometry. Shadows casted by objects is one such example. This results in decrease in foreground segmentation accuracy.

In order to address this issue, we propose to use two color spaces; YCbCr and RGB. This is contrary to all existing techniques that employ only one color space. The use of two color spaces is motivated by human visual system. The human visual system provides color vision by using two types of cells; rods and cones. Rods are used for vision in low light levels known as *scotopic*, in which color vision is not possible. At intermediate light levels(0.01 - 1 cd/m^2), our vision is *mesopic*, in which both rods and cones are active. In *mesopic* light conditions color discrimination is poor. At high levels or above (>1cd/m^2), our vision becomes *photopic*, where cone activity is best and allows for good color discrimination [26].

Like the human visual system, in which the cells for different lighting levels work together, we employ RGB and YCbCr color spaces. RGB color space plays its part under poor lighting conditions since chromatic information is uniformly distributed across RGB channels, Whereas under sufficient lighting condition, the incorporation of illumination invariant channels (Cb and Cr) provide a robust BG/FG classification. Hence, the two color spaces complement each other resulting in better foreground segmentation.

To support our claim, a detailed quantitative analysis is presented in section 4 by comparing segmentation accuracy when the two color spaces are used together and when both are tested individually.

3.2. Background Modeling

BG modeling is a very crucial and one of most important steps in a BS process. We are convinced that if the model being built is accurate, even weak binary classifiers can produce comparable segmentation results. The most widely adapted BG modeling approaches, in its basic form is to build a multi-modal pixel-wise statistical background model. Such approaches suffer due to 2 reasons; first, it is difficult to determine the correct number of modes for modeling 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 follow the conventional approach but we retain the inter-pixel spatial dependencies and build a more simpler single-mode instead of multi-modal pixel wise model. More specifically, we build a Background Model Bank(BMB) comprising of multiple BG models instead of a single BG model. To form BMB, each training image is treated as a vector. All images are then grouped together into N clusters using the K-means algorithm. This is followed by an estimation of a pixel-wise single-mode Gaussian model $\left(\mu_{D_n}(X), \sigma_{D_n}^2(X)\right)$ for each color component D of each cluster n. It should be noted that component means one of the color channels (R, G, B, Y, Cb and Cr) of color spaces.

The concept of multiple BG models allow us to capture scene more accurately while keeping spatial dependencies intact. Another important aspect is that it is computationally comparable to conventional approach, since for classification, first we choose a model at frame level and then for pixel wise comparison, choice of probability distributions are restricted to chosen model while remaining BG models are ignored. The only additional cost is choosing the model at frame level but at the same time computational cost is reduced since each pixel is represented by single-mode instead of multi-mode distribution.

The approach of multiple BG models has proved to capture scene diversity and camera variations more robustly and allowed us to employ simple binary classifiers for pixel classification in comparison to complex and multi-modal techniques. This is evident from the results on various challenging test sequences.

3.3. Binary Classifiers

In our proposed scheme, we use three different types of Binary Classifiers(BC) based on the type of comparison and how they are thresholded. In a typical BS process, this step is known as foreground detection. The details of BCs are described in this section.

3.3.1. Pixel Level Binary Classifier(PBC)

This BC performs a pixel-level comparison between each pixel in each color channel with the corresponding BG pixel distribution from the chosen BMB model. The threshold is also based solely on the pixel distribution itself. Specifically, for pixel classification we have eq 3.2.

$$D_{mask_1}(X) = \left[\left|I_D(X) - \mu_{D_n}(X)\right| \geq c_{PW} \,.\, \sigma^2_{D_n}(X)\right] \tag{3.2}$$

where $I_D(X)$ is the input image, $\left(\mu_{D_n}(X), \sigma^2_{D_n}(X)\right)$ is the chosen BMB model, and c_{PW} is a parameter as discussed in system parameters section.

3.3.2. Pixel and Frame Level Binary Classifier(PFBC)

This BC is a hybrid approach in which comparison is done on pixel level but threshold is calculated at frame-level. The motivation behind this approach is that estimation of threshold at frame level allows us to account for inter-pixel dependencies and produces more accurate masks in comparison to if pixels are thresholded independently. For this purpose, we consider the global spatial statistics of the average BG frame given by eq 3.3 and eq 3.4:

$$\mu_n = \frac{1}{|X|} \sum_{x \in X} \mu_{D_n}(x) \tag{3.3}$$

and

$$\sigma_n = \sqrt{\frac{1}{|X|-1} \sum_{x \in X} \left(\mu_{D_n}(x) - \mu_n\right)^2} \tag{3.4}$$

and compute the foreground mask based on the deviation from this global statistics using eq 3.5:

$$D_{mask_2}(X) = [|I_D(X) - \mu_n(X)| \geq c_{SI} \cdot \sigma_n] \tag{3.5}$$

c_{SI} is a parameter as discussed in system parameters section.

3.3.3. Frame Level Binary Classifier(FBC)

This BC performs both comparison and thresholding at frame-level. It uses the Quotient Image $QI_D(X)$ and its inverse $IQI_D(X)$ between the input and the BG average given by eq 3.6 and eq 3.7:

$$QI_D(X) = \frac{I_D(X)}{\mu_{D_n}(X)} \tag{3.6}$$

and

$$IQI_D(X) = \frac{1}{QI_D(X)} \tag{3.7}$$

The motivation behind this BC is twofold: first is to keep pixel spatial dependencies intact and secondly, we assume that the chosen BG model is ideally the same as input image and therefore, BG pixels should have their QI and IQI values equal to 1 while remaining pixels represent FG. We use both QI and IQI so as to avoid using both a lower and an upper threshold. As the majority of the pixels are background, the mean value provides a good lower threshold to identify foreground pixels given by eq 3.8 and eq 3.9:

$$D_{mask_3}(X) = [QI_D(X) \geq c \cdot \langle QI_D(X) \rangle] \tag{3.8}$$

and

$$D_{mask_4}(X) = [IQI_D(X) \geq c \cdot \langle IQI_D(X) \rangle] \tag{3.9}$$

where c is introduced to account for non-idealities between input image and average BG model and discussed in system parameters section.

3.3.4. Discussion on Binary Classifiers

Based on comprehensive analysis and evaluation, each of the BC has its own strengths and weaknesses. Let TP = True positive, TN = True negative, FP = False Positive and FN = False Negative.

PBC is most accurate in terms of TP and FN but results in most number of FP and least TN. PFBC produces less TP and FN than PBC but at same time is more accurate by producing less FP and more TN than PBC. FBC produces least TP and most FN but is most accurate in terms of FP and TN. In terms of accuracy, we can easily conclude the superiority of PFBC over other two BCs, followed by FBC over PBC. The superiority of both PFBC and FBC over PBC also emphasizes the importance of inter-pixel spatial dependencies.

In totality, when these simple and computationally inexpensive BCs are treated independently and then combined together in a framework, they complement each other well to produce more accurate foreground mask. This is contrary to all existing techniques in which a single but complex classifier is used.

3.4. Process Overview

In this section, we give an overview of the complete system and how different components of system work together. The proposed system consists of seven steps as shown in Figure 3.1. The first two steps are a part of training phase involving training images denoted by M. Training images are expected to comprise of scene and camera variations. We split training images into two sets and then use them for BMB formation and parameter training. Each step of proposed system is detailed below.

Step 1: BMB Formation

The first step is to build a BMB comprising of N number of background models from training images. For BMB, it is not necessary that training images are foreground free. This is one of fundamental requirements for any BS algorithm, since for every scene, it is impossible to obtain training images without foreground. The criteria to set N is discussed in parameter setting section.

Step 2: Automatic Parameter Training

The proposed system requires a number of parameters. In order to determine these parameters, a number of training images with foreground and their respective ground truths are used. The details of these parameters and the criteria on how we set them is detailed in system parameters section. Once the BMB is formed and optimal parameters are determined, the training phase ends.

Step 3: Component BG Model Selection

Next step is to select an appropriate BG Model for each of the color channels of an input image. The selection criterion is based on minimizing the total error given by eq 3.10:

$$E = \arg\min_{n=1,\ldots,N} \ (I_D(X) - \mu_{D_n}(X))^2 \tag{3.10}$$

where D represents each of the color channels and $I_D(X)$ is the input frame.

Step 4: Components Mask Generation

In this step, the color channels and their respective selected BG models are passed onto BCs, each generating a respective binary component mask. We denote the foreground mask for color channel D generated by the k^{th} BC as $D_{mask_k}(X)$.

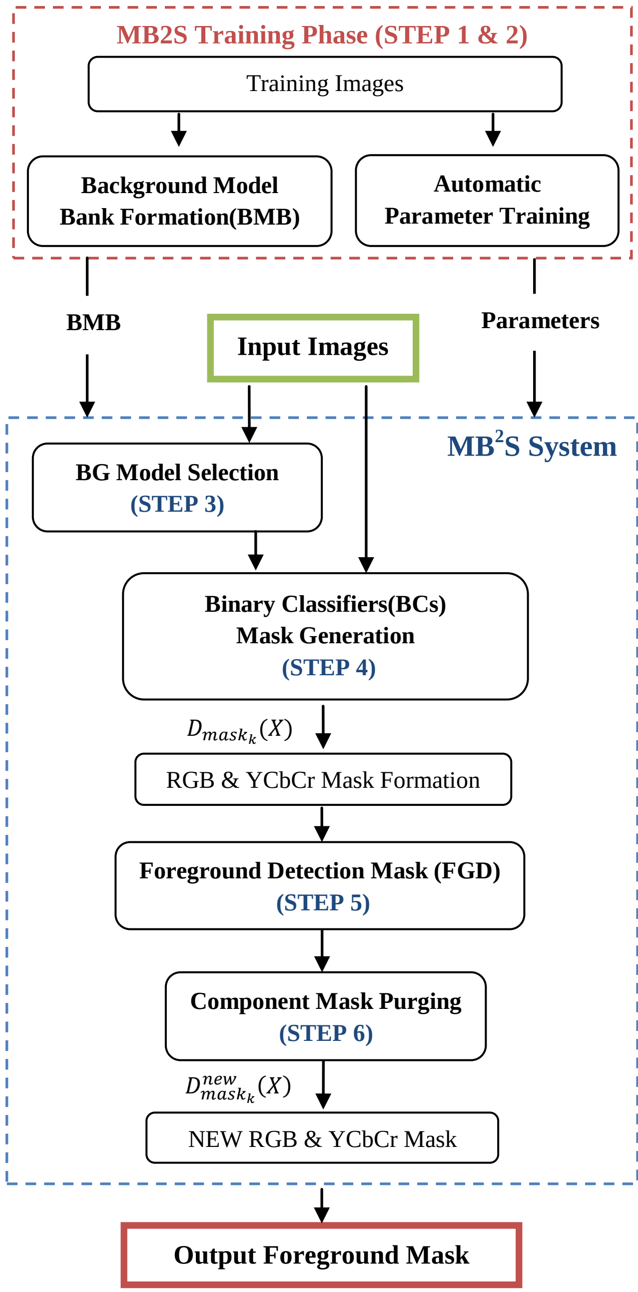


Figure 3.1. MB^2S Background Subtraction System.

Step 5: Foreground Detection Mask

Following the generation of component masks, we aggregate the component masks using eq 3.11 and eq 3.12:

$$RGB_{mask}(X) = \left[\sum_{D=R,G,B}(D_{mask_k}(X)) > \tau_{RGB}\right] \tag{3.11}$$

$$YCbCr_{mask}(X) = \left[\sum_{D=Y,Cb,Cr}(D_{mask_k}(X)) > \tau_{YCbCr}\right] \tag{3.12}$$

The aggregation computes an overall likelihood on whether a pixel belongs to FG or BG based on the results of all BCs. The procedure to determine τ_{RGB} and τ_{YCbCr} is described in system parameters section. We then dilate these masks and multiply them to obtain the Foreground Detection (FGD) mask given by eq 3.13:

$$FGD_{mask}(X) = \text{Dilate}(RGB_{mask}(X)) \cdot \text{Dilate}(YCbCr_{mask}(X)) \tag{3.13}$$

The FGD mask is not the final mask. The relaxed thresholds and the dilation are to ensure that all true foreground pixels are captured in the FGD mask.

Step 6: Component Mask Purging

The FGD mask is then applied to each of the component masks 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 given by eq 3.14:

$$D_{mask_k}^{new}(X) = D_{mask_k}(X) \cdot \ FGD_{mask}(X) \tag{3.14}$$

Step 7: Output FG Mask

In the final step of the process, all of the $D_{mask_k}^{new}(X)$ are combined to form $RGB_{mask}^{new}(X)$ and $YCbCr_{mask}^{new}(X)$ masks given by eq 3.15 and eq 3.16.

$$RGB_{mask}^{new}(X) = \left[\sum_{D=R,G,B} (D_{mask_k}^{new}(X)) > t_{RGB}\right] \tag{3.15}$$

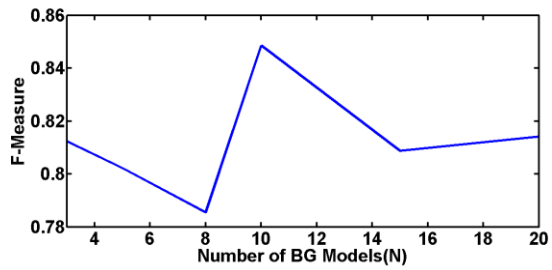
$$YCbCr_{mask}^{new}(X) = \left[\sum_{D=Y,Cb,Cr} (D_{mask_k}^{new}(X)) > t_{YCbCr}\right] \tag{3.16}$$

The thresholds t_{RGB} and t_{YCbCr} differ from those in step 5 in that this final step provides tighter thresholds to eliminate BG pixels that are erroneously included in previous step. The final output FG mask is simply obtained by the logical OR of these two masks.

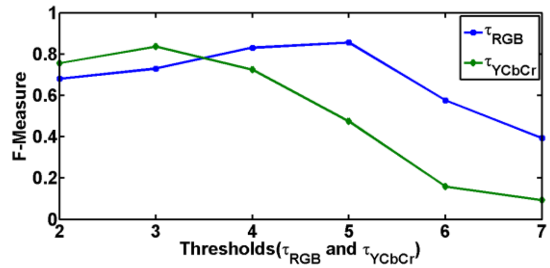
3.5. System Parameters

The system has six parameters; N, c_{PW}, c_{SI}, c, τ_{RGB} and τ_{YCbCr}. Among the six parameters, τ_{RGB}, τ_{YCbCr} and N are most critical in order of their importance. The evaluation on over 50 test sequences has show very little variance for remaining three parameters and use of default values are recommended. The default values are also discussed in this section.

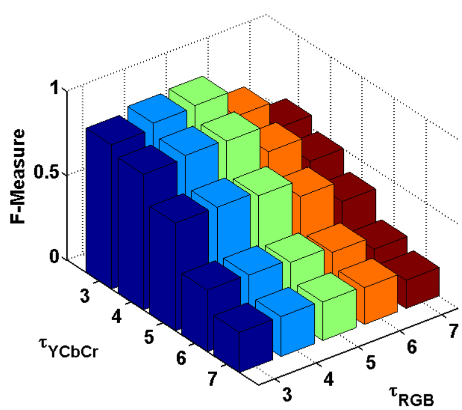
In order to understand the impact of τ_{RGB}, τ_{YCbCr} and N on segmentation accuracy, we present a detailed analysis of these parameters on one of the test sequences using F-measure metric. The F-measure metric is defined in chapter 4. Figure 3.2 depicts the variation of F-measure against N, τ_{RGB} and τ_{YCbCr} while fixing remaining parameters to default values.



(a)



(b)



(c)

Figure 3.2. System Parameter Variations. (a) F-Measure variation with number of BG models (N). (b) F-Measure variation with τ_{RGB} and τ_{YCbCr} individually. (c) F-Measure variation with both τ_{RGB} and τ_{YCbCr} combined.

It would be natural to think of a linear relationship between N and F-measure but looking at Figure 3.2a reveals that such a relationship does not hold and hence it is important to determine N for any given scenario. Furthermore, increasing N beyond certain limit deteriorates the F-measure. Similar to N, optimal value of both τ_{RGB} and τ_{YCbCr} are dependent on the scene. In general, increasing both the values increases F-measure but after a certain value the F-measure plummets. Another important observation is that optimal setting of τ_{YCbCr} is lower than that of τ_{RGB}. In Figure 3.2c, we illustrate the effect on F-measure, when τ_{RGB} and τ_{YCbCr} are varied i.e. the case when both color spaces are used. In this case, it is important to note that determining the values of τ_{RGB} and τ_{YCbCr} independently does not necessarily result in optimal combination and therefore may affect the segmentation accuracy. Thus, it is important to find the right combination of τ_{RGB} and τ_{YCbCr}.

In order to set these parameters, we devise three types of parameter settings; default, optimal and standard. Each type of setting has its own advantages and disadvantages.

3.5.1. Default Setting

In this mode, the system does not run the Automatic Parameter Training step rather default parameter values are used. This option allows quick system deployment and also removes the hassle of training step but may not produce optimal results. The default system parameters are: $N=15$, $c_{PW} = 350$, $c_{SI}=0.75$, $c=1.2$, τ_{RGB}=5, and τ_{YCbCr}= 4. These settings are based on results obtained for more than 50 videos sequences of diverse nature and has proved to perform well.

3.5.2. Optimal Setting

This option runs the Automatic Parameter Training step and determines the setting of parameters that produce optimal results. In order to determine the parameters N, c_{PW}, c_{SI},

c, τ_{RGB}, and τ_{YCbCr}, a number of training images with foreground and their ground truth are passed on to the algorithm. A three step procedure is then followed:

Step 1: The number of background models N in BMB ranges from 2 to 50. For each N, all of the BCs are run over training images to determine c_{PW}, c_{SI} and c parameters. Since, the BCs are independent of each other, their parameters are determined exclusively based on the masks generated from their respective BS modules. For each value of c_{PW}, c_{SI} and c, the masks generated from respective BCs are evaluated using F-measure metric. The value of c_{PW}, c_{SI} and c that results in highest F-measure is chosen and set.

Step 2: After fixing c_{PW}, c_{SI} and c, we determine τ_{RGB} and τ_{YCbCr} parameters. Both threshold values are varied in combination and similar to c_{PW}, c_{SI} and c, the value of τ_{RGB} and τ_{YCbCr} that results in highest F-measure is chosen.

Step 3: The parameters t_{RGB} and t_{YCbCr} are simply one less than τ_{RGB} and τ_{YCbCr} respectively.

Once the procedure is completed for all possible values of N, the optimal setting of parameters is determined by simply choosing the combination of parameters that offers the highest F-measure over training images. This completes the parameter setting and training.

3.5.3. Standard Setting

In this type of setting, all parameters are same as default settings except τ_{RGB} and τ_{YCbCr}. Both of these parameters are determined using the step 2 of procedure outlined in optimal setting. This is done by simply fixing the other parameters to default values and then passing the training images and their respective ground truth. By considering two most critical parameters; τ_{RGB} and τ_{YCbCr}, this setting offers quick but more accurate foreground segmentation than default setting. Table 3.1 reports the typical range of each of the parameters we found over 53 test sequences.

Table 3.1. Typical range of parameters.

Parameter	Minimum Value	Maximum Value	Typical Value
N	3	50	15
M	100	1000	300
c_{PW}	300	400	350
c_{SI}	0.6	0.98	0.75
c	1	1.5	1.2
τ_{RGB}	4	7	6
τ_{YCbCr}	3	6	4

3.6. Real time operation

The proposed system is currently implemented in Matlab and with default settings, it is able to achieve real time operation with 11 fps for images with a resolution of 320 x 240 on an Intel core i7 PC with 16GB RAM.

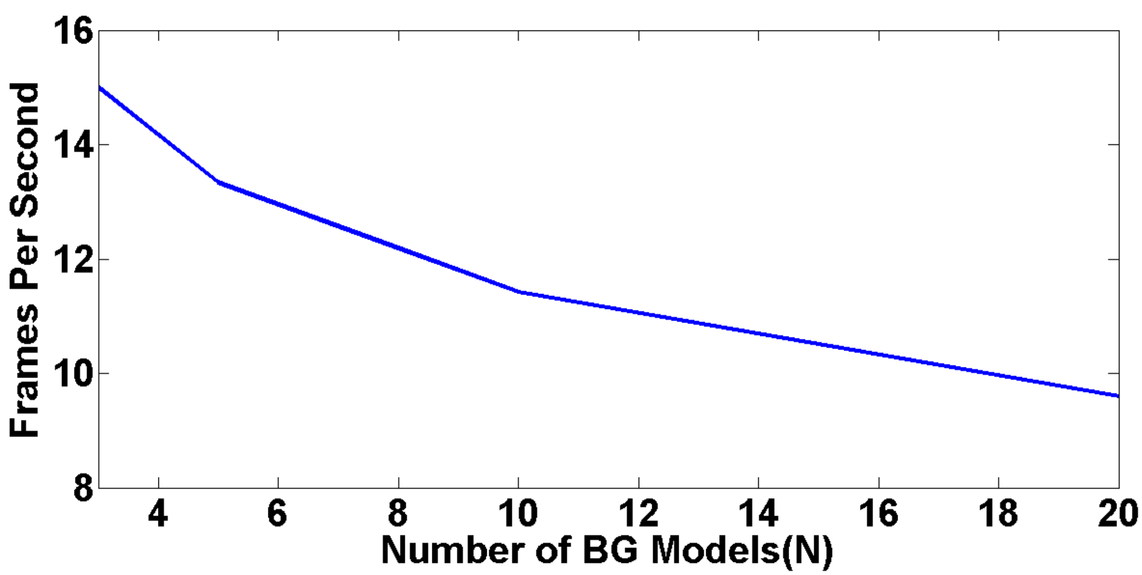


Figure 3.3. Variation of Frames Per Second(FPS) with increase in Number of BG Models (N).

Figure 3.3 shows how the system performance is affected with increase in number of background models N. Clearly, there is a decrease in FPS with increasing N but it does not drops down sharply. With implementation in C/C++, naturally the processing time is further going to be decreased resulting in increase in FPS.

In addition to this, although system uses 2 color spaces; RGB and YCbCr but for simple scenes or based on user choice, only one of the color spaces can be used. This will cut down the number of per-pixel operations by half and therefore significantly increase the speed but at the cost of segmentation accuracy.

Chapter 4

Experiments and Results

In this chapter, we compare proposed system with existing state of the art algorithms on publicly available test sequences. Two type of datasets are included; changedetection[22][23] and ESI[16].

The reason to choose these datasets is to avoid bias unlike other techniques and also offer comprehensive insight of strengths and weaknesses of our system. The two datasets, parameter setting and evaluation criteria are detailed in following sections:

4.1. Changedetection dataset

The changedetection dataset[23] is one of the most comprehensive datasets available for evaluating BS algorithms and has become a defacto standard. It comprises of 11 categories: Baseline(BL), Dynamic Background(DB), Camera Jitter(CJ), Intermittent Object Motion(IOM), Shadow(S), Thermal(TH), Bad Weather(BW), Low Framerate(LFR), Night Videos(NV), Pan Tilt Zoom(PTZ) and Turbulence(TB). Each category comprises of 4 to 6 videos and they total to 53 videos sequences. For details of these categories we refer authors to its official webpage at [23].

4.1.1. Evaluation Metrics

For fair comparison, we use metrics recommended by authors in [22][23]. Seven different metrics have been used:

1. $Recall\ (Re)\colon \frac{TP}{TP+FN}$
2. $Specificity(Sp)\colon \frac{TN}{TN+FP}$

3. $False\ Positive\ Rate(FPR)$: $\frac{FP}{FP+TN}$

4. $False\ Negative\ Rate(FNR)$: $\frac{FN}{FP+TN}$

5. $Percentage\ of\ Wrong\ Classifications(PWC)$: $100 * \frac{(FN+FP)}{(FN+FP+TN+TP)}$

6. $Precision(Pr)$: $\frac{TP}{TP+FP}$

7. $F - measure$: $2 * \frac{Pr.Re}{Pr+Re}$

For each video in a category, the seven metrics are calculated and then average value of each metric for all videos in that category is calculated. This is followed by calculating average rank which is defined as:

$$Average\ ranking: (rank{:}Re + rank{:}Sp + rank{:}FPR\ +\ rank{:}FNR\ +\ rank{:}PWC \\ +\ rank{:}FMeasure\ +\ rank{:}Pr)/7$$

4.1.2. Parameter Setting

In this section we discuss the parameter setting of test sequences for all of 10 categories. Table 4.1 presents the parameter settings for all test sequences. These parameters are calculated based on standard parameter setting procedure explained in System Parameters section. For each test sequence, we use 20 images with foreground and their respective ground truths from training images. Although more than 20 images could have been used for determining the optimal parameter setting but to simplify the evaluation procedure over large dataset, we limit the number to 20. For details of parameter used by other techniques, we refer readers to the website at [23].

Table 4.1. Parameter values for Changedetection Dataset.

Category	**Test Sequence**	**MB^2S-RGB**	**MB^2S-YCbCr**	**MB^2S-Standard**	
		τ_{RGB}	τ_{YCbCr}	τ_{RGB}	τ_{YCbCr}
BL	Highway	7	5	7	5
	Office	7	5	7	5
	Pedestrians	7	4	7	4
	PETS2006	7	5	7	5
DB	Boats	5	5	5	4
	Canoe	4	5	4	4
	Fountain01	6	4	6	4
	Fountain02	7	5	7	4
	Overpass	6	5	6	4
	Fall	5	4	5	4
CJ	Badminton	5	4	5	4
	Boulevard	5	4	5	4
	Sidewalk	3	3	3	3
	Traffic	5	4	5	4
IOM	AbandonedBox	6	5	5	4
	Parking	5	4	5	4
	Streetlight	5	4	5	5
	Sofa	5	5	5	4
	Tramstop	5	4	5	4
	winterDriveway	6	6	5	4
TH	Corridor	7	4	7	4
	Library	5	4	4	4
	park	6	4	5	4
	diningRoom	6	6	6	4
	lakeSide	4	3	4	3
BW	Blizzard	5	5	5	5
	Skating	5	5	5	5
	snowFall	5	5	5	5
	wetSnow	5	5	5	5
LFR	Port	6	4	6	4
	tramCrossroad	5	4	5	4
	tunnelExit	5	4	5	4
	Turnpike	4	4	4	4
NV	bridgeEntry	4	4	5	3
	busyBoulvard	4	4	5	3
	fluidHighway	6	5	6	5
	streetCorneratNight	5	4	5	4
	tramStation	6	5	6	5
	winterStreet	5	4	5	4
PTZ	continuousPan	6	5	6	5
	intermittentPan	6	5	6	5
	twoPositionPTZ	5	4	5	4
	zoomInzoomOut	5	5	6	5
T	Turbulence0	6	4	6	4
	Turbulence1	6	4	6	4
	Turbulence2	6	4	6	4
	Turbulence3	6	3	6	3

4.1.3. Quantitative Evaluation

In this section, we compare proposed system with existing state of the art algorithms. [23] presents a detailed comparison of 14 state of the art algorithms on changedetection dataset. It contains overall as well as results for individual categories. Here, we only mention top 5 techniques but based on the statistics available on [23], comparison is done against all 14 algorithms.

Table 4.2. MB^{2}S Quantitative Evaluation.

Method	Category	Avg rank	Avg Re	Avg Sp	Avg FPR	Avg FNR	Avg PWC	Avg F-Measure	Avg Pr	Rank
MB^{2}S-Standard	BL	3.71	0.9396	0.9977	0.0023	0.0605	0.4587	0.9371	0.9348	**3**
MB^{2}S-default		8.85	0.9822	0.9914	0.0086	0.0179	0.8885	0.8974	0.8304	8
MB^{2}S-RGB		7.57	0.8692	0.9974	0.0026	0.1308	0.7449	0.8944	0.9224	7
MB^{2}S-YCbCr		9.42	0.9438	0.9948	0.0052	0.0562	0.6874	0.8823	0.8312	9
MB^{2}S-Standard	CJ	2.14	0.8318	0.9921	0.0079	0.1682	1.4641	0.8394	0.8503	**1**
MB^{2}S-default		2.28	0.7718	0.9923	0.0077	0.2282	1.5956	0.8046	0.8561	1
MB^{2}S-RGB		4.57	0.7601	0.9873	0.0127	0.2399	2.2404	0.7709	0.7944	4
MB^{2}S-YCbCr		2.85	0.7849	0.9917	0.0083	0.2151	1.6906	0.8457	0.7931	3
MB^{2}S-Standard	BW	10.42	0.7126	0.9936	0.0064	0.2874	1.0964	0.6995	0.7078	**13**
MB^{2}S-default		11.42	0.8095	0.9857	0.0143	0.1905	1.6956	0.6242	0.5177	14
MB^{2}S-RGB		10.14	0.7264	0.9938	0.0062	0.2736	1.0797	0.7062	0.7234	13
MB^{2}S-YCbCr		12.85	0.6127	0.9938	0.0062	0.3873	1.2317	0.6355	0.6862	14
MB^{2}S-Standard	DB	4.71	0.8080	0.9989	0.0011	0.1921	0.3434	0.7862	0.8002	**4**
MB^{2}S-default		5.71	0.8750	0.9950	0.0050	0.1250	0.6520	0.7346	0.7007	5
MB^{2}S-RGB		8	0.6642	0.9987	0.0013	0.3358	0.5065	0.7067	0.8030	7
MB^{2}S-YCbCr		7.71	0.7482	0.9970	0.0030	0.2529	0.5560	0.7459	0.7512	7
MB^{2}S-Standard	IOM	7	0.8134	0.9460	0.0540	0.1866	5.7699	0.6194	0.5681	**4**
MB^{2}S-default		7	0.8134	0.9460	0.0540	0.1866	5.7699	0.6194	0.5681	4
MB^{2}S-RGB		7.57	0.7391	0.9577	0.0423	0.2609	5.2722	0.6135	0.5787	6
MB^{2}S-YCbCr		7.57	0.8052	0.9347	0.0653	0.1948	6.7623	0.5986	0.5361	6
MB^{2}S-Standard	LFR	4	0.6749	0.9971	0.0029	0.3250	1.2204	0.6611	0.6981	**1**
MB^{2}S-default		5.14	0.6834	0.9968	0.0032	0.3166	1.3981	0.6205	0.6803	3
MB^{2}S-RGB		5.28	0.6712	0.9958	0.0042	0.3288	1.1915	0.6512	0.6729	3
MB^{2}S-YCbCr		7.28	0.6434	0.9947	0.0053	0.3566	1.7530	0.5662	0.6294	7
MB^{2}S-Standard	NV	6.28	0.5652	0.9786	0.0214	0.4348	2.9718	0.4235	0.3664	**4**
MB^{2}S-default		7.71	0.4536	0.9852	0.0148	0.5464	2.5945	0.3905	0.3931	7
MB^{2}S-RGB		5.28	0.6153	0.9771	0.0229	0.3847	3.0424	0.4389	0.3630	4
MB^{2}S-YCbCr		6.71	0.5463	0.9783	0.0217	0.4537	3.1147	0.3980	0.3502	4
MB^{2}S-Standard	PTZ	4.85	0.6744	0.9430	0.0570	0.3254	5.9919	0.3630	0.3491	**3**
MB^{2}S-default		5.14	0.8283	0.8882	0.1118	0.1718	11.304	0.2831	0.2360	4
MB^{2}S-RGB		6.14	0.6226	0.9283	0.0717	0.3774	7.5276	0.3310	0.3516	4
MB^{2}S-YCbCr		6.28	0.7140	0.8426	0.1573	0.2859	15.758	0.0187	0.0095	5
MB^{2}S-Standard	T	9	0.8547	0.9755	0.0245	0.1453	2.8995	0.7445	0.6681	**10**
MB^{2}S-default		9.42	0.8244	0.9681	0.0319	0.1756	3.5358	0.7030	0.6469	11
MB^{2}S-RGB		9.71	0.8018	0.9738	0.0262	0.1982	3.2405	0.6998	0.6417	12
MB^{2}S-YCbCr		9.71	0.9294	0.9574	0.0426	0.0706	4.3145	0.6687	0.5352	12
MB^{2}S-Standard	TB	8.42	0.7420	0.9870	0.0130	0.2580	1.4230	0.5517	0.5630	**8**
MB^{2}S-default		10.14	0.7047	0.9851	0.0149	0.2952	1.7196	0.5062	0.5127	12
MB^{2}S-RGB		11.42	0.6228	0.9890	0.0110	0.3772	1.3817	0.3609	0.3192	15
MB^{2}S-YCbCr		9.57	0.8244	0.9830	0.0170	0.1756	1.7575	0.5134	0.4653	11

Table 4.2 details complete results of our algorithm for four different cases: standard parameter setting, default parameter setting, using RGB and YCbCr color spaces individually. These are denoted by MB^2S-Standard, MB^2S-default, MB^2S-RGB and MB^2S-YCbCr respectively. Such a comparison will later help us to analyze the robustness that is offered by using both color spaces together and separately.

In 7 out of 10 categories, the proposed system is placed among top 4 with 1st position in CJ and LFR categories. Table 4.3 presents an overall comparison of proposed system with top 5 algorithms; Flux Tensor with Split Gaussian models(FTSG)[24], suBSENSE[25], CwisarDH[33], Spectral-360[34] and Bin Wang Apr 2014 [35]. For overall comparison, Average Ranking across all Categories(ARC) is calculated, which is given by eq 4.1:

$$ARC\text{:}\frac{Sum\ of\ ranks\ for\ all\ categories}{number\ of\ categories} \tag{4.1}$$

Our proposed system achieves an ARC of 5.1 and is placed at 4th position out of 14 existing state of the art algorithms. Also note that with default setting or use of one color space, the system is placed at 6th position. Note that this position is out of 14 existing state of the art.

Table 4.3. Overall Comparison.

Method	**ARC**	**Position**
MB^2S-Standard	5.1	4
MB^2S-default	6.9	6
MB^2S-RGB	7.5	6
MB^2S-YCbCr	7.8	6
FTSG[24]	2.1	1
suBSENSE[25]	2.7	2
CwisarDH[33]	4.4	3
Spectral-360[34]	5.2	5
Bin Wang Apr 2014[35]	6.8	6

4.1.4. Qualitative Results

Figure 4.1 presents some sample results of proposed system for different categories of changedetection dataset. Complete set of results for all categories will be made available at our website [36].

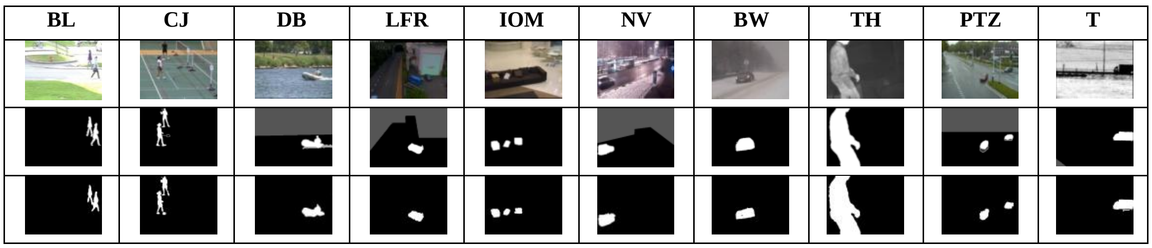


Figure 4.1. Foreground Segmentation results of example frames from changedetection dataset. (top row) input image, (middle row) ground truth and (bottom row) MB^{2}S-standard segmentation result.

4.1.5. Discussion on Results

There are several key points that are highlighted through comprehensive evaluation and results. First, the results support the claim that using RGB and YCbCr color spaces produce more accurate results in comparison to when they are employed individually. This is clear from overall higher position and ARC of both MB^{2}S-Standard and MB^{2}S-default than MB^{2}S-RGB and MB^{2}S-YCbCr.

Second important point is the performance of MB^{2}S-RGB and MB^{2}S-YCbCr for Night Videos(NV) and Bad Weather category. NV has low lighting conditions and BW has poor color discrimination problem. From Table 4.3, we can see that MB^{2}S-RGB has an average ranking of 5.28, which is not only higher than that of 6.71 of MB^{2}S-YCbCr but also higher than 6.28 and 7.71 of MB^{2}S-Standard and MB^{2}S-default, respectively. Likewise for BW, MB^{2}S-RGB has higher average ranking than other three. This supports our earlier claim that RGB is more robust under low lighting conditions or when color discrimination is poor.

Third, in general, RGB performs well in simple background scenes with minimum noise, whereas YCbCr is more robust against noise. This is evident from higher average ranking of MB^2S-RGB in categories such as BL and LFR and higher average ranking of MB^2S-YCbCr in CJ and DB categories.

Fourth, despite use of standard parameter setting and fixed number of BG models N, our proposed system performs well in 7 categories, whereas it performs poorly in 3 categories; Thermal, Bad Weather and Turbulence. This has resulted in overall position to drop down to 4th. The main reason is that in some of video sequences the scene changes over time and model update is required. This is lacking in our current system and has resulted in poor performance in aforementioned categories. For example in case of thermal, in one of video sequences when a person sitting on a chair stands up after a while and leaves, the higher temperature of chair results in misclassifications of chair as foreground. Another example is from one of test sequences in bad weather in which when snow is cleared from pathway, it becomes foreground and remains foreground since model is not updated resulting in poor performance.

Lastly, with the incorporation of BG model update and use of optimal parameter setting, the proposed system is expected to outperform the existing state of the art on all categories.

4.2. 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 occurs often such as opening and closing of door, switching light on and off etc. Changedetection dataset lacks any such category, therefore, we include ESI dataset and instead of comparing with general BS algorithms, we compare its performance with algorithms that specialize in dealing with this challenge. In our opinion, ESI dataset[16] is the most challenging publicly available test dataset in terms of sudden illumination changes.

ESI dataset comprises of 5 of test sequences; sofa, walking, chair, scene1 and scene2 [1]. They have 382, 734, 573, 750 and 154 frames respectively. For evaluation purposes, 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.

4.2.1. Evaluation Metrics

For quantitative evaluation of ESI dataset, we use three metrics as defined earlier; precision, recall 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.

4.2.2. Parameter Setting

For parameter training, optimal setting procedure is used as described in System Parameters section. Table 4.4 reports the number of training images M used for making background models and parameters used for each test sequence.

Table 4.4. Parameter Values for ESI test sequences.

Sequence	**M**	**N**	$\mathbf{C_{PW}}$	$\mathbf{C_{SI}}$	**C**	$\boldsymbol{\tau_{RGB}}$	$\boldsymbol{\tau_{YCbCr}}$
House	200	35	350	0.65	1.2	5	4
Scene1	250	30	300	0.7	1.25	4	4
Scene2	300	35	400	0.8	1.1	5	4

4.2.3. Quantitative Evaluation

The scores of our proposed approach for all test sequences are tabulated in Table 4.5. A comparison of existing state-of-the-art techniques with our proposed approach on the three test sequences: house, scene1 and scene2 [1] is depicted in Figure 4.2. The techniques include; Eigen background based Statistical Illumination (ESI) [1], Statistical Illumination (SI) [3], Eigen Background (EB) both dynamic and fixed [20][6], Tonal Alignment (TA) [14] and Adaptive Background Mixture Model (ABMM)[19]. The results for these techniques are obtained from [1].

Table 4.5. Precision, Recall and F-Measure for MB^2S.

Sequence	**Precision%**	**Recall%**	$\mathbf{F_1}$ **score**
House	78.46	78.67	78.56
Scene1	83.99	83.07	83.53
Scene2	73.48	75.97	74.70

4.2.4. Qualitative Results

For qualitative results, we choose the ESI technique as benchmark for comparison purposes. Figure 4.3, Figure 4.4 and Figure 4.5 not only present comparative results of our approach on some of example frames from house, scene1 and scene2 test sequences, but also depict 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[36].

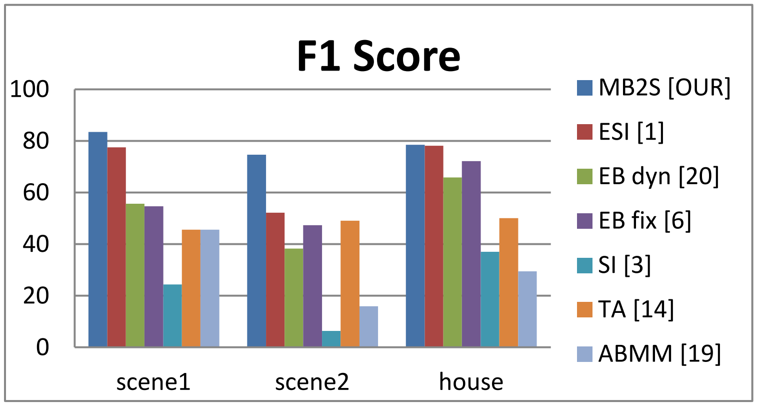


Figure 4.2. F-Measure of Test Sequences.

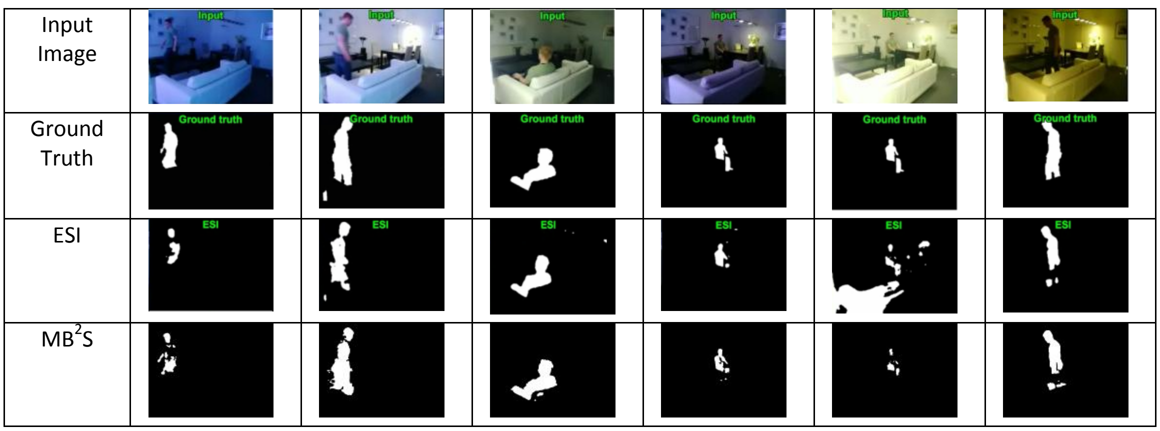


Figure 4.3. Foreground Segmentation results of example frames from test sequence *house*.

Figure 4.4. Foreground Segmentation results of example frames from test sequence *scene1*.

Figure 4.5. Foreground Segmentation results of example frames from test sequence *scene2*.

4.2.5. Discussion on ESI dataset Results

For scene1, scene2 and house test sequences, our proposed approach outperforms all of the other techniques. It should be noted that for scene2, where most of the techniques fail badly, our proposed approach outperforms other schemes with significant difference. We also calculated average F_1 score of our proposed approach and second best technique, which is ESI. The average F_1 score of our approach turned out to be 78.93 and higher than that of ESI's average F_1 score of 69.24.

Chapter 5

Conclusion and Future Work

We have presented a universal BG subtraction system that exploits multiple uni-modal Gaussian BG models and combines the strengths of pixel and frame based binary classifiers in a single framework. The use of two color spaces; RGB and YCbCr has proved that both color spaces combined together provide more accurate foreground segmentation in comparison to using one of the color spaces. Comprehensive evaluation of proposed system over 11 different challenges demonstrates its capability for use in real life applications with an overall 4th position.

In current implementation of our algorithm, it lacks model update, which is part of our future work. Model update is not only expected to improve our position but more importantly make it more robust for real life applications. Another important part of future work is code optimization and implementation of algorithm in C/C++.

For fair comparison, the source code and results will be available at our website[36]. In addition to this, the results will also be made available at changedetection website[23].

Bibliography

[1] L. P. J. Vosters, C. Shan, and T. Gritti." Background subtraction under sudden illumination changes," in *Proc. IEEE Conf. AVSS*, 2010, pp. 384–391.

[2] 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, Jun 2011.

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

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

[5] C. Stauffer and W. Grimson. "Learning patterns of activity using real-time tracking". *IEEE Trans. on Pattern Anal. Mach. Intell*., vol. 22, no. 8, pp. 747–757, 2000

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

[7] S.-C. Cheung and C. Kamath, "Robust techniques for background subtraction in urban traffic video*",* in *Proc. SPIE Video Comm. Image Process.,* 2004, pp. 881-892.

[8] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, "Wall flower: principles and practice of background maintenance", in *Proc. IEEE Conf. ICCV*, 1999, pp. 255–261.

[9] C. Stauffer and W. Grimson. "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Conf. CVPR*, 1999, pp. 2246-2252.

[10] A. M. Elgammal, D. Harwood, and L. S. Davis. "Non-parametric model for background subtraction,". in *Proc. ECCV*, 2000, pp. 751–767.

[11] S. Brutzer, B. Ḧoferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in Proc. IEEE Conf. CVPR, 2011, pp. 1937–1944.

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

[13] K. Skifstad and R. Jain. "Illumination independent change detection for real world image sequences," Comp. Vision, Graphics, and Image Process., vol. 46, no. 3, pp. 387–399, 1989.

[14] L. Di Stefano, F. Tombari, and S. Mattoccia. "Robust and accurate change detection under sudden illumination variations," in *Proc. ACCV Workshop on Multi-dimensional and Multi-view Image Process*., 2007.

[15] F. Tombari, L. Di Stefano, and S. Mattoccia. "A robust measure for visual correspondence," in *Proc. ICIAP*, 2007, pp 376–381.

[16] ESI Dataset [Online]. Available: https://sites.google.com/site/tommasogritti/publications/background-subtraction-data.

[17] L. Li and M. K. H. Leung. "Integrating intensity and texture differences for robust change detection," *IEEE Trans. on Pattern Anal. Mach. Intell.,* vol. 11, no. 2, pp. 105–112, 2002.

[18] X. Zhao, W. He, S. Luo, and L. Zhang. "MRF-based adaptive approach for foreground segmentation under sudden illumination change," in *Proc. ICICS*, 2007, pp 1–4.
[19] P. Kaewtrakulpong and R. Bowden. "An improved adaptive background mixture model for real time tracking with shadow detection," in *Proc. European Workshop on Advanced Video Based Surveillance Systems*, 2001.
[20] B. Han and R. Jain. "Real-time subspace-based background modeling using multi-channel data," in *Proc. Int. Symposium on Visual Computing*, 2007, pp. 162–172.
[21] T. Bouwmans, "Recent Advanced Statistical Background Modeling for Foreground Detection: A Systematic Survey", *Recent Patents on Computer Science*, Vol. 4, no. 3, pp. 147-176, Sept 2011.
[22] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, changedetection.net: A new change detection benchmark dataset, in *Proc. IEEE Workshop on Change Detection*, 2012.
[23] Changedetection Dataset [Online]. Available: www.changedetection.net.
[24] R. Wang, F. Bunyak, G. Seetharaman and K. Palaniappan "Static and Moving Object Detection Using Flux Tensor with Split Gaussian Models", in *proc of IEEE Workshop on Change Detection*, 2014, to be published.
[25] P-L St-Charles, G-A Bilodeau and R. Bergevin "Flexible Background Subtraction With Self-Balanced Local Sensitivity" in *proc of IEEE Workshop on Change Detection*, 2014, to be published.
[26] Theo Gevers, Arjan Gijsenij, Joost van de Weijer, Jan-Mark Geusebroek, *Color in Computer Vision: Fundamentals and Applications*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2012. pp. 13-18.
[27] C. Jin, G. Chang, W. Cheng, and H. Jiang, "Background extraction and update method based on histogram in YCbCr color space", in *Proc. Int. Conf. E-Business and E-Government*, 2011, pp. 1 - 4.
[28] Kristensen F., Nilsson P., Öwall V. "Background Segmentation Beyond RGB," in *Proc. ACCV*, 2006, pp. 602-612.
[29] T. Horprasert, D. Harwood, L.S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proc. IEEE ICCV Frame-Rate Workshop*, 1999.
[30] Chen Z., Pears N., Freeman M., Austin J. "Background subtraction in video using recursive mixture models, spatio-temporal filtering and shadow removal", in *Proc. Int. Symposium on Visual Computing*, 2009, pp. 1141–1150.
[31] M. Balclar, F. Karabiber, A.C. Sonmez. "Performance Analysis of Lab2000HL Color Space for Background Subtraction", *in Proc. IEEE Int. Symposium INISTA*, 2013, pp. 1-6.
[32] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati,"Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 25, no. 10, pp. 1337-1442, 2003.
[33] M. De Gregorio and M. Giordano "Change Detection with Weightless Neural Networks", in *proc of IEEE Workshop on Change Detection*, 2014, to be published.
[34] M.Sedky, M.Moniri and C. C. Chibelushi "Spectral-360: A physical-based technique for change detection", in *proc of IEEE Workshop on Change Detection*, 2014, to be published.

[35] B. Wang and P. Dudek "A Fast Self-tuning Background Subtraction Algorithm", in *proc of IEEE Workshop on Change Detection*, 2014, to be published.
[36] https://sites.google.com/a/g.uky.edu/hasansajid/research
[37] P. Varcheie, M. Sills-Lavoie, and G.-A. Bilodeau, "A multiscale region-based motion detection and background subtraction algorithm", *Sensors*, vol. 10, pp. 1041–1061, Jan. 2010.
[38] Zivkovic, Z., van der Heijden, F ."Efficient adaptive density estimation per image pixel for the task of background subtraction", *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006.
[39] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction", in *Proc. IEEE Int. Conf. Image Process.*, 2004, pp. 3061–3064.
[40] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model", in Proc. Real-Time Imag., vol. 11, no. 3, pp. 172–185, 2005.
[41] M. Seki, T. Wada, H. Fujiwara, and K. Sumi, "Background subtraction based on cooccurrence of image variations", in *Proc. IEEE Int. Conf. CVPR*, 2003, pp. 65–72.
[42] P. Power and J. Schoonees, "Understanding background mixture models for foreground segmentation," in *Proc. Image Vis. Comput.*, 2002, pp. 267–271.

Vita

Hasan Sajid was born in Lahore, Punjab, Pakistan.

EDUCATION:

National University of Sciences and Technology, Pakistan 10/2007

B.S. in Mechatronics Engineering

PROFESSIONAL POSITIONS:

Research Assistant, MIA Laboratory, University of Kentucky, 2014

Team Lead(Mechatronics), National University of Sciences and Technology, 2011

HONORS AND AWARDS:

US State Dept. Fulbright scholarship for MS program, University of Kentucky, 2012

Power and Energy Institute Scholarship